

## Часть 1

Лектор - профессор Колчанов Николай Александрович, заведующий лабораторией Теоретической генетики ИЦиГ СО РАН, заместитель директора по науке.

В течение первого семестра предполагается чтение курса лекций, «Введение в информационную биологию». Этот курс лекций читается в связи с открытием на Факультете естественных наук новой кафедры - Кафедры информационной биологии. (Рис. 1).

Что же такое информационная биология и почему возникла необходимость создания такой кафедры? Как знают многие из Вас, в 1953 году Уотсон и Крик расшифровали структуру двойной спирали ДНК. Этот год многие рассматривают как дату возникновения молекулярной биологии. В последующие годы в молекулярной биологии и генетике со все возрастающей скоростью происходили важнейшие открытия и революционные изменения. К ним относятся: создание методов расшифровки пространственной структуры белков с помощью рентгено-структурного анализа; создание методов расшифровки (чтения) аминокислотных и нуклеотидных последовательностей; создание методов генетической инженерии; разработка методов трансгенеза, клонирования (рис.2.)

С начала 90-х годов прошлого века наступила эпоха массовой расшифровки геномов, увенчанная таким выдающимся достижением как расшифровка генома человека. На этом рисунке вы можете видеть изображение одной из фабрик по автоматической расшифровке геномов. Подавляющее большинство процедур клонирования и секвенирования (расшифровки) геномной ДНК полностью автоматизированы и на обслуживании оборудования, стоимостью до нескольких сотен миллионов долларов, обеспечивающего собственно расшифровку геномной ДНК работает, как правило, лишь немного людей. Окончательные результаты секвенирования хранятся в памяти компьютера в виде последовательности нуклеотидов (рис.3).

В результате появления принципиально новых экспериментальных технологий в молекулярной биологии и генетике за последние 20 лет произошел информационный взрыв. В результате осуществления крупномасштабных геномных проектов расшифрованы структуры геномной ДНК тысяч вирусов, многих сотен бактерий, геномы дрожжей, червя *S.elegans*, плодовой мушки дрозофилы, ведется интенсивная расшифровка геномов ряда сельскохозяйственных животных, расшифрованы геномы ряда растений. Важнейшим событием начала 2001 года было завершение черновой расшифровки генома человека (рис.4).

Объемы получаемых данных поражают воображение. Для того, чтобы вы физически представили их, предположим, что мы записываем геномную последовательность на страницах книги, по 2 тысячи символов на страницу. Тогда геном *E. Coli* длиной более 5 миллионов п.о. может быть записан в книге объемом 500 страниц. Для записи генома дрожжей длиной около 10 миллионов пар оснований потребуется 10 таких книг, для записи генома дрозофилы размером 100 миллионов п.о. требуется 100 таких томов, а для записи информации о геноме человека длиной около 3 миллиардов п.о. - порядка 3000 тысяч томов, то есть огромная библиотека (рис.5).

Рассмотрим несколько подробнее геном человека, содержащий около 35 тысяч генов. При его расшифровке помимо собственно нуклеотидной последовательности получены данные

о цитогенетических и физических картах хромосом, их нуклеотидных последовательностях, локализации генов, устойчивых полиморфизмах, то есть мутациях, присутствующих в локальных популяциях человека с частотами не менее 3-5%. К настоящему времени выявлено не менее 1.5 миллиона мутационных полиморфизмов, по которым геномы людей отличаются друг от друга. Суммарный объем только первичных данных по различным аспектам организации генома человека составляет десятки терабайт (рис.6).

К настоящему времени расшифрованы аминокислотные последовательности миллионов белков и с использованием методов рентгеноструктурного анализа и ядерного магнитного резонанса определены пространственные структуры более 15 тысяч белков (рис.7).

Революционный прорыв в изучении экспрессии генов был достигнут в связи с разработкой технологий ДНК-чипов, которые позволяют количественно измерять экспрессию десятков тысяч генов одновременно в отдельном организме или даже отдельной клетке. Каждая точка на микропластинке, изображенной на рисунке, соответствует отдельному гену. Она содержит определенный олигонуклеотид, обладающий способностью специфически взаимодействовать на основе комплементарных взаимодействий с мРНК, кодируемой соответствующим геном. В каждой точке чипа сигнал гибридизации между мРНК и олигонуклеотидом может быть считан определенным устройством. Тем самым может быть получена информация не только о том, какие гены экспрессируются в исследуемой клетке (ткани, органе), но и о количественном уровне экспрессии (рис. 8).

Итак, современная биология стала производителем беспрецедентно огромных объемов экспериментальных данных, накапливаемых в компьютерных базах данных. Осмысливание этих данных принципиально невозможно без привлечения современных информационных технологий и эффективных методов анализа и моделирования биологических систем и процессов. Именно в ответ на этот вывод и возникает новая наука - информационная биология. Объектами исследований информационной биологии являются генетические макромолекулы - ДНК, РНК, белки, фундаментальные генетические процессы - репликация, транскрипция, трансляция, генетические сети, функционирование которых обеспечивает выполнение всех функций организмов (рис. 9).

Информационные ресурсы, которыми оперирует современная биология, охватывают все экспериментально получаемые данные, характеризующие организацию биологических систем от молекулярного уровня (первичные и пространственные структуры белков, последовательности генов и геномов, генетические, цитогенетические и физические карты геномов), включая экспериментальные данные об индивидуальных генотипических и физиологических особенностях индивидуумов, информацию о путях передачи сигналов и генетических сетях, и заканчивая эволюционно-популяционными особенностями организации биологических систем (популяционный эволюционный полиморфизм, данные по молекулярной эпидемиологии и т.д) (рис.10).

Как следует из сказанного, принципиальная особенность современной биологии состоит в том, что она, образно говоря, стоит на «двух китах» - экспериментальных исследованиях и использовании высоких информационных технологий, реализованных в огромном разнообразии программного обеспечения для анализа этих данных и моделирования биологических систем и процессов. Существенно, что эти информационные технологии применяются на всех этапах экспериментальной работы, начиная от планирования эксперимента и заканчивая интерпретацией полученных экспериментальных результатов (рис.11).

О значимости и масштабах исследований в области информационной биологии свидетельствует тот факт, что только в ближайшие 4 года в индустриально развитых странах мира (США, Англия, Япония, Германия, Франция) в информационные технологии, ориентированные на молекулярную биологию, генетику, биотехнологию, медицину, агробиологию будет инвестировано более 30 миллиардов долларов (рис.12).

Не менее мощные источники биологической информации возникают при исследовании экосистемного и биосферного уровней организации живой материи. Без информационных технологий немислимы, в частности, современные исследования в области биоразнообразия, требующие систематизации и компьютерного описания флоры и фауны, результаты которых накапливаются в компьютерных БАЗАХ ДАННЫХ. На этом рисунке вы видите в качестве примера описание одного из растений в базе данных редких и исчезающих видов растений Сибири (рис.13).

Еще одним мощным источником информации об экосистемах и глобальном состоянии биосферы является космический мониторинг, для обработки результатов которого требуются суперкомпьютеры с огромной оперативной и внешней памятью. На этом рисунке, в частности, вы видите модельную карту пожароопасных зон юга Красноярского края, построенную по результатам космического мониторинга. Наиболее пожароопасные зоны отмечены желтым цветом (рис.14).

Кафедра информационной биологии создается для подготовки специалистов, владеющих методами биоинформатики, теоретического анализа и компьютерного моделирования, необходимыми для решения широкого круга фундаментальных и прикладных проблем молекулярной биологии, молекулярной генетики, клеточной биологии, физиологии, биофизики, общей биологии, биомедицины, фармакологии, экологии, возникающих на стыке с математикой, информатикой и физикой (рис.15).

Базируясь на факультете естественных наук, Кафедра будет, прежде всего, готовить выпускников по специальности «Биология» со СПЕЦИАЛИЗАЦИЕЙ по информационной биологии. Первый набор студентов-биологов на Кафедру информационной биологии будет проходить на 3-м курсе во втором семестре настоящего учебного года. Одновременно с этим, кафедра будет выступать в качестве центра координации и организации учебного процесса для студентов и магистрантов других факультетов НГУ (ММФ, ФФ, ФИТ), желающих специализироваться в информационной биологии, то есть на стыке между биологией, физикой, математикой и информатикой. Предусмотрено чтение дополнительных или альтернативных курсов, проведение семинаров и практических занятий, руководство дипломными работами в области информационной биологии. Распределившись на соответствующие кафедры этих факультетов, студенты математики, физики и информатики смогут в рамках индивидуальных учебных планов, согласованных со своими деканатами, получать необходимые знания, слушая курсы и проходя практические занятия на кафедре информационной биологии. Соответствующие договоренности об этом между всеми факультетами уже имеются (рис.16).

Базовым институтом Кафедры информационной биологии будет Институт цитологии и генетики СО РАН, в котором работает более 70 ученых этой специальности. Центром этой активности в институте является возглавляемая мною лаборатория теоретической генетики, которая выполняет и координирует более 10 крупных Российских и международных проектов в области биоинформатики. Вся исследовательская деятельность лаборатории проводится в тесной кооперации с более чем 10 биологическими, тематическими, физическими и информационными институтами Сибирского отделения РАН, специалисты которых работают с нами в рамках общих проектов (рис.17).

На следующих рисунках (рис.18-21), заканчивающих вводную часть первой лекции, представлен перечень основных учебных дисциплин, которые будут читаться студентам кафедры информационной биологии.

И еще одно важное замечание. Курс лекций будет читаться коллективом ученых институтов Новосибирского Научного Центра, работающих совместно работающих над крупными проблемами информационной биологии. По сути дела, эти ученые будут представлять около 10 Институтов СО РАН, в которых ведутся исследования в области информационной биологии (биоинформатики).

## Часть 2

Я начну вторую часть лекции с рассмотрения центрального объекта информационной биологии - биологических самовоспроизводящихся систем. Это понятие включает три сущности, которые мы рассмотрим одну за другой: (1) система; (2) самовоспроизводящаяся система; (3) биологическая самовоспроизводящаяся система.

Итак, что такое система? любая совокупность элементов, взаимодействующих друг с другом на основе обмена потоками вещества, энергии и (или) информации. Таким образом, диагностическим критерием для рассмотрения совокупности элементов как системы, является наличие между ними взаимодействий определенных типов (рис.23).

Системы бывают двух типов: (а) замкнутые, когда отсутствует взаимодействие между ними и окружающей средой и (б) открытые, когда через определенные элементы (входы и выходы) осуществляется обмен потоками вещества, энергии и информации с окружающей средой (рис.24).

Не претендуя на общность, под самовоспроизводящейся будем понимать любую систему, способную к воспроизведению своих копий на основе содержащейся в ней наследственной информации. По определению, самовоспроизводящиеся системы являются открытыми и неравновесными. Воспроизведение подобных себе копий является одним из результатов взаимодействия этих систем со средой. В настоящее время известно два типа самовоспроизводящихся систем: (а) биологические и (б) социальные (рис.25).

Социальных СВ систем я коснусь несколько позднее. Что же касается биологических СВ систем, то их характерной особенностью является запись подавляющей части наследственной информации с использованием двух типов кодирующих биополимеров ДНК или/и РНК. Способность к самовоспроизведению относится к числу фундаментальных свойств биологических систем, независимо от уровня их сложности и особенностей организации.

На этом рисунке (рис.26) приведены примеры реальных биологических самовоспроизводящихся систем различного уровня сложности: (1) •вирус или фаг - простейшая СВ система, воспроизводящая свои копии на основе информации, записанной в их геномах, но с использованием молекулярно- генетического аппарата клетки-хозяина; (2) •одноклеточный про- или эукариотический организм - СВ система минимального уровня сложности, способная к полностью автономному воспроизведению своих копий;

(3) •пара разнополых особей - минимальная единица самовоспроизведения для животных и растений с половым процессом размножения; (4) •популяция - св система, способная к устойчивому (по отношению к изменениям окружающей среды) воспроизведению; (5) •биогеоценоз - совокупность популяций, существующих на определенной территории, принадлежащих различным биологическим видам и объединенных иерархией трофических связей; (6) •биосфера земли - глобальная СВ системами иерархически наиболее высокого уровня (рис.26).

Для всех биологических СВ систем, независимо от особенностей их структурной организации характерно наличие ряда инвариантных подсистем, выполняющих одни и те же функции, абсолютно необходимые для их жизнедеятельности (Рис. 27): (1) наследственная память, (2) система считывания информации из наследственной памяти, (3) подсистема самовоспроизведения, инструктируемая наследственной памятью совокупность механизмов, структур и процессов, обеспечивающих воспроизведение копий СВ системы, (4) исполняющая подсистема - совокупность механизмов, структур и процессов, обеспечивающих выполнение всех жизненно-важных функций СВ системы, также инструктируемая наследственной памятью. Исполняющая подсистема - совокупность механизмов, структур и процессов, обеспечивающих функционирование СВ системы. (реализуется через функцию генов сетей); (5) репарирующая подсистема, обеспечивающая восстановление поврежденных блоков наследственной памяти; (6) подсистема безопасности, обеспечивающая защиту СВ системы от проникновения в нее других СВ систем и чужеродной генетической информации; (7) подсистема рецепции, оценивает величины параметров внешней и внутренней среды, значимые для функционирования и воспроизведения СВ системы. Важнейшую роль в функционировании СВ систем играет (8) подсистема регуляции, обеспечивающая разворачивание в пространстве и во времени процессов, инструктируемых наследственной памятью.

Являясь открытой, СВ система получает два типа возмущающих воздействий со стороны окружающей среды: (1) возмущения через входные каналы (качественные и количественные изменения потоков вещества, энергии и информации); (2) мутации наследственной памяти системы.

В качестве напоминания привожу следующие рисунки (рис.28, 29), на которых отображены потоки генетической информации и фундаментальные генетические процессы в современном представлении.

Центральным блоком любой СВ системы является ее наследственная память (рис.30). Наследственная память необходима СВ системе для: 1) воспроизведения ее копий; 2) обеспечения функционирования СВ системы в типичных для нее условиях среды; 3) реакции СВ системы на изменяющиеся условия внешней среды; 4) замены состарившихся или поврежденных соматических элементов на новые, и выполнения ряда других функций. Наследственная память биологических СВ систем подразделяется на два типа: (1) генетическая, записанная в геномной ДНК или РНК и (2) эпигенетическая, записанная в виде концентраций определенных веществ в клетке. Эпигенетическая память подразделяется на два типа - позиционированная и непозиционированная (рис.31).

В многоклеточных СВ системах генетическая память подразделяется на память соматических и генеративных элементов. Генеративные элементы обеспечивают хранение и реализацию наследственной информации, необходимой для воспроизведения копии СВ

системы. Соматические элементы обеспечивают функционирование и поддержание структурной организации СВ систем (рис.32).

В большинстве биологических СВ систем материальным носителем генетической памяти является дезоксирибонуклеиновая кислота - ДНК. ДНК большинства организмов - это длинная двухцепочечная молекула. В отдельной нити ДНК мономеры - пуриновые основания (аденин -А и гуанин - G) и пиримидиновые основания (цитозин С и тимин Т) прикреплены к полимерному остову, состоящему из чередующихся остатков фосфата (Р) и сахара (дезоксирибозы). Таким образом, молекула ДНК формируется из мономеров четырех типов и на информационном уровне ей может быть сопоставлен генетический текст в алфавите {А,Т,С,Г} (рис.33).

Основания аденин и тимин с одной стороны и гуанин и цитозин - с другой, обладают способностью к комплементарному взаимодействию на основе формирования водородных связей (рис 33,А). В результате формируется стабильная двухнитевая молекула ДНК, в которой комплементарные нити связаны водородными связями.

На рисунке 34,А представлено схематическое изображение двойной спирали ДНК. Поперечные перекладки - комплементарные пары оснований, «боквинны» - сахарофосфатный остов. На рисунке 34,Б представлено стереохимическое строение двойной спирали ДНК (3.4 п.о. на один виток спирали). В спирали ДНК имеется две бороздки - малая и большая.

Удвоение ДНК-матрицы (воспроизведение генетической информации) происходит в результате фундаментального генетического процесса - репликации. Репликация проходит по полуконсервативному механизму, так что по каждой из нитей ДНК достраивается ей комплементарная. Не останавливаясь на деталях, отметим, что в ходе этого процесса происходит плавление ДНК, сопровождающееся разрывом водородных связей, и по каждой из нитей ДНК достраивается комплементарная ей вторая нить двойной спирали ДНК. Итак, физико-химический смысл репликаций - матричный синтез молекулы ДНК по ДНК. Информационный смысл репликации- самовоспроизведение генетической ДНК-памяти в исходном алфавите {А, Т, С, Г} (рис.35).

По мере увеличения сложности СВ систем все большую роль в их функционировании и развитии играет эпигенетическая наследственная память. На качественном уровне эпигенетическая память может быть определена как переданная клетке-потомку от родительской клетки концентрация определенного вещества (веществ), определяющая начальные условия функционирования молекулярно-генетических систем клетки). Эпигенетическая память бывает двух типов: (1) позиционированная, когда концентрация вещества (носителя эпигенетической информации) важна в строго определенных местах клетки и (2) непозиционированная, когда концентрация вещества важна вне зависимости от его положения в клетке (рис.36). На рисунке представлен известный пример распределения в яйцеклетке дрозофилы морфогенов (носителей эпигенетической информации), значимых для ранних стадий дробления яйцеклетки и активации определенных групп генов. При созревании яйцеклетки в организме матери формируются градиенты ряда веществ: (1) градиент РНК(белков) гена bicoid; 2) градиент белка nanos; (3) градиент белка torso и некоторых других. Известно, что морфогены взаимодействуют с регуляторными участками ряда генов, изменяя их функцию. Поэтому понятно, что после оплодотворения яйцеклетки и ее дробления до стадии бластодермы, каждая клетка в ней будет занимать определенное место по отношению к исходным градиентам морфогенов. В

силу этого активность ряда генов в клетках-потомках, возникших из различных частей яйцеклетки, будет существенно различна.

Как отмечалось выше, еще одним инвариантным блоком СВ систем является считывающе-перекодирующее устройство - совокупность механизмов, структур и процессов, ответственных за считывание информации из наследственной памяти СВ систем и ее перекодирование в различные типы генетических сообщений, необходимых для функционирования СВ системы (рис.37).

Первый из процессов перекодирования генетической информации - это транскрипция, осуществляемая РНК-полимеразой. В результате транскрипции на основе информации, закодированной в ДНК, и соответствующей определенному гену, синтезируется молекула РНК. РНК образуют второй (наравне с ДНК), класс матричных биополимеров, то есть генетических макромолекул, в которых генетическая информация кодируется в виде последовательностей нуклеотидов. Физико-химический смысл транскрипции - матричный синтез РНК по ДНК. Информационный смысл - перекодирование фрагмента генетического ДНК-текста, соответствующего гену (или оперону - в случае прокариот) и записанного в наследственной памяти СВ системы в алфавите {А, Т, G, С }, в генетический РНК-текст, записанный в алфавите {А, U, G, С } в динамической памяти СВ системы. В ходе транскрипции РНК-полимераза денатурирует двухцепочечную спираль ДНК и вновь восстанавливает ее, поддерживая кодирующую и матричную цепи в состоянии необходимом для матричного синтеза РНК (рис.38).

Следующий фундаментальный генетический процесс, имеющий отношение к преобразованию генетической информации - это сплайсинг. Он является абсолютно необходимым этапом экспрессии большинства генов эукариот. Гены эукариот имеют экзон-интронную структуру, в которой кодирующие участки - экзоны, чередуются с некодирующими - интронами. Интроны транскрибируются наравне с экзонами, поэтому пре-мРНК содержит участки, транскрибированные как с интронов, так и с экзонов. В процессе сплайсинга участки пре-мРНК, которые соответствуют интронам, вырезаются, а соседние экзоны «сшиваются». Поэтому окончательный продукт сплайсинга - зрелая мРНК, содержит только кодирующие участки - транскрипты экзонов. Таким образом, физико-химический смысл сплайсинга - это : вырезание фрагментов из пре-мРНК. Информационный смысл - удаление из генетического РНК-текста фрагментов, не несущих генетической информации. Биологический смысл экзон интронной организации генов эукариот будет обсуждаться в одной из следующих лекций. Здесь же я скажу, что интроны, хотя и не являются белок-кодирующими последовательностями, выполняют множество важных функций, связанных, в первую очередь с регуляцией экспрессии генов (рис.39).

Третий класс матричных биополимеров (генетических макромолекул) - это белки. Белки синтезируются на основе информации, закодированной в мРНК в результате процесса трансляции, и образованы из мономеров 20 типов - аминокислот. Объединяясь стандартным способом, аминокислоты формируют полипептидные цепи - аминокислотные последовательности (рис.40).

Важнейшую функцию в СВ системах выполняет система репарации (рис.41), осуществляющая исправление дефектов наследственной памяти, возникающих как в ходе воспроизведения наследственной памяти, так и при ее оперативном функционировании. В биологических СВ системах основу ее функционирования обеспечивают ферментные комплексы репарации ДНК. Генетическая память биологических СВ может повреждаться, прежде всего, в ходе оперативного функционирования. На физико-химическом уровне возникновение повреждений генетической памяти проявляется в химических

модификациях ДНК, возникновении некоплементарных пар ДНК, химическом разрыве ковалентных связей в сахаро-фосфатном остове ДНК и т.д. Существуют специальные ферментные системы, устраняющие повреждения ДНК. Например, система эксцизионной репарации E. Coli работает следующим образом (рис.42). Опознав повреждение ДНК, эксцизионная система делает надрезы в поврежденной цепи ДНК на небольшом расстоянии как справа, так и слева от места повреждения. Вырезанный одонитевой фрагмент ДНК удаляется и разрушается экзонуклеазами. После этого возникшая брешь заполняется ферментом ДНК-полимераза 1, а ДНК-лигаза зашивает оставшиеся разрывы.

Сложная проблема возникает при исправлении ошибок, возникающих в генетической памяти в процессе воспроизведения - репликации ДНК. Возникшая некоплементарная пара С-С должна быть удалена и заменена на одну из двух комплементарных пар С-Г или Г-С. Как же решается проблема выбора правильной комплементарной пары? Родительская ДНК метилирована по аденину в олигонуклеотидном сигнале GATC. Именно наличие метилированных аденинов и отличает родительскую нить ДНК от вновь синтезированной нити в процессе репликации. Существует короткий промежуток времени между репликацией и метилированием ДНК. В это время репарационные системы распознают нарушения комплементарности и исправляют вновь синтезированную нить по родительской. Через небольшой промежуток времени происходит метилирование вновь синтезированных нитей ДНК, после чего они маркируются как родительские, что необходимо для правильной репарационной коррекции на следующем цикле репликации (рис.43).

Еще один важный блок биологических СВ систем - подсистема генерации генетической изменчивости (рис.44). Наличие необходимого уровня изменчивости генетической памяти - критическое условие для выживания СВ систем при резких изменениях условий окружающей среды (рис.45). Одним из фундаментальных генетических процессов генерации генотипической изменчивости является гомологичная рекомбинация (рис.46). Она происходит между гомологичными молекулами ДНК. Как показано на этой схеме она может, например, инициироваться одонитевым разрывом в одной из нитей ДНК. За этим следует плавление дуплекса ДНК вблизи разрыва, и освобождение одонитевого фрагмента ДНК. Взаимодействуя с гомологичным участком двойной спирали ДНК, этот свободный фрагмент вытесняет из него гомологичную ему нить, создавая тем самым переходную структуру для процесса рекомбинации.

Не останавливаясь на деталях этого процесса, укажем только, что в итоге происходит обмен крупными блоками ДНК между гомологичными хромосомами (рис.47). Таким образом, на основе гомологичных рекомбинаций происходит перетасовке генетического материала в пределах гомологичных хромосом. Фактически, в эволюционной шкале времени хромосома - это непостоянная структура, образуемая временно связанными блоками. Такое непостоянство, обусловленное рекомбинациями, позволяет формироваться новым группам сцепления генов в геномах потомков, обеспечивая тем самым возможность проявления новых фенотипических признаков.

Следует подчеркнуть, что элементарной единицей эволюции СВ систем является популяция. Популяция - это совокупность СВ систем, обменивающихся наследственной информацией и образующих самовоспроизводящуюся единицу иерархически более высокого уровня. Популяция существенно повышает вероятность сохранения наследственной информации СВ систем при изменяющихся условиях внешней среды. Чем ближе у конкретной особи величина фенотипической характеристики  $X_i$  к величине  $X_0$ , оптимальной для данных средовых условий, тем больше ее приспособленность  $W_i$ . Генотипическое разнообразие является необходимым условием невырождения популяций при изменяющихся условиях внешней среды. Действительно, при резких изменениях



среды оптимальным становится другое значение критической фенотипической характеристики. Особи, адаптированные к старым условиям среды, теряют свою приспособленность и гибнут. В этой ситуации имеют шанс выжить только «изгои», то есть те особи, которые в исходных условиях имели существенно отличающиеся от нормы значения фенотипической характеристики и, как следствие этого, низкую приспособленность. Именно эти изгои и обеспечат выживание популяции в новых средовых условиях с последующей эволюцией фенотипических характеристик к новым оптимальным условиям среды. В этом случае, чем выше темпы мутирования, тем больше размах возникающей фенотипической изменчивости, и как следствие этого, вероятность выживания в новых средовых условиях (рис.45).

Подсистема безопасности (рис.48), обеспечивает: (1) защиту СВ системы от проникновения в нее чужеродной генетической информации, (2) подавляет развитие паразитических подсистем в пределах данной СВ системы; (3) предотвращает «смешивания» наследственной памяти данной СВ системы с наследственной памятью других СВ систем, то есть осуществляет функцию распознавания и уничтожения «чужого» (рис.49). Одним из примеров реализации подсистем безопасности является иммунная система, распознающая чужеродные антигены (рис.50). Функция иммунной системы основана на функционировании кластеров иммуноглобулиновых генов. В результате комбинирования V, D, J и C модулей генерируется огромное разнообразие экспрессирующихся иммуноглобулиновых генов, кодирующих различные варианты антител, в том числе и такие, которые обладают способностью специфически связывать и нейтрализовать чужеродные антигены. Разнообразие иммуноглобулинов увеличивается на много порядков за счет процессов соматического мутагенеза иммуноглобулиновых генов.

Следует подчеркнуть, что разнообразие антител, способных нейтрализовать гигантское разнообразие потенциально возможных антигенов, принципиально не может быть закодировано в геноме любого размера. В этом случае в наследственной памяти генеративных клеток СВ системы хранится не сама по себе информация об антителах, а «программа» генерации этого разнообразия. Существенно при этом, что генерация разнообразия антител происходит не в наследственной памяти генеративных клеток, а в соматических клетках

Рассмотрим в заключение этого краткого обзора основных компонент СВ систем исполняющие и регуляторные подсистемы (рис.51). Они реализуются через функцию генных сетей (рис.52). Генная сеть - группа координированно функционирующих генов, контролирующих физиологические, биохимические и молекулярные функции организмов. В качестве примера на рисунке 53 показан центральный фрагмент генной сети, обеспечивающей биосинтез холестерина и постоянство его концентрации в клетках. Путь биосинтеза холестерина из ацетил-коэнзима А контролируется по механизму отрицательной обратной связи. Центральный регулятор генов, кодирующих ферменты этого пути - транскрипционный фактор SREBP, активирующий транскрипцию каскада этих генов и тем самым усиливающий продукцию холестерина. Фактор SREBP образуется из предшественника - preSREBP под действием стерол-зависимой протеазы. При повышении уровня холестерина активность протеазы подавляется, что снижает скорость образования фактора SREBP и его концентрацию. Тем самым снижается активность генов, кодирующих ферменты этого пути, и уровень холестерина нормализуется. Так работает отрицательная обратная связь, контролирующая концентрацию холестерина. Всего эта генная сеть включает несколько сотен генов.

Теперь поговорим о программах исследования геномов, чтение которых стало рутинной процедурой молекулярной генетики (рис.54). Ряд геномов уже прочитан, другие, например, геном человека, близки к расшифровке. В связи с этим многие ученые еще не

так давно думали, что битва за выяснение молекулярных основ наследственности уже выиграна. Однако, реальность оказалась другой: перед глазами предстают бесконечные последовательности нуклеотидов, смысл которых понятен на доли процента.

Основой процесса трансляции является генетический код (рис.55). Я не буду останавливаться на многих замечательных свойствах генетического кода, так как этим проблемам будет посвящена специальная лекция. Здесь же отмечу его главные свойства (рис.56,57):

- (1) генетический код триплетен, то есть каждой аминокислоте соотносится набор определенных троек нуклеотидов;
- (2) Генетический код неразрывен, то есть между нуклеотидами, входящими в триплеты отсутствуют пробелы;
- (3) Все кодоны однозначны, то есть каждый кодирует единственную аминокислоту;
- (4) Совместно транслируемые кодоны не перекрываются и не имеют разделительных знаков в тексте: 5'-AUG GGG GGA GAU GAG CUA-3'
- (5) Генетический код вырожден, то есть большинство аминокислот кодируется несколькими кодонами
- (6) Различные аминокислоты кодируются различным количеством кодонов (от 1 до 6)
- (7) Генетический код содержит знаки пунктуации, отмечающие начало и конец рамки трансляции: 5'-AUG GGG GGA GAU CUA-UAG: 3'
- (8) Генетический код универсален.

Молекулярный аппарат трансляции содержит следующие базовые компоненты (рис.58):

- (а) мРНК, содержащая информацию об аминокислотной последовательности белка;
- (б) набор тРНК, антикодоны которых комплементарно взаимодействуют с кодонами мРНК;

(в) рибосома, состоящая из двух субъединиц, и имеющая два активных сайта А и Р, соответственно;

(г) набор аминоктил-тРНК-синтетаз, обеспечивающих специфическое распознавание соответствующих тРНК и присоединяющих к их 3'-концам определенные аминокислоты.

Схематически процесс трансляции представлен на рисунке 59. тРНК, к которой присоединена растущая полипептидная цепь, находится в Р-сайте рибосомы и взаимодействует своим антикодоном с комплементарным кодоном мРНК. На следующем этапе заряженная тРНК, то есть тРНК, к которой присоединена определенная аминокислота, взаимодействует с своим антикодоном с соседним кодоном мРНК (в 3'-направлении), находящемся в А-сайте рибосомы. Далее растущая полипептидная цепь присоединяется к аминокислоте, находящейся в А-сайте, а рибосома сдвигается на один триплет к 3'-концу мРНК, так что тРНК, к которой присоединена растущая полипептидная цепь, вновь оказывается в Р-сайте рибосомы. После чего весь цикл трансляции повторяется.

Таким образом, физико-химический смысл процесса трансляции - это синтез полипептидной цепи по РНК-матрице. Информационный смысл - перекодирование генетического сообщения о структурно-функциональной организации белка, записанного в РНК-тексте в алфавите {A, U, G, C} в белковый текст, записанный в 20-буквенном аминокислотном алфавите {A, C, D, E, F, ..., W}. Биологический смысл перекодирования информации состоит в том, что белки - матричные биополимеры, образованные из аминокислот, обладают существенно большим спектром биологических активностей по сравнению с двумя другими типами матричных биополимеров - ДНК и РНК. Видимо поэтому в ходе эволюции они были отобраны как важнейшие компоненты исполняющих устройств биологических СВ систем, в то время как ДНК и РНК - были отобраны, главным образом, для выполнения функций хранения и передачи генетической информации. Более детально эволюционные мотивации распределения функциональной нагрузки между тремя типами матричных биополимеров будут рассмотрены в одной из следующих лекций.

На рисунке 60 приведен простейший вариант канонической схемы, в рамках которой рассматриваются вопросы кодирования в связи с надежностью (помехоустойчивостью) информации. Эта схема включает следующие элементы: (1) источник информации; (2) коды записи сообщений; (3) канал передачи информации; (4) приемник информации. Проблема состоит в том, что любой канал передачи информации работает с шумом. В технических и естественных биологических системах преимущество имеют те способы кодирования информации, которые обеспечивают максимальную надежность (помехоустойчивость) передаваемой информации.

Формальные математические подходы, реализуемые в рамках теории информации, начиная от классических работ Клода Шеннона - основоположника теории информации, внесшего в 30-х 40-х годах прошлого века решающий вклад в становление этой теории,

до современных подходов, разрабатывающихся в настоящее время, будут описаны в одной из следующих лекций. Здесь же я хочу рассказать о некоторых качественных аспектах этой проблемы.

Прежде всего, что понимается под каналами передачи информации в биологических СВ системах? Укажем, по крайней мере, на три типа таких каналов:

(1) передача информации по внутренним коммуникациям СВ системы в ходе ее функционирования.

(2) Передача информации от родителя (родителей) к потомку в процессе самовоспроизведения системы.

(3) Передача информации клетке-потомку от родительской клетки;

Отметим, что процессы (2) и (3) обязательно сопряжены с копированием наследственной памяти СВ системы, в первую очередь, генетической наследственной памяти, представленной геномной ДНК. В качестве шума в этом случае можно рассматривать мутации, приводящие на этапе репликации ДНК к искажению генетической информации, передаваемой потомку.

В этой связи следует отметить, что триплетный генетический код, в силу особенностей своей организации, обладает высокой помехоустойчивостью. Укажем на два существенных обстоятельства. Во-первых, известно, что аминокислоты в белках представлены неравномерно. Наиболее часто встречающиеся аминокислоты - это глицин и аланин, а наиболее редко встречающиеся - триптофан и метионин. Оказалось, что существует замечательная корреляция между частотами использования аминокислот в белках и количеством кодирующих их кодонов. Это обеспечивает существенную помехоустойчивость генетического кода, так как при мутации по кодону часто встречающейся аминокислоты имеется более высокая вероятность, что этот кодон промутирует в кодон, кодирующий эту же аминокислоту (рис.61).

Анализ свойств генетического кода показывает на еще одну важную особенность, обеспечивающую его помехоустойчивость при одиночных нуклеотидных заменах. Как известно, все аминокислоты можно разделить на полярные, несущие заряженные или поляризованные группы, и гидрофобные. Наиболее сильное повреждающее влияние на функцию белков оказывают мутации, приводящие к изменению физико-химического смысла аминокислоты, то есть, заменам полярных аминокислот на гидрофобные и наоборот. Оказалось, что структура реального генетического кода минимизирует повреждающее влияние мутаций, обеспечивая повышенную частоту аминокислотных замен с сохранением физико-химических свойств аминокислот (рис.62). Действительно, при произвольном замещении любого аминокислотного остатка на любой, то есть без

учета структуры генетического кода, средняя разница в гидрофобности аминокислот составляет 1280 ккал/моль. В то же время, с учетом структуры генетического кода, которая определяет возможности перехода от одной аминокислоты к другой при одиночных нуклеотидных заменах, различие в гидрофобности составляет лишь 870 ккал/мол.

Расшифровка генетического кода - триумф генетики 60-х годов. В то время было трудно представить иные коды, кроме триплетного. Однако, в 70-е годы в проблему генетического кодирования вторглась новая тема. На следующих рисунках (64-69) представлены некоторые характеристики геномов про- и эукариот и строение их генов (оперонов).

Оказалось, что геномы эукариот насыщены некодирующей ДНК: повторами, интронами, межгенными спейсерами, а на долю триплетного кода приходится не более 5% генома. Первоначально некодирующую ДНК рассматривали как паразитическую. Однако, постепенно стала формироваться иная точка зрения (рис. 70). “Преобладающие в геномах эукариот некодирующие последовательности, вероятно, кодируют нечто иное, что не требует привлечения традиционного триплетного кода. Иными словами, кроме триплетного кода и трансляционной машины, в клетке, по-видимому, имеются другие коды и средства их чтения. При этом под кодом понимается любой тип нуклеотидного контекста, значимый для выполнения определенной биомолекулярной функции”. Изучение природы кодов функционирования генетических систем - одна из задач новой науки - информационной генетики. Другие ее задачи: предсказание фенотипических характеристик организмов по информации, закодированной в геномах и создание теоретических основ конструирования генетических систем с заданными свойствами. На рис. 71 представлены примеры нетриплетных кодов записи генетической информации, занимающих не менее 50% геномной ДНК эукариот. Эти коды будут рассмотрены нами подробно в следующей лекции.