Для биологов геном-это то, что они видят под микроскопом. Это набор хромосом Ххр и Yхр, две пары хромосом: одна от мамы другая от папы. Геном для биоинформатика- это набор текстовых файлов, записанных непрерывным набором букв. Говоря о геноме, обычно проводят аналогию с расшифровкой древних рукописей, когда текст известен, а язык — нет. Эта задача неразрешима до тех пор, пока у нас нет никаких представлений о содержании текста. Однако, если мы хотя бы примерно представляем, о чем этот текст, то появляется надежда на его осмысление. В биоинформатике ситуация лучше, чем при расшифровке древних письмен, хотя это тоже трудная задача. Здесь на помощь приходит математика. В гигантском массиве информации с помощью современных математических методов ищутся скрытые закономерности, которые и позволяют находить гены и предсказывать их свойства.

**Геном**- совокупность всех генов организма. Анализ геномов приносит множество новой информации. В настоящее время расшифровано более 200 геномов различных бактерий, каждый из которых содержит несколько тысяч генов. Для того чтобы охарактеризовать один ген, требуется несколько месяцев напряженной работы экспериментаторов. С другой стороны, для того чтобы достаточно подробно описать один бактериальный геном средствами биоинформатики, достаточно примерно месяца работы небольшой группы исследователей.

В геноме человека около 35 тыс. генов (всего в 10 раз больше, чем у бактерии, и в 2 раза больше, чем у плодовой мушки), а количество синтезируемых белков гораздо больше. В чем же дело? Оказывается, что очень часто один ген кодирует несколько разных форм белка. За это отвечает явление, названное ***альтернативным сплайсингом***. Биоинформатика впервые показала, что количество генов, имеющих альтернативный сплайсинг, очень велико. Осталось загадкой, как все это регулируется.

В клетке не все гены должны работать одновременно. Для того чтобы гены работали, как слаженный оркестр, необходимо, чтобы гены включались только тогда, когда их работа необходима. Этим заведует система регуляции генов, анализ которой позволил обнаружить принципиально новые способы регуляции –***рибопереключатели***.

Еще одно направление — исследование эволюции всего живого. Здесь тоже есть много открытий, например, горизонтальный перенос генов между видами. Биоинформатика в некоторых случаях позволяет не только показать эти случаи, но также и датировать их. Биология и биоинформатика являются не только способами познания мира, но имеют и прикладное значение, прежде всего в медицине и биотехнологии. Если рассматривать строение нашего тела детально, то это будет выглядеть так: органы состоят из тканей, ткани из клеток. Клетка в свою очередь имеет довольно сложное строение. Нас интересует ядро.

В ядрах находятся ***хромосомы***, которые по внешнему виду похожи на маленьких червячков: хромосома состоит в основнои из **ДНК**-дезоксирибонуклеиновой кислоты, молекулярной основы генома. ДНК-это длинная молекула, выглядит как закрученная двойная спираль. В ядре любой соматической клетки человека имеется 23 пары хромосом. Каждая хромосома содержит одну молекулу ДНК из двух цепочек. Каждая цепочка построена из нуклеотидов, в каждом по одному из азотистых оснований. Всего азотистых оснований четыре: аденин, тимин, гуанин, цитозин. Для краткости их обозначают первыми буквами слов – А, Т, Г, Ц. Эти буквы – алфавит ДНК. Каждая молекула ДНК представляет собой целый ряд генов. Так как ДНК построена из двух цепочек, то еe длина измеряется числом пар оснований, а каждую пару обозначают словом «буква».

Две цепочки из нуклеотидов с помощью пар оснований соединяются вместе и образуют целую молекулу ДНК. Пары оснований комплементарны друг другу и соединены между собой водородными связями только такими сочетаниями: А-Т или Т-А; Г-Ц или Ц-Г. На протяжении всей длины молекулы ДНК можно видеть такие сочетания оснований. Но порядок размещения пар оснований или их последовательность для каждого гена свой. Именно в последовательности пар оснований или «букв» и заключена генетическая информация. Это инструкция о том, как должна быть построена клетка каждого типа, и как она должна функционировать. Это относится и к организму в целом. То есть генетическая информация всех генов – это «рецепт» построения и функционирования организма человека. Из этого ясны цели открытия строения генома:

1. выявить последовательность пар оснований каждого гена вдоль длины молекулы ДНК;
2. определить место каждого гена и его границы по длине молекулы ДНК. Это необходимо было сделать в каждой хромосоме – с 1-й по 23-ю.

Весь процесс открытия строения генома человека состоял из этапов:

1. разделение молекул ДНК на большие фрагменты и секвенирование (от лат. sequi – следовать), т.е. определение последовательности пар оснований или «букв»;
2. картирование, т.е. идентификация генов и локализация места их расположения на хромосомах.

На первом этапе было обнаружено, что ДНК человека в 23-х хромосомах состоит из 3,2 млрд. пар оснований, т.е. химических «букв». Каждая пара оснований – это «кирпичик», из которых формируются гены. На втором этапе ученые расположили эти пары оснований в правильной последовательности, что позволило «прочитать» каждый ген и произвести инвентаризацию генов и межгенных участков. Так была составлена карта топографии генов на хромосомах. Но эти данные будут еще уточняться и дополняться.

Прямая выгода от знания строения генов станет очевидна не сразу, пока не будут выяснены функции каждого гена в клетке, т.е. что он делает?

До открытия строения генома ученые знали функции ряда генов. В целом учeные знают о функциях не более одной трети генов, а о двух третях их – ничего. То есть учeным ещe предстоит выполнить важнейший третий этап – выяснить функции каждого гена в клетке каждого типа в норме. Важно узнать, какие гены в такой клетке работают, а какие «молчат». Затем ученые то же сделают в дефектной клетке того же типа, например, в раковой клетке. По разнице можно выявить ген-причину, но это лишь один из путей обнаружения раковой клетки.

Что пока дало открытие строения генома?

* + - Оказалось, что в геноме человека всего 26-30 тысяч генов, а не больше, как предполагалось.

Очевидно, не зря в клетках человека обнаружено «расщепление» генов. За счет этого с гена может копироваться не одна иРНК, а больше – в среднем до 10 таких молекул. Отсюда продукт гена – белок не один, а несколько – от 3 до 10 разных белков То есть в схему этапов реализации генетической информации от гена к его основному продукту, следует внести уточнение: ген – иРНК – белки, вместо – «белок» в прежней схеме Считают, что у человека «один ген отвечает за группу функций» или «за одну функцию отвечает сразу целая группа генов». Это будет изучено, так как есть методы определения функций генов.

* Лишь 2% всех генов участвует в синтезе белков в клетке, остальные – 98% длины ДНК до недавнего времени считалось – не являются генами. Это длинные последовательности нуклеотидов, в которых часто повторяется – АГГ, АГГ… Функции этой части генома только начинают изучаться. Но уже ясно, что это не «мусор», как некоторые считали, а часть ДНК, которая регулирует экспрессию генов и их связей между собой.
  + - Небольшое число генов в геноме человека имеет важное практическое значение – легче разобраться в роли генов, которые могут вызывать разные болезни, в частности, рак.
    - Оказалось, что около половины генов являются общими и для человека, и для червячка «C. elegans». Более того, в геноме человека только 300 генов, которых нет в геноме полевой мыши.

Для изучения функций генов широко используется клетки в культуре. Принцип: с помощью введения в клетки антисмысловой РНК блокировать, а теперь можно малой молекулой РНКи разрушить в клетке иРНК, т.е. копию изучаемого гена. По изменению свойств такой клетки можно определить функции этого гена. Открытие общих с человеком генов у червячка и мыши делает эти организмы новыми, очень нужными моделями для определения функций неизвестных у человека генов, но одинаковых с ними. Ведь изучать функции генов на этих организмах проще и легче, чем на человеке. Их гены можно легко изменять, т.е. вызывать мутации или выключать ген, следя за изменениями свойств клетки или свойств организма. На мышах можно использовать и такой прием: в половых клетках удалять один, либо два, либо больше генов и изучать организм таких мышей-мутантов. Это позволит понять какой ген или гены за что отвечают, изучать взаимодействие генов между собой.

* + - Разница между генами мыши и человека не превышает 1%, а между разными людьми – 0,01%.

На выяснение функций всех генов в клетках каждого типа человека уйдет много лет. Когда это будет достигнуто, ученые смогут на уровне нормальной клетки каждого типа установить стандарт или идеал нормы: а) последовательности пар оснований для каждого гена и б) функций каждого гена в клетке.

Генетическая информация, содержащаяся в ДНК и в иРНК, заключена в последовательности расположения нуклеотидов в молекулах. Каким же образом иРНК кодирует (шифрует) первичную структуру белков, т. е. порядок расположения аминокислот в них? Суть кода заключается в том, что последовательность расположения нуклеотидов в иРНК определяет последовательность расположения аминокислот в белках. Этот код называют генетическим, его расшифровка - одно из великих достижений науки. Носителем генетической информации является ДНК, но так как непосредственное участие в синтезе белка принимает иРНК - копия одной из нитей ДНК, то генетический код записан на «языке» РНК.

Код триплетен. В состав РНК входят 4 нуклеотида: А, Г, Ц, У. Если бы мы попытались обозначить одну аминокислоту одним нуклеотидом, то можно было бы зашифровать лишь 4 аминокислоты, тогда как их 20 и все они используются в синтезе белков. Двухбуквенный код позволил бы зашифровать 16 аминокислот (из 4 нуклеотидов можно составить 16 различных комбинаций, в каждой из которых имеется 2 нуклеотида).

В природе же существует трехбуквенный, или триплетный, код. Это означает, что каждая из 20 аминокислот зашифрована последовательностью 3 нуклеотидов, т. е. триплетом, который получил название кодон. Из 4 нуклеотидов можно создать 64 различные комбинации, по 3 нуклеотида в каждой. Почти каждая аминокислота шифруется более чем одним кодоном (от 2 до 6). Это видно из таблицы генетического кода.  Код однозначен. Каждый триплет шифрует только одну аминокислоту.

Между генами имеются знаки препинания. Каждый ген кодирует одну полипептидную цепочку. Поскольку в ряде случаев иРНК является копией нескольких генов, они должны быть отделены друг от друга. Поэтому в генетическом коде существуют три специальные триплета (УАА, УАГ, УГА), каждый из которых обозначает прекращение синтеза одной полипептидной цепи. Таким образом, эти триплеты выполняют функцию знаков препинания. Они находятся в конце каждого гена. Внутри гена нет знаков препинания. Поскольку генетический код подобен языку. Код универсален. Код един для всех живущих на Земле существ. У бактерий и грибов, злаков и мхов, муравья и лягушки, окуня и пеликана, черепахи, лошади и человека одни и те же триплеты кодируют одни и те же аминокислоты.

Реализация генетического кода в клетке происходит в два этапа:

* + - синтез молекулы матричной, или информационной, РНК (соответственно мРНК, или иРНК) на соответствующем участке ДНК; при этом последовательность нуклеотидов ДНК "переписывается" в нуклеотидную последовательность мРНК ;
    - синтез белка при котором последовательность нуклеотидов мРНК переводится в соответствующую последовательность аминокислот (см. Трансляция).

Впервые идея о существовании генетического кода сформулирована А.Дауном и Дж.Гамовым в 1952-1954, которые показали, что последовательность нуклеотидов, однозначно определяющая синтез той или иной аминокислоты, должна содержать не менее трех звеньев. Позднее было доказано, что такая последовательность состоит из трех нуклеотидов, названных кодоном или триплетом. Т.к. молекулы нуклеиновых кислот, на которых происходит синтез мРНК или белка состоят из остатков только четырех разных нуклеотидов, кодонов, отличающихся между собой, может быть всего 64. Все синтезируемые в процессе трансляции белки построены из остатков 20 аминокислот (так называемых кодируемых).

Генетический код специфичен: это означает, что каждый кодон кодирует только одну аминокислоту. Генетический код называют вырожденным, поскольку 61 кодон кодирует всего 20 аминокислот. Поэтому почти каждой аминокислоте соответствует более чем один кодон. Вырожденность генетического кода неравномерна: для аргинина, серина и лейцина она шестикратна (т.е. для каждой из этих аминокислот имеется по шесть кодонов), тогда как для многих других аминокислот (тирозина, гистидина, фенилаланина и др.) лишь двукратна. Две аминокислоты (метионин и триптофан) представлены единственными кодонами. Кодоны-синонимы почти всегда отличаются друг от друга по последнему из трех нуклеотидов, тогда как первые два совпадают. Таким образом, код аминокислоты определяется в основном первыми двумя "буквами". Вырожденность генетического кода имеет важное значение для повышения устойчивости генетической информации.

С механизмами трансляции связана еще одна особенность генетического кода: он неперекрывающийся. Кодоны транслируются всегда целиком; для кодирования невозможно использование элементов одного из них в сочетании с элементами соседнего. "Рамкой", ограничивающей транслируемый кодон и перемещающейся скачком сразу на три нуклеотида, служит антикодон тРНК, который представляет собой триплет нуклеотидов, комплементарный одному из кодонов и обусловливающий специфичность к нему. Таким образом, наблюдается линейное соответствие между последовательностью кодирующих триплетов и расположением остатков аминокислот в синтезируемом полипептиде, т.е. код имеет линейный непрерывающийся порядок считывания.

Важнейшее свойство генетического кода - его однонаправленность. Кодоны информативны только в том случае, если они считываются в одном направлении - от первого нуклеотида к последующим.

Генетический код универсален для всех живых существ. Могут быть только небольшие видовые изменения, возникшие, вероятно, при эволюции и дифференцировке клеток. Это связано с вырожденностью кода и проявляется в преимуществ. использовании разных кодонов одной и той же аминокислоты и в различиях в структуре соответствующих тРНК в разных организмах или в разных тканях одного организма.

Далеко не все гены человека известны, о функциях многих генов нет данных. Задача биоинформатики заключается в том, чтобы найти ранее неизвестные гены и описать их предположительную функцию. – В будущем ученые надеются перейти с уровня понимания того, какими генами вызывается болезнь, на уровень понимания того, как она возникает – как гены работают и как они между собой взаимодействуют. Это даст возможность приблизиться к более точному предсказанию болезней. Еще одно интересное направление касается действия лекарств. Оказывается, для каждого человека скорость этого действия, а значит, и эффективность индивидуальна. Это тоже генетически обусловленный факт. Эту скорость с определенной степенью достоверности тоже можно будет предсказать по анализу генов. Таким образом, знание генетической конституции позволит определить, что необходимо для лечения и в какой дозе. С практической точки зрения это поможет стать человеку умным и здоровым. Увеличится продолжительность и качество жизни. И это предстоит сделать нашему поколению.