Лекция 3. Структура гена: прерывистые и непрерывные кодирующие последовательности, размеры и расположение регуляторов.

1 Структура гена.

2 Прерывистые и непрерывные кодирующие последовательности.

1. Размеры и расположение регуляторов.
2. **Структура гена**. Ген- единица [наследственной](https://www.britannica.com/dictionary/hereditary) информации, занимающая фиксированное положение (локус) на [хромосоме](https://www.britannica.com/science/chromosome). Гены достигают своего эффекта, направляя синтез [белков](https://www.britannica.com/science/protein).

[У эукариот](https://www.britannica.com/science/eukaryote) (таких как [животные](https://www.britannica.com/animal/animal) , [растения](https://www.britannica.com/plant/plant) и [грибы](https://www.britannica.com/science/fungus) ) гены содержатся в [ядре](https://www.britannica.com/science/nucleus-biology) клетки. [Митохондрии](https://www.britannica.com/science/mitochondrion) (у животных) и [хлоропласты](https://www.britannica.com/science/chloroplast) (у растений) также содержат небольшие подмножества генов, отличных от генов, обнаруженных в ядре. [У прокариот](https://www.britannica.com/science/prokaryote) (организмов, не имеющих отдельного ядра, например, [бактерий](https://www.britannica.com/science/bacteria) ) гены содержатся в одной хромосоме, свободно плавающей в [цитоплазме](https://www.britannica.com/science/cytoplasm) клетки . Многие бактерии также содержат [плазмиды](https://www.britannica.com/science/plasmid) — внехромосомные генетические элементы с небольшим числом генов.

Число генов в геноме организма (полный набор хромосом) значительно варьируется между видами. Например, в то время как [геном человека](https://www.britannica.com/science/human-genome) содержит приблизительно 20 000–25 000 генов, геном бактерии Escherichia coli O157:H7 содержит ровно 5416 генов. Arabidopsis thaliana — первое [растение](https://www.britannica.com/plant/plant) , для которого была восстановлена ​​полная геномная последовательность — имеет примерно 25 500 генов; его геном является одним из самых маленьких известных растений. Среди [существующих](https://www.merriam-webster.com/dictionary/extant) независимо реплицирующихся организмов бактерия Mycoplasma genitalium имеет наименьшее количество генов, всего 517.

***Химическая структура генов.*** Гены состоят из дезоксирибонуклеиновой кислоты ([ДНК](https://www.britannica.com/science/DNA) ), за исключением некоторых [вирусов](https://www.britannica.com/science/virus) , гены которых состоят из тесно связанного [соединения,](https://www.merriam-webster.com/dictionary/compound) называемого [рибонуклеиновая кислота](https://www.britannica.com/science/RNA) ( [РНК](https://www.britannica.com/science/RNA) ). Молекула ДНК состоит из двух цепей [нуклеотиды](https://www.britannica.com/science/nucleotide) , которые обвиваются друг вокруг друга, напоминая скрученную лестницу. Стороны лестницы состоят из сахаров и фосфатов, а перекладины образованы связанными парами азотистых оснований. Эти основания — [аденин](https://www.britannica.com/science/adenine) (A), [гуанин](https://www.britannica.com/science/guanine) (G), [цитозин](https://www.britannica.com/science/cytosine) (C) и [тимин](https://www.britannica.com/science/thymine) (T). A на одной цепи связывается с T на другой (таким образом, образуя ступеньку лестницы A–T); аналогично, C на одной цепи связывается с G на другой. Если связи между основаниями разорваны, две цепи [раскручиваются](https://www.britannica.com/dictionary/unwind) , и свободные нуклеотиды внутри [клетки](https://www.britannica.com/science/cell-biology) прикрепляются к открытым основаниям теперь разделенных цепей. Свободные нуклеотиды выстраиваются вдоль каждой цепи в соответствии с правилом спаривания оснований — A связывается с T, C связывается с G.

Этот процесс приводит к созданию двух идентичных молекул ДНК из одной исходной и является методом, с помощью которого наследственная информация передается от одного поколения клеток к другому.



***Рисунок 1.***[*Ген; интрон и экзон*](https://cdn.britannica.com/96/114896-050-3F22219B/Genes-promoter-regions-production-introns-exons-gene.jpg)*Гены состоят из промоторных областей и чередующихся областей интронов (некодирующих последовательностей) и экзонов (кодирующих последовательностей) (источник: Britannica, The Editors of Encyclopaedia. "gene".*Encyclopedia Britannica*, 25 Oct. 2024,* [*https://www.britannica.com/science/gene. Accessed 26 October 2024*](https://www.britannica.com/science/gene.%20Accessed%2026%20October%202024)*).*

Ген является фундаментальной единицей наследования и конечным детерминантом всех фенотипов. ДНК нормальной человеческой клетки содержит приблизительно от 30 000 до 120 000 генов, только часть из них используется (или *экспрессируется* ) в любой конкретной клетке в любой момент времени. Например, гены, специфичные для эритроидных клеток, такие как гены гемоглобина, не экспрессируются в клетках мозга. Идентичность каждого гена, экспрессируемого в конкретной клетке в определенный момент времени, и уровень его экспрессии определяются как *транскриптом*.

Согласно «центральной догме» молекулярной биологии, ген оказывает свое действие, транскрибируя свою ДНК *в* мРНК, которая, в свою очередь, *транслируется* в белок, конечный эффектор действия гена. Таким образом, молекулярные биологи часто исследуют «экспрессию» или «активацию» гена, под которыми подразумевается процесс транскрипции ДНК в РНК или трансляции РНК в белок. Процесс транскрипции включает создание идеальной РНК-копии гена с использованием ДНК гена в качестве шаблона. Трансляция мРНК в белок — несколько более сложный процесс, поскольку структура белка гена закодирована *в* мРНК, и это структурное сообщение должно быть расшифровано во время трансляции.

***Функциональные компоненты гена.*** Каждый ген состоит из нескольких функциональных компонентов, каждый из которых участвует в различных аспектах процесса экспрессии гена ( Рисунок 2). Однако, в общем, существуют две основные функциональные единицы: область *промотора* и область *кодирования*.



***Рисунок 2.*** *Экспрессия гена. ДНК гена транскрибируется в мРНК, которая, в свою очередь, транслируется в белок. Функциональные компоненты гена схематически изображены здесь (источник: Polyak K, Meyerson M. Overview: Gene Structure. In: Kufe DW, Pollock RE, Weichselbaum RR, et al., editors. Holland-Frei Cancer Medicine. 6th edition. Hamilton (ON): BC Decker; 2003. Available from:* [*https://www.ncbi.nlm.nih.gov/books/NBK12983/*](https://www.ncbi.nlm.nih.gov/books/NBK12983/)*)*

Промоторная область контролирует, когда и в какой ткани экспрессируется ген. Например, промоторы гена глобина отвечают за их экспрессию в эритроидных клетках, а не в клетках мозга. Как достигается эта тканеспецифическая экспрессия? В ДНК промоторной области гена есть определенные структурные элементы, *нуклеотидные последовательности*, которые позволяют гену экспрессироваться только в соответствующей клетке. Это элементы в гене глобина, которые инструктируют эритроидную клетку транскрибировать мРНК глобина с этого гена. Эти структуры называются *цис* -действующими элементами, поскольку они находятся на той же молекуле ДНК, что и ген. В некоторых случаях другие специфичные для типа ткани *цис* -действующие элементы, называемые *энхансерами*, находятся на той же молекуле ДНК, но на большом расстоянии от кодирующей области гена. В соответствующей клетке *цис* -действующие элементы связывают белковые факторы, которые физически отвечают за транскрипцию гена. Эти белки называются *транс* -действующими факторами, поскольку они находятся в ядре клетки, отдельно от молекулы ДНК, несущей ген. Например, клетки мозга не будут иметь правильных *транс* -действующих факторов, которые связываются с промотором β-глобина, и, следовательно, клетки мозга не будут экспрессировать глобин. Однако у них будут *транс* -действующие факторы, которые связываются с промоторами генов, специфичными для нейронов.

Структура белка гена определяется кодирующей областью гена. Кодирующая область содержит информацию, которая направляет эритроидную клетку на сборку аминокислот в правильном порядке для создания белка β-глобина. Как определяется этот порядок аминокислот? Как подробно описано ниже, ДНК представляет собой линейный полимер, состоящий из четырех различимых субъединиц, называемых *нуклеотидами*. В кодирующей области гена линейная последовательность нуклеотидов *кодирует* аминокислотную последовательность белка. Этот генетический код находится в форме триплета, так что каждая группа из трех нуклеотидов кодирует одну аминокислоту. 64 триплета, которые могут быть образованы 4 нуклеотидами, превышают 20 отдельных аминокислот, используемых для создания белков. Это делает код вырожденным и позволяет кодировать некоторые аминокислоты несколькими различными триплетами. Теперь можно определить последовательность нуклеотидов любого гена. Транслируя код, можно вывести прогнозируемую аминокислотную последовательность для белка, кодируемого геном.

***Структурные особенности.*** Основные повторяющиеся единицы полимера ДНК — нуклеотиды (Рисунок 3). Нуклеотиды состоят из инвариантной части, пятиуглеродного дезоксирибозного сахара с фосфатной группой, и вариабельной части, основания*.* Из четырех оснований, которые появляются в нуклеотидах ДНК, два являются пуринами, аденином (A) и гуанином (G), и два являются пиримидинами, цитозином (C) и тимином (T). Нуклеотиды связаны друг с другом в полимере через свои фосфатные группы, оставляя основания свободными для взаимодействия друг с другом посредством водородных связей. Это *спаривание оснований* является специфическим, так что A взаимодействует с T, а C взаимодействует с G. ДНК обычно является двухцепочечной, то есть два линейных полимера ДНК выровнены таким образом, что основания двух цепей обращены друг к другу. Спаривание оснований делает это выравнивание специфическим, так что одна цепочка ДНК является идеально комплементарной копией другой. Эта комплементарность означает, что каждая нить ДНК несет информацию, необходимую для создания точной копии самой себя.



***Рисунок 3.*** *Структура двухцепочечной ДНК с парными основаниями (источник: : Polyak K, Meyerson M. Overview: Gene Structure. In: Kufe DW, Pollock RE, Weichselbaum RR, et al., editors. Holland-Frei Cancer Medicine. 6th edition. Hamilton (ON): BC Decker; 2003. Available from:* [*https://www.ncbi.nlm.nih.gov/books/NBK12983/*](https://www.ncbi.nlm.nih.gov/books/NBK12983/)*)*

В каждой цепи полимера ДНК замены фосфата чередуются между 5′ и 3′ атомами углерода молекул дезоксирибозы. Таким образом, в ДНК есть направленность: генетический код читается в направлении от 5′ к 3′. В двухцепочечной ДНК цепь, которая несет транслируемый код в направлении от 5′ к 3′, называется смысловой *цепью* , в то время как ее комплементарный партнер — *антисмысловой* цепью.

***Общая структура.*** У эукариот кодирующие области большинства генов не являются непрерывными. Скорее, они состоят из областей, которые транскрибируются в мРНК, экзоны*,* которые прерываются участками ДНК, которые не появляются в зрелой мРНК, *интроны*. Функции интронов точно не известны. Их сохранение в эволюции подразумевает определенную цель. Однако их общая физическая структура может быть важнее их конкретных нуклеотидных последовательностей, поскольку нуклеотидные последовательности интронов расходятся в эволюции быстрее, чем последовательности экзонов. В целом, ДНК, содержащая гены, составляет меньшинство от общей ДНК. Между генами находятся обширные участки нетранскрибированной ДНК, которые, как предполагается, играют важную структурную роль.

В ядре ДНК не присутствует в виде голой нуклеиновой кислоты. Вместо этого ДНК находится в тесной связи с рядом вспомогательных белков, таких как гистоны, и в этой форме называется *хроматином*. Хотя, многие из вспомогательных белков ДНК не имеют известной конкретной функции, они, как правило, по-видимому, участвуют в правильной упаковке ДНК. Например, двойная спираль ДНК обычно закручена сама на себя, образуя сверхспиральную структуру. Эта структура должна частично раскручиваться во время репликации и транскрипции ДНК. Некоторые из вспомогательных белков, например, топоизомеразы и гистонацетилазы, участвуют в регуляции этого процесса.

Гены определяют структуру белков, которые отвечают за фенотип, связанный с определенным геном. Хотя ядро ​​каждой человеческой клетки содержит от 30 000 до 120 000 генов, только часть из них экспрессируется в любой данной клетке в любой данный момент времени. Промотор (с энхансером или без него) — это часть гена, которая определяет, когда и где он будет экспрессироваться. Кодирующая область — это часть гена, которая диктует аминокислотную последовательность белка, кодируемого геном. ДНК — это линейный полимер нуклеотидов. Обычно нуклеотидные основания одной цепи ДНК взаимодействуют с основаниями другой цепи (A с T, C с G), образуя двухцепочечную ДНК. В ядре клетки ДНК связана со вспомогательными белками, образуя структуру, называемую хроматином.

1. **Прерывистые и непрерывные кодирующие последовательности.** Генетическая информация, закодированная в молекулах ДНК, определяет строение и функции всех живых организмов. Реализация этой информации происходит через синтез белков, который обеспечивается кодирующими последовательностями генов. В зависимости от организации этих последовательностей различают непрерывные и прерывистые гены, что отражает разные механизмы экспрессии генов в прокариотах и эукариотах. В данной лекции мы рассмотрим особенности этих типов кодирующих последовательностей, их биологическое значение и влияние на регуляцию генов.

Непрерывные кодирующие последовательности представляют собой гены, в которых информация для синтеза белка записана в виде единого, неразрывного участка ДНК. Такие гены характерны для прокариотических организмов — бактерий и архей. У них отсутствуют интроны, и вся последовательность ДНК транскрибируется в мРНК, которая затем непосредственно транслируется в белок. Это упрощает процессы транскрипции и трансляции, позволяя прокариотам быстро реагировать на изменения окружающей среды. Кроме того, у прокариот часто встречаются опероны — группы генов, связанных по функции, которые транскрибируются вместе и регулируются совместно. Это обеспечивает координированную экспрессию генов, необходимых для определенных метаболических путей.

Примером непрерывных генов служит оперон лактозы (lac-оперон) у Escherichia coli, который регулирует метаболизм лактозы. Также к ним относятся гены, кодирующие рибосомные белки, необходимые для синтеза белков в клетке. Такая организация генетического материала позволяет прокариотам экономить ресурсы и быстро приспосабливаться к изменениям условий окружающей среды.

Прерывистые кодирующие последовательности характерны для эукариотических организмов — животных, растений, грибов и протистов. В таких генах информация разделена на экзоны и интроны. Экзоны содержат кодирующую информацию и сохраняются в зрелой мРНК, тогда как интроны — некодирующие последовательности — удаляются в процессе сплайсинга. Первичный транскрипт РНК, или пре-мРНК, содержит как экзоны, так и интроны. Для образования зрелой мРНК интроны должны быть вырезаны, а экзоны — соединены вместе. Этот процесс осуществляется сплайсосомой — сложным комплексом из РНК и белков.

Одним из ключевых преимуществ прерывистых генов является возможность альтернативного сплайсинга. Это процесс, при котором экзоны могут комбинироваться в различной последовательности или некоторые из них могут быть исключены из зрелой мРНК. В результате из одного гена могут синтезироваться различные белки, что значительно увеличивает функциональное разнообразие организма. Альтернативный сплайсинг играет важную роль в развитии многоклеточных организмов и позволяет регулировать экспрессию генов в различных тканях и на разных этапах развития.

Биологическое значение прерывистых и непрерывных кодирующих последовательностей связано с эффективностью экспрессии генов и адаптационными возможностями организмов. У прокариот непрерывные гены обеспечивают быстрый и энергоэффективный синтез белков, что важно для быстрого роста и размножения в изменяющихся условиях среды. Отсутствие интронов снижает энергетические затраты на сплайсинг и позволяет транскрипции и трансляции происходить одновременно.

У эукариот прерывистые гены предоставляют возможность более сложной регуляции генов. Интроны могут содержать регуляторные элементы, влияющие на экспрессию генов, а альтернативный сплайсинг позволяет получать различные белки из одного гена, что способствует функциональному разнообразию и адаптации к разным условиям. Кроме того, интроны могут играть роль буфера, защищая кодирующие последовательности от мутаций и способствуя эволюционной гибкости генома.

Эволюционные аспекты организации генов также представляют интерес. Существуют две основные теории происхождения интронов: "интроны-раньше" и "интроны-позже". Согласно первой, интроны присутствовали в древних генах и были утрачены прокариотами в процессе эволюции для повышения эффективности экспрессии генов. Вторая теория предполагает, что интроны появились уже после разделения прокариот и эукариот, обеспечивая усложнение регуляторных механизмов и функциональное разнообразие у последних.

Знания о прерывистых и непрерывных кодирующих последовательностях имеют важное практическое значение. В генетической инженерии понимание структуры генов необходимо для их успешного клонирования и экспрессии в различных организмах. При создании трансгенных организмов важно учитывать особенности сплайсинга и регуляции генов, чтобы обеспечить корректную экспрессию введенных генов.

В медицинской генетике мутации в интронах или ошибки в процессе сплайсинга могут приводить к различным заболеваниям. Диагностика таких патологий требует глубокого понимания структуры генов и механизмов их экспрессии. Генная терапия, направленная на исправление генетических дефектов, также основывается на точном знании структуры и функции генов.

В биотехнологии и фармакологии разработки лекарственных средств, влияющих на процессы сплайсинга, открывают новые возможности в лечении рака и вирусных инфекций. Производство рекомбинантных белков для медицинских и промышленных целей требует учета особенностей кодирующих последовательностей и процессов экспрессии генов в различных системах.

В заключение, прерывистые и непрерывные кодирующие последовательности отражают фундаментальные различия в организации генетической информации между прокариотами и эукариотами. Непрерывные гены обеспечивают прокариотам быстрый и эффективный синтез белков, тогда как прерывистые гены эукариот позволяют сложную регуляцию экспрессии генов и повышенное функциональное разнообразие через альтернативный сплайсинг. Понимание этих механизмов имеет ключевое значение для биологии, медицины и биотехнологии, открывая новые перспективы в исследовании живых организмов и разработке инновационных подходов к лечению заболеваний и созданию биологических продуктов.

1. **Размеры и расположение регуляторов.** Генетическая информация, закодированная в ДНК, управляется не только белок-кодирующими последовательностями, но и регуляторными элементами, которые контролируют экспрессию генов. Эти регуляторы играют ключевую роль в пространственно-временной координации генетической активности, обеспечивая правильное функционирование клеток и всего организма. В этой лекции мы рассмотрим размеры и расположение регуляторных элементов в геноме, опираясь на современные научные исследования.

Регуляторные элементы ДНК включают промоторы, энхансеры, сайленсеры, инсуляторы и другие последовательности, взаимодействующие с транскрипционными факторами и регулирующими белками. Размеры этих элементов могут существенно различаться, варьируя от нескольких десятков до нескольких тысяч нуклеотидов. Промоторы, например, являются относительно короткими последовательностями, обычно состоящими из 100–200 пар оснований у прокариот и несколько больших размеров у эукариот.

Промоторы располагаются непосредственно перед стартовой точкой транскрипции гена и служат основным местом связывания РНК-полимеразы и основных транскрипционных факторов. У эукариотических организмов промоторы часто содержат такие элементы, как TATA-бокс, расположенный примерно за 25–35 пар оснований до места начала транскрипции. Этот элемент играет важную роль в правильном позиционировании РНК-полимеразы II и инициации транскрипции.

Энхансеры и сайленсеры являются дистальными регуляторными элементами, которые могут находиться на значительном расстоянии от регулируемого гена — как вверх, так и вниз по течению ДНК, а иногда даже внутри интронов других генов. Энхансеры способны повышать уровень транскрипции гена, в то время как сайленсеры подавляют его. Размеры энхансеров обычно составляют от 200 до 1000 пар оснований, но могут быть и более протяженными. Благодаря петлеобразованию ДНК, энхансеры и сайленсеры могут физически взаимодействовать с промоторами, несмотря на значительное линейное расстояние между ними.

Инсуляторы — это особые регуляторные элементы, которые ограничивают взаимодействие между энхансерами и промоторами, предотвращая нежелательную активацию или репрессию соседних генов. Они обычно имеют размер от 500 до 2000 пар оснований и содержат сайты связывания специфических белков, таких как CTCF (CCCTC-binding factor) у эукариот. Расположение инсуляторов между энхансером и промотором может блокировать передачу регуляторных сигналов, обеспечивая точную пространственную регуляцию генов.

Трехмерная организация генома играет критическую роль в функционировании регуляторных элементов. Современные методы изучения хроматиновой структуры, такие как Hi-C и ChIA-PET, выявили существование топологически ассоциированных доменов (TADs), внутри которых происходят активные взаимодействия между регуляторными элементами и их целевыми промоторами. Эти домены могут охватывать от десятков до сотен килобаз ДНК и обеспечивают функциональную изоляцию генетических элементов, влияя на их расположение и эффективность взаимодействия.

Эпигенетические модификации, такие как метилирование ДНК и посттрансляционные модификации гистонов, также влияют на размеры и функциональную доступность регуляторных элементов. Метилирование цитозинов в промоторных областях обычно ассоциируется с подавлением генной экспрессии. Такая эпигенетическая регуляция может изменять структуру хроматина, делая регуляторные элементы более или менее доступными для связывания транскрипционных факторов.

Генетические вариации в размерах и расположении регуляторных элементов могут иметь значительные биологические последствия. Множество исследований ассоциации генома (GWAS) показали, что многие полиморфизмы, связанные с риском развития различных заболеваний, находятся именно в некодирующих регуляторных областях. Например, вариации в энхансерах гена MYC связаны с повышенным риском развития некоторых видов рака, что подчеркивает важность точного расположения регуляторов.

Технологии редактирования генома, такие как CRISPR/Cas9, предоставили новые возможности для изучения функции регуляторных элементов путем их целенаправленного удаления или модификации. Изменение размеров энхансеров или перемещение инсуляторов может приводить к изменению экспрессии генов и помочь в выявлении их конкретных функций. Такие подходы также открывают перспективы для разработки генотерапевтических методов лечения наследственных и онкологических заболеваний.

Важность размеров и расположения регуляторных элементов становится особенно очевидной при изучении процессов развития и дифференцировки клеток. Точные пространственно-временные модели экспрессии генов обеспечиваются сложной сетью взаимодействий между различными регуляторными элементами. Нарушения в этой сети могут приводить к развитию патологий и аномалий развития.

В заключение, размеры и расположение регуляторных элементов являются фундаментальными параметрами, определяющими сложность и точность генетической регуляции. Понимание этих аспектов имеет критическое значение для молекулярной биологии, генетики, медицины и биотехнологии. Современные научные исследования продолжают раскрывать сложность регуляторных сетей и их влияние на здоровье и заболевания человека, что открывает новые горизонты для диагностических и терапевтических инноваций.



***Рисунок 2.*** *Представление регуляторных регионов в геноме животных*

**Вопросы для самоконтроля:**

1. **Объясните различия между прерывистыми и непрерывными кодирующими последовательностями. Какие организмы характеризуются каждым типом, и как это влияет на процессы транскрипции и трансляции?**
2. **Что такое альтернативный сплайсинг, и какое биологическое значение он имеет у эукариотических организмов? Приведите примеры, иллюстрирующие его роль в повышении протеомного разнообразия.**
3. **Опишите роль регуляторных элементов ДНК, таких как промоторы, энхансеры и инсуляторы. Как их размеры и расположение влияют на экспрессию генов?**
4. **Как трехмерная организация генома и формирование топологически ассоциированных доменов (TADs) влияют на взаимодействие между регуляторными элементами и генами? Приведите примеры того, как это может влиять на генную экспрессию.**
5. **Обсудите эволюционные теории происхождения интронов: "интроны-раньше" и "интроны-позже". Какие научные данные поддерживают каждую из этих гипотез?**
6. **Каким образом изменения в размерах или расположении регуляторных элементов могут привести к заболеваниям? Приведите примеры генетических вариаций в регуляторных областях, связанных с патологией у человека.**