

МУСОРНАЯ ДНК. ENCODE. ЗНАЧЕНИЕ.
ИСПОЛЬЗОВАНИЕ НЕКОДИРУЮЩЕЙ
ДНК. НЕКОДИРУЮЩАЯ
ФУНКЦИОНАЛЬНАЯ РНК

01-705 гр

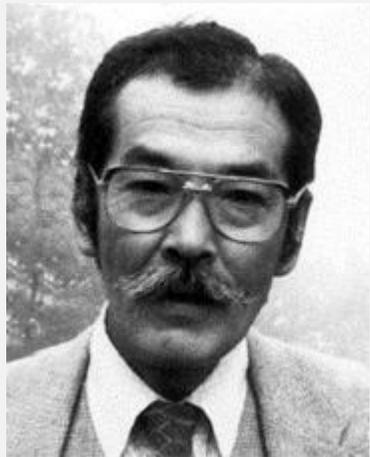
Свитко С., Муллакаева А.

НЕКОДИРУЮЩАЯ ДНК (МУСОРНАЯ ДНК)
(АНГЛ. NON-CODING DNA) — ЧАСТИ
ГЕНОМНОЙ ДНК ОРГАНИЗМОВ, КОТОРЫЕ НЕ
КОДИРУЮТ ПОСЛЕДОВАТЕЛЬНОСТИ БЕЛКОВ.



ИСТОРИЯ ТЕРМИНА

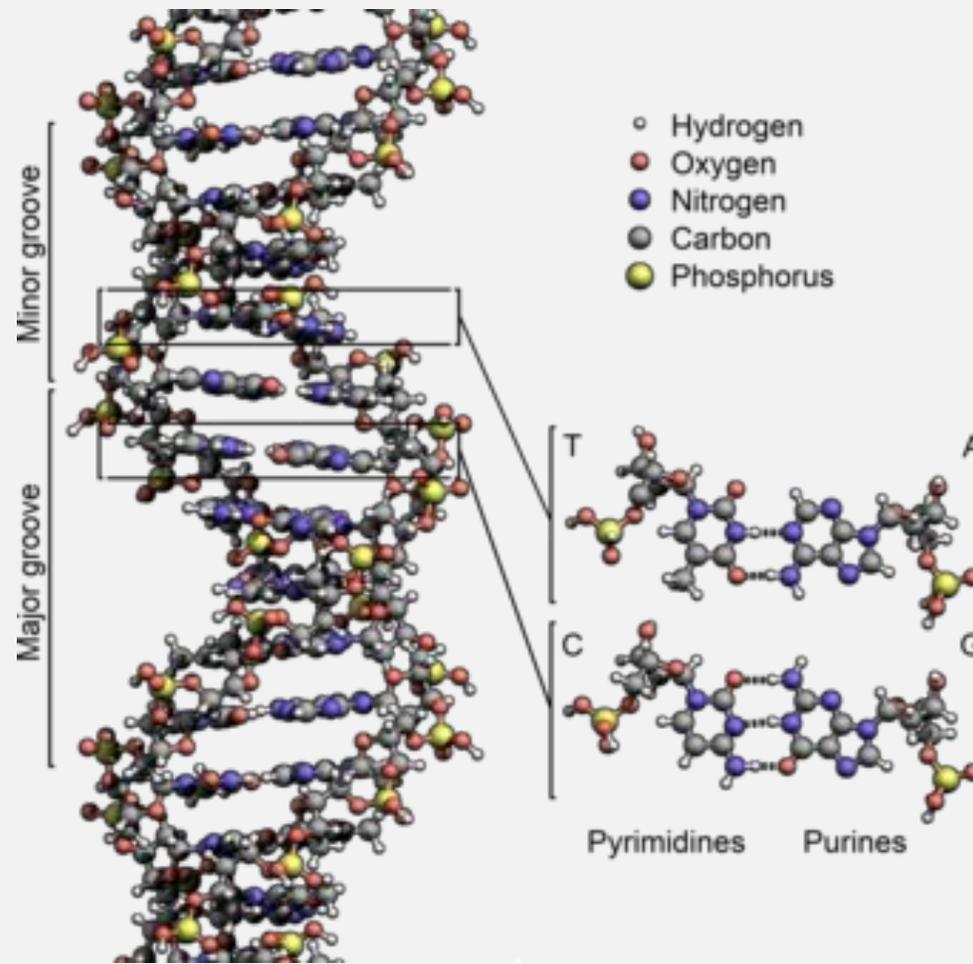
- Термин был впервые предложен японским ученым-генетиком Сусуму Оно в 1972 году.



- Он заметил, что генетический груз нейтральных мутаций находится на верхнем пределе значений для функционирующих локусов, которые могли быть ожидаемыми исходя из типичной частоты мутаций. Сусуму предсказал, что геномы млекопитающих не могут содержать более чем *30 000* локусов из-за давления естественного отбора, так как «стоимость» мутационной нагрузки вызвала бы неизбежное снижение приспособленности, и в конечном счете вымирание. Этот прогноз остается верным, геном человека содержит приблизительно *20 000* генов. Другим подтверждением теории Оно служит наблюдение, что даже близкородственные виды могут иметь очень разные (отличающиеся на порядок) по размеру геномы.

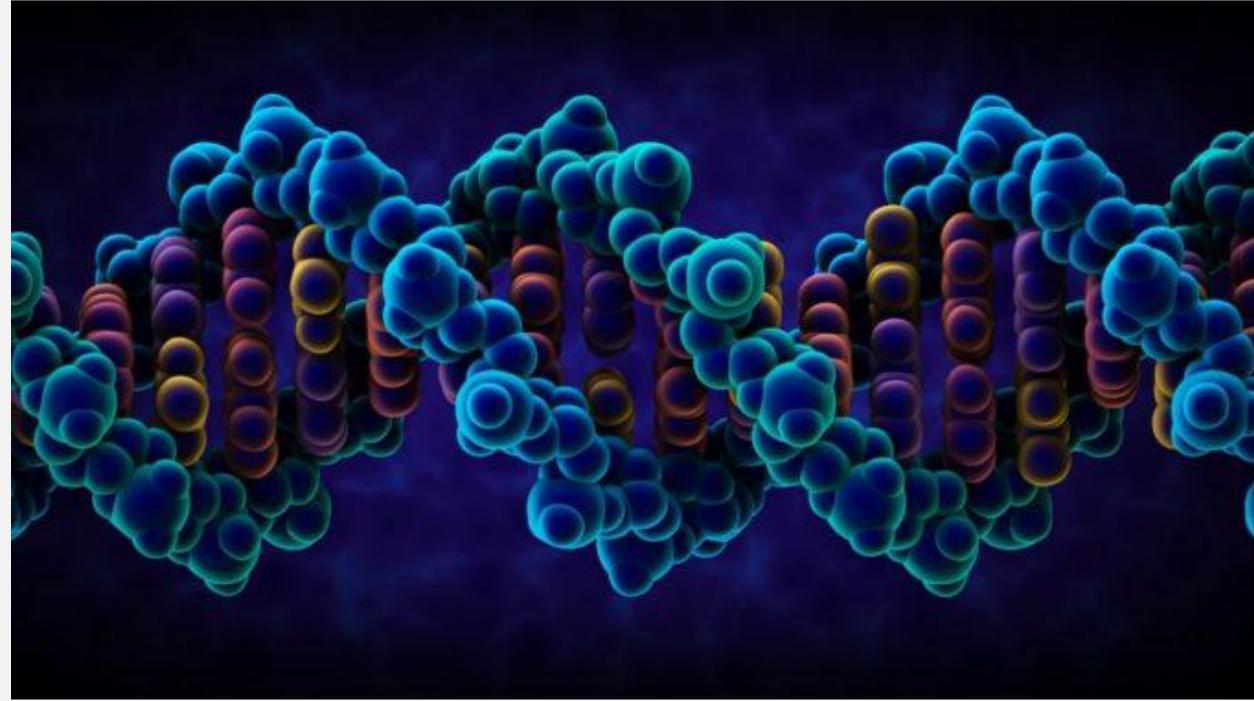
НЕКОДИРУЮЩАЯ ДНК

- Единой концепции эволюционной роли и возникновения «мусорной» ДНК пока нет, однако существует мнение о том, что некодирующая ДНК эукариот представляет собой остатки некодирующих последовательностей ДНК, возникших при становлении жизни. Прокариоты были вынуждены сократить размер своих геномов для того, чтобы уменьшить количество ДНК, в которой могут происходить мутации, в то время как эукариоты «пошли по пути» диплоидности и регулярного полового процесса.



НЕКОДИРУЮЩАЯ ДНК

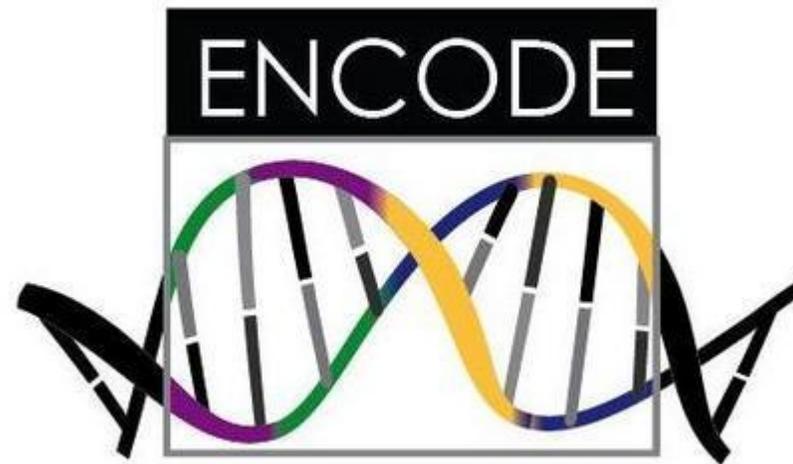
- Ранее считалось, что около 95 % последовательностей ДНК генома человека можно отнести к мусорной ДНК. Однако в 2012 году в публикациях проекта «Энциклопедия элементов ДНК» (ENCODE) было показано, что доля мусорной ДНК сильно завышена, и до 80 % генома имеет биохимические функции



НЕКОДИРУЮЩАЯ ДНК

- Исследования ENCODE показывают, что некоторые последовательности «мусорной ДНК», скорее всего, должны иметь неизвестную нам функциональную активность и что процесс *экзаптации* (смены функциональности) фрагментов первоначально нефункциональной ДНК было обычным явлением на протяжении всей эволюции

National Human Genome Research Institute



ENCODE

- **Энциклопедия элементов ДНК (англ. The Encyclopedia of DNA Elements, ENCODE)** — международный исследовательский проект, созданный в сентябре 2003 года. Организован и финансируется американским Национальным институтом исследований генома человека (англ. National Human Genome Research Institute, NHGRI). Задуманный как продолжение проекта «Геном человека», **ENCODE** ставит целью проведение полного анализа функциональных элементов генома человека. Все результаты, получаемые в ходе реализации проекта, публикуются в общедоступных базах данных.
- 5 сентября 2012 года первые результаты проекта были опубликованы в виде 30 взаимосвязанных публикаций на сайтах журналов «Nature», «Genome Biology[en]» и «Genome Research[en]». Эти публикации показывают, что по крайней мере 80 % генома человека является биологически активным, до этого господствовало представление, что большая часть ДНК является «мусорной». Однако такие поспешные выводы подвергаются критике со стороны многих учёных, которые указывают на отсутствие необходимых доказательств функциональности этих элементов

ENCODE

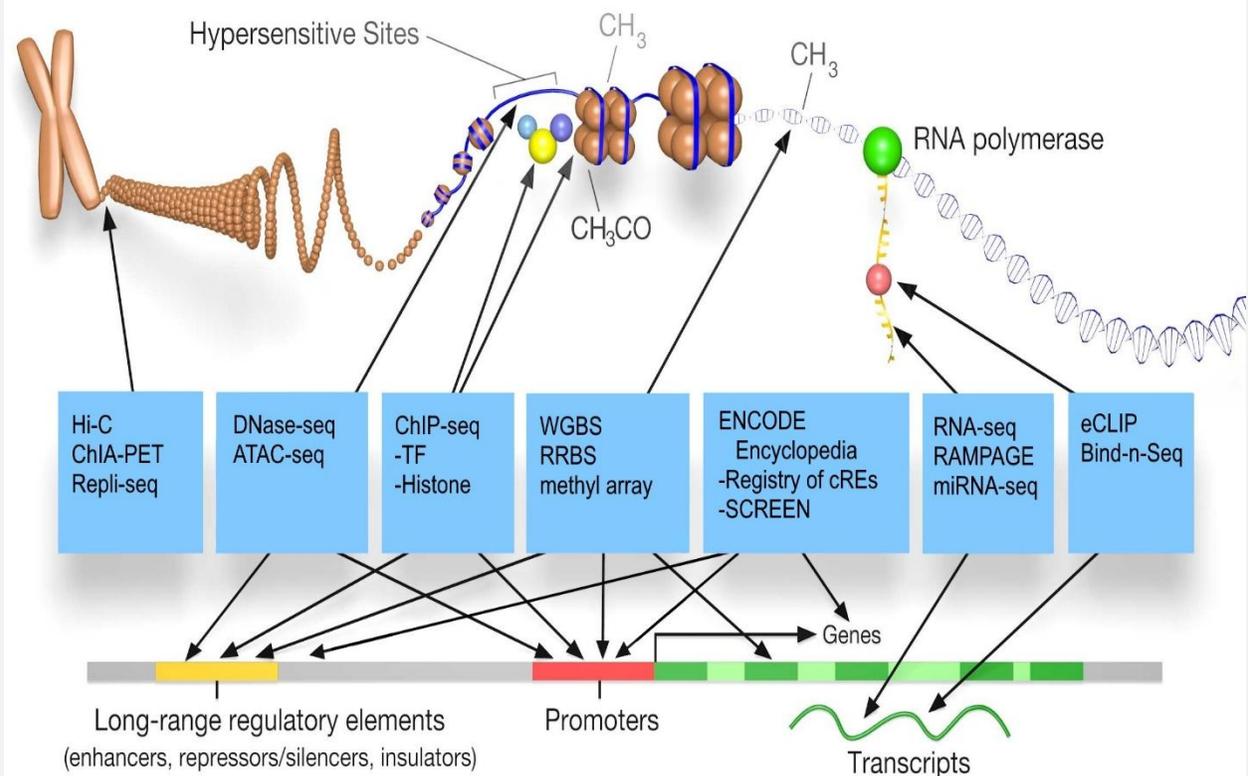
- Геном человека по приблизительным оценкам содержит 20 000 белок-кодирующих генов, и на их долю приходится всего около 1,5 % ДНК генома человека. Первостепенная задача проекта ENCODE — определить функцию остальной части генома, большая часть которой традиционно рассматривалась как «мусор» (например, ДНК, которая не транскрибируется).
- Примерно 90 % однонуклеотидных полиморфизмов в геноме человека (которые связаны с различными заболеваниями) найдено за пределами белок-кодирующих областей.

Однонуклеотидный полиморфизм — отличия последовательности ДНК размером в один нуклеотид в геноме представителей одного вида или между гомологичными участками гомологичных хромосом.

- Активность и экспрессия белок-кодирующих генов может регулироваться регуломом — различными элементами ДНК, такими как промотор, регуляторные последовательности и участки хроматина, а также модификации гистонов. Считается, что изменения в регуляторных областях могут нарушать экспрессию белка и функционирование клетки и, таким образом, приводить к заболеваниям. Определяя расположение регуляторных элементов и их влияние на транскрипцию, можно выяснить связь между изменением уровней экспрессии конкретных генов и развитием заболеваний.

ENCODE

- ENCODE задуман как всеобъемлющий ресурс, который позволит научному сообществу лучше понять, как геном может влиять на здоровье человека, и простимулирует развитие новых методов профилактики и лечения заболеваний.
- На сегодняшний день проект помогает в обнаружении новых регуляторных элементов ДНК, позволяя по-новому понять организацию и регуляцию наших генов и генома, а также того, как изменения в последовательности ДНК могут влиять на развитие заболеваний. Один из главных результатов проекта — описание того, что для 80 % человеческого генома на данный момент показана связь хотя бы с одной биохимической функцией. Большинство этой некодирующей ДНК участвует в регуляции экспрессии кодирующих генов. Кроме того, экспрессия каждого кодирующего гена контролируется множеством регуляторных участков, расположенных как вблизи, так и на расстоянии от гена. Эти результаты демонстрируют, что генная регуляция оказывается намного сложнее, чем считалась ранее.



ENCODE

- Проект **ENCODE** реализуется в три этапа: *начальная фаза, фаза развития технологии и продуктивная фаза.*
- В течение начальной фазы ENCODE оценивал стратегии для определения различных типов элементов генома. Цель начального этапа состояла в определении набора процедур, которые в совокупности позволили бы точно и наиболее подробно охарактеризовать большие регионы генома человека, с учетом экономической рентабельности и высокой эффективности процесса. Начальная фаза должна была выявить пробелы в наборе инструментов для определения функциональных последовательностей, а также показать, если какие-то из используемых методов окажутся неэффективными или непригодными для масштабного применения.
- В сентябре 2007 года началось финансирование продуктивной фазы проекта ENCODE. На этом этапе целью был анализ генома целиком и проведение "дополнительных исследований в промышленных условиях».
- Как и в начальной фазе, работа продуктивной фазы была организована в виде открытого консорциума. В 2007 году, когда начальная стадия была закончена, проект нарастил мощности в значительной степени благодаря секвенированию нового поколения. Было обработано действительно много данных, исследователи получили около 15 терабайт сырой информации.
- К 2010 году проектом ENCODE было получено *более 1000 полногеномных наборов данных.* Взятые вместе эти данные демонстрируют, какие участки транскрибируются, какие участки, по-видимому, контролируют экспрессию генов, используемых в клетках определённого типа, и какие участки взаимодействуют с большим набором белков.

НЕКОДИРУЮЩАЯ ДНК

- Помимо термина «некодирующая ДНК» существует также альтернативное название «мусорная» ДНК. Однако оно не совсем верно, так как в «некодирующей» ДНК присутствуют транспозоны (кодирующие белки, функция которых пока не установлена), а также некоторые регуляторные элементы.
- По одной из версий, некодирующая ДНК, по крайней мере частично, используется при производстве различных видов РНК, а именно тРНК, рРНК, микроРНК, малые ядерные РНК, малые ядрышковые РНК. Все эти РНК участвуют в критически важных процессах жизнедеятельности клеток и даже многоклеточных организмов



ВИДЫ ПОСЛЕДОВАТЕЛЬНОСТЕЙ НЕКОДИРУЮЩЕЙ ДНК

- **Некодирующая функциональная РНК**

Некодирующие РНК — это функциональные молекулы РНК которые не транслируются в белки. Примеры некодирующих РНК включают рРНК, тРНК, рiРНК и микроРНК.

МикроРНК, предположительно, контролируют трансляционную активность приблизительно 30 % всех белок-кодирующих генов в млекопитающих и могут быть жизненно важными составляющими при развитии или лечении различных заболеваний, включая рак, сердечно-сосудистые заболевания, и иммунного ответа на инфекцию

- **Цис- и Транс-регуляторные элементы**

Цис-регуляторные элементы — это последовательности контролирующие транскрипцию близлежащего гена. Цис-элементы могут быть расположены в нетранслируемой области или внутри интронов. Транс-регуляторные элементы контролируют транскрипцию генов на дальних расстояниях.



ВИДЫ ПОСЛЕДОВАТЕЛЬНОСТЕЙ НЕКОДИРУЮЩЕЙ ДНК

- **Интроны**

Интроны — это некодирующие участки гена, транскрибируемые в последовательности предшественников мРНК (пре-мРНК), но полностью удаляемые при сплайсинге в течение процесса созревания матричной РНК. Много интронов представляют собой мобильные генетические элементы



- **Псевдогены**

Псевдогены — это последовательности ДНК, сходные с обычными генами, которые утратили их способность кодировать белок или больше не экспрессирующиеся в клетке. Псевдогены возникают при ретротранспозиции или дупликации функциональных генов, и становятся неработающими «ископаемыми генами» вследствие мутаций препятствующих транскрипции гена, а также мутаций внутри региона промотора, или полностью меняющих трансляцию гена, такие как возникновение стоп-кодона или сдвиг рамки считывания

В то время как закон необратимости эволюции говорит о том, что потеря функции псевдогенами должна быть постоянна, молчащие гены могут на самом деле сохранять функцию на протяжении нескольких миллионов лет и могут «реактивироваться» восстановив белок-кодирующую последовательность и значительное число бывших псевдогенов активно транскрибируется. Так как псевдогены могут меняться, как предполагается, без эволюционных ограничений, они могут служить рабочей моделью типичных и частых различных спонтанных генетических мутаций

ВИДЫ ПОСЛЕДОВАТЕЛЬНОСТЕЙ НЕКОДИРУЮЩЕЙ ДНК

- **Повторы, транспозоны и вирусные элементы**

Транспозоны и ретротранспозоны — мобильные генетические элементы. Ретротранспозонные повторяющиеся последовательности, включают длинные диспергированные повторы (LINEs) и короткие диспергированные повторы (SINEs), составляют большую часть геномной последовательности во многих видах. Alu-повторы, классифицируемые как короткие диспергированные повторы, наиболее распространенные подвижный элемент в геноме человека. Найдены некоторые примеры того, что SINEs оказывают влияние на транскрипционный контроль некоторых белок-кодирующих генов.

Последовательности эндогенных ретровирусов являются продуктами обратной транскрипции геномов ретровирусов и их встройке в геном клеток зародышевой линии. Мутации внутри этих обратно-транскрибируемых последовательностей могут инактивировать геном вируса.

Более 8 % человеческого генома ведет свое начало от (в основном уже распавшихся) последовательностей эндогенных ретровирусов, из них свыше 42 % узнаваемо произошли от ретротранспозонов, в то время как другие 3 % могут быть идентифицированы как остатки ДНК транспозоны. Большую часть из оставшейся половины генома которая не имеет на данный момент ясного происхождения, как предполагается, ведет свое происхождение от подвижных элементов, бывших активными очень много лет назад (> 200 миллионов лет) но случайные мутации сделали их неузнаваемыми.

- **Теломеры**

Теломеры — это области повторяющейся ДНК на концах хромосом, которые обеспечивают их защиту от укорачивания в течение репликации ДНК.



ЗНАЧЕНИЕ НЕКОДИРУЮЩЕЙ ДНК

- Существует мнение, что наличие большого количества некодирующей ДНК **стабилизировало геном** в плане мутаций (снизилась частота «попадания» мутации на действующий ген). Это явилось условием для возникновения многоклеточных организмов.
- Многие некодирующие последовательности ДНК имеют важные биологические функции, о чём свидетельствуют исследования сравнительной геномики (англ. comparative genomics) в которых сообщается о некоторых регионах некодирующей ДНК, которые высококонсервативны (англ. Conserved non-coding sequence), иногда в масштабе времени, составляющем сотни миллионов лет, что подразумевает, что эти некодирующие области находятся под сильным эволюционным давлением и позитивным отбором. Например, в геномах человека и мыши, которые разошлись от общего предка 65-75 миллионов лет назад, белок-кодирующие последовательности ДНК составляют лишь около 20 % консервативной ДНК, а оставшиеся 80 % консервативной ДНК представлены в некодирующих областях. Сцепленное наследование часто выявляет области хромосом, связанные с заболеванием, не имеющих признаков функциональных вариантов кодирующих генов внутри региона, что указывает на то, что **варианты последовательностей, вызывающие заболевание, лежат в некодирующей ДНК**.
- Также показано, что генетический полиморфизм некодирующих последовательностей играет роль в **восприимчивости к инфекционным болезням**, таких как гепатит С. Кроме того, показано, что генетический полиморфизм некодирующих последовательностей способствует восприимчивости к саркоме Юинга — высокоагрессивному детскому костному раку.
- Некоторые специфические последовательности некодирующей ДНК могут быть особенно важными для **поддержания структуры хромосом**, функционирования центромеры и узнавания гомологичных хромосом в мейозе.

ЗНАЧЕНИЕ НЕКОДИРУЮЩЕЙ ДНК

• 1. Защита генома

Некодирующая ДНК отделяет длинными промежутками гены друг от друга, так что мутация в одном гене или в участке хромосомы, не приводит к «мутации сдвига рамки считывания» на всём протяжении хромосомы. Когда сложность генома является относительно высокой, подобно геному человека, не только отдельные гены, но также и отдельные части гена разделены некодирующими участками - интронами, защищающими всю кодирующую последовательность гена, минимизируя изменения, вызванные мутацией.

Было высказано предположение, что некодирующая ДНК может снижать вероятность повреждения гена в течение кроссинговера хромосом

2. Генетические переключатели

Некоторые последовательности некодирующей ДНК выступают в качестве генетических «переключателей» которые определяют где и когда будут экспрессироваться гены. Например, молекула длинной некодирующей РНК (lncRNA), как было показано, помогает в предотвращении развития рака груди, предотвращая «залипание» генетического переключателя

• 3. Регуляция экспрессии генов

Некоторые некодирующие ДНК последовательности определяют уровень экспрессии различных генов

4. Сайты связывания транскрипционных факторов

Некоторые некодирующие последовательности ДНК определяют место связывания факторов транскрипции. Факторы транскрипции — это белки, которые связываются со специфическими некодирующими последовательностями ДНК, тем самым управляя переносом (или транскрипцией) генетической информации от ДНК к мРНК. Факторы транскрипции действуют в совершенно разных местах генома у разных людей.



ИСПОЛЬЗОВАНИЕ НЕКОДИРУЮЩЕЙ ДНК

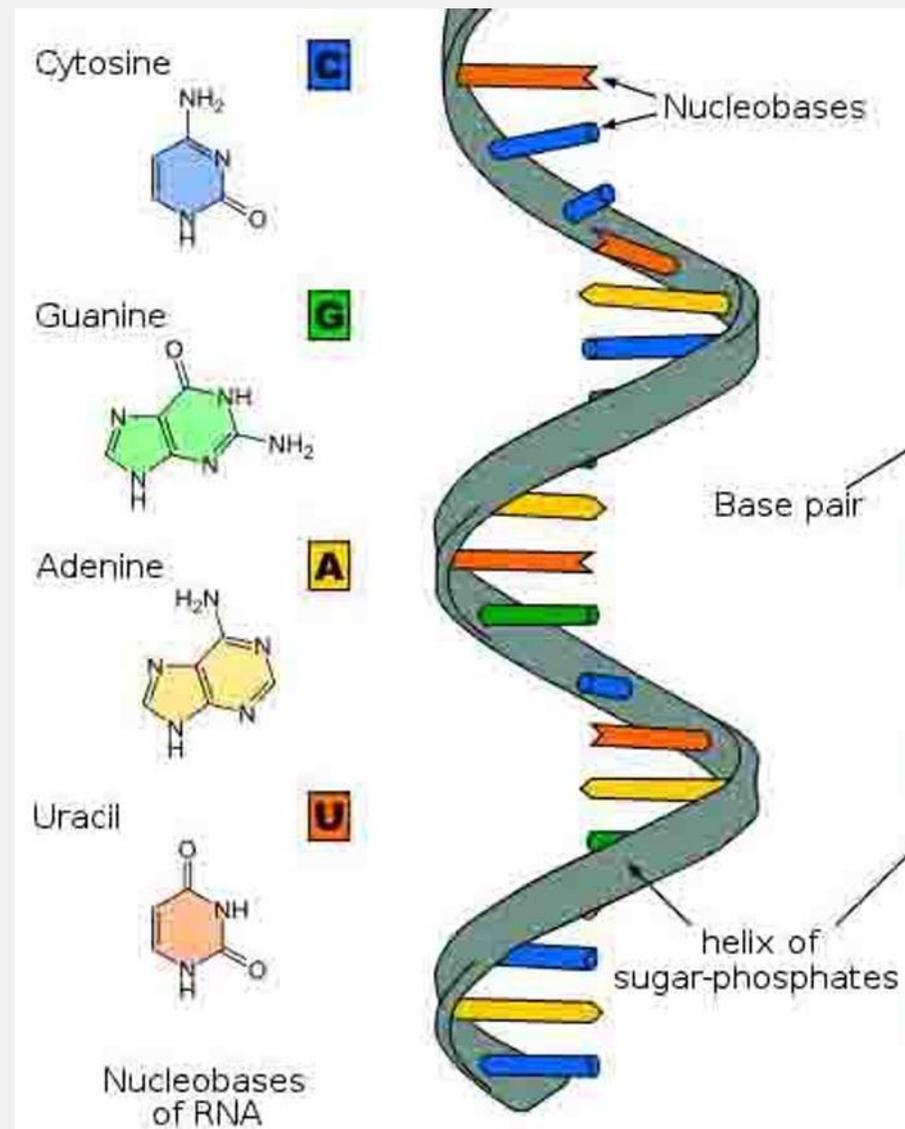
- **В эволюционной теории**

Общие последовательности некодирующей ДНК являются главным свидетельством происхождения от общего предка.

Последовательности псевдогенов, по всей видимости, накапливают мутации с большей скоростью, чем кодирующие последовательности, в связи с потерей селективного давления естественного отбора. Это позволяет создавать мутантные аллели, которые обладают новыми функциями и которые могут быть подхвачены естественным отбором; таким образом, *псевдогены могут служить в качестве материала для эволюции и могут рассматриваться как «протогены»*

НЕКОДИРУЮЩИЕ РНК

- **Некодирующие РНК (non-coding RNA, ncRNA)** — молекулы РНК, которые не транслируются в белки. Ранее использовавшийся синоним — «малые РНК» (smRNA, small RNA) — вышел из употребления, так как некоторые некодирующие РНК могут быть очень большими, например, Xist.
- Последовательность ДНК, на которой транскрибируются некодирующие РНК, часто называют **РНК-геном**.



НЕКОДИРУЮЩИЕ РНК

К некодирующим РНК относят:

- транспортные РНК (тРНК),
 - рибосомные РНК (рРНК),
 - малые ядерные РНК (мяРНК, snRNA),
 - малые ядрышковые РНК (мякРНК, snoRNA),
 - антисмысловые РНК (aRNA),
 - микроРНК (miRNA),
 - малые интерферирующие РНК (siRNA),
 - piРНК (piwiRNA, piRNA),
 - длинные некодирующие РНК (lncRNA) — Xist, Evf, Air, CTN, PINK, TUG1.
- Последние транскриптомные технологии (секвенирование РНК) и методы ДНК-микрочипов выявили более 30000 длинных некодирующих РНК (англ. long ncRNA). Предполагается, что они «руководят» избирательной посадкой репрессорных комплексов polycomb на определённые участки хроматина, а также направляют ДНК-метилтрансферазы и, возможно, регулируют избирательное метилирование ДНК. Более того, как оказалось, составляющие значительную часть (~20 % от общего количества длинных некодирующих РНК в геномах млекопитающих), так называемые дивергентные дднРНК, которые транскрибируются в противоположном направлении к соседним белок-кодирующим генам, регулируют транскрипцию этих соседних смежных генов

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

- 1. Патрушев Л. И. Экспрессия генов. — М.: Наука, 2000
- 2. <http://www.nature.com/encode/#/threads>
- 3. <http://www.zbi.ee/fungal-genomesize/index.php>
- 4. *Castillo-Davis C. I.* The evolution of noncoding DNA 2005
- 5. *Gregory, T.R.* Genome size evolution in animals // The Evolution of the Genome / T.R. Gregory (ed.). — San Diego : Elsevier, 2005.