

На правах рукописи

УСМАНОВА
Надежда Маратовна

**РОЛЬ РЕТРОПОЗОНОВ ALU-СЕМЕЙСТВА В РЕГУЛЯЦИИ ГЕННОЙ
ЭКСПРЕССИИ В КЛЕТКАХ ЛИНИЙ НЕК293 И А549 ЧЕЛОВЕКА**

03.00.25 – гистология, цитология, клеточная биология

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата биологических наук

Санкт-Петербург
2008

Работа выполнена в Институте цитологии РАН, Санкт-Петербург

Научный руководитель:

кандидат биологических наук, доцент
Кзаков Василий Иванович
Институт цитологии РАН,
Санкт-Петербург

Официальные оппоненты:

доктор биологических наук, профессор
Воробьев Владимир Иосифович

доктор биологических наук, профессор
Горбунова Виктория Николаевна

Ведущая организация:

ВНИИ Генетики и разведения
сельскохозяйственных животных,
Санкт-Петербург – Пушкин

Защита состоится 25 апреля 2008 года в 13 часов на заседании диссертационного совета Д.002.230.01 при Институте цитологии РАН по адресу: 194064, Санкт-Петербург, Тихорецкий пр., д.4.
e-mail: cellbio@mail.cytspb.rssi.ru

С диссертацией можно ознакомиться в библиотеке Института цитологии РАН

Автореферат разослан 21 марта 2008 г.

Ученый секретарь диссертационного совета
кандидат биологических наук

Е.В. Каминская

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность исследования. Секвенирование генома человека позволило установить, что ~55% нуклеотидной последовательности составляют различные повторяющиеся элементы [1]. Изначально эти последовательности относили к «эгоистической ДНК», представляющей собой нефункциональную часть генома. В последние годы накапливаются доказательства того, что те или иные повторяющиеся элементы генома выполняют различные функции, как на генном, так и на геномном уровнях [2, 3]. Ретропозоны Alu семейства относятся к классу SINE повторов (короткие диспергированные нуклеотидные элементы) и представлены более 10^6 копий в геноме человека. В настоящее время имеются данные, указывающие на то, что Alu повторы могут участвовать в регуляции генной экспрессии на нескольких уровнях [4, 5]. Например, Alu повторы содержат в своем составе множественные CpG динуклеотиды, являющиеся сайтами метилирования в геноме. Посредством метилирования осуществляется не только регуляция генной экспрессии, но и контроль эмбрионального развития, а также подавление амплификации мобильных элементов генома [6]. Многие исследования показали наличие в Alu повторах сайтов связывания транскрипционных факторов. Экспериментально доказано связывание некоторых из этих транскрипционных факторов, например, Yin Yang 1 (YY1), Estrogen receptor α (ER α), Stimulating protein 1 (Sp1), с консенсусной последовательностью Alu [7]. Инсерции Alu повторов в экзоны белок-кодирующих генов, в области экзон-интронных границ, а также незаконная гомологичная рекомбинация между различными копиями Alu могут приводить к возникновению наследственных заболеваний [4]. Показано, что Alu РНК стимулирует трансляцию всех репортерных мРНК в бесклеточной системе трансляции [8]. Регуляция стабильности транскрипта может осуществляться за счет связывания Alu повтора, локализованного в 3'-нетранслируемой области, с SRP9/14 субъединицей сигнал-распознающей частицы [9].

Регуляция генной экспрессии является важным этапом в поддержании клеточного гомеостаза. Она лежит в основе клеточной пролиферации и дифференцировки в ходе эмбрионального развития. Изменение экспрессии генов в ответ на внеклеточные стимулы обуславливает адаптацию организмов к условиям окружающей среды [10].

В свете вышеизложенного, изучение роли Alu повторов в регуляции генной экспрессии представляется актуальной задачей. Такие исследования позволят не только детализировать механизмы регуляции, но и сделать выводы касательно функций повторяющихся элементов генома.

Цели и задачи исследования. Цель данной работы заключалась в исследовании регуляции экспрессии генов посредством Alu повторов на уровне транскрипции и на уровне формирования факультативного гетерохроматина. Для достижения данной цели были поставлены следующие задачи:

- 1) При помощи методов компьютерного анализа выбрать гены для изучения роли Alu повторов, локализованных в промоторах, в регуляции экспрессии генов;
- 2) Экспериментально показать сайленсерную или энхансерную функцию Alu повторов, расположенных в промоторах выбранных генов, в постоянных клеточных линиях человека;
- 3) На примере AluYa5 повтора, содержащегося в 16-м интроне гена ангиотензинпревращающего фермента (АПФ), показать функциональную роль Alu повторов, локализованных в интронах;
- 4) На примере инактивации X хромосомы показать возможность функционирования Alu повторов и SINE повторов мыши в качестве вспомогательных элементов при формировании факультативного гетерохроматина.

Основные положения, выносимые на защиту:

1. Alu повторы, локализованные в промоторах генов лиганда циклофилина, регулирующего кальциевый ответ (CAML) и дезоксирибонуклеазы II (DNAse II), участвуют в регуляции их транскрипции в клетках линий HEK293 и A549 человека.
2. Инсерция AluYa5 повтора в 16-й интрон гена АПФ коррелирует с уровнем фенилаланина и смеси лейцин+изолейцин в плазме крови людей.
3. Alu-ассоциированные кластеры сайтов связывания транскрипционных факторов (ССТФ) способствуют формированию в ходе эволюции регуляторных элементов в промоторах и интронах некоторых человеческих генов.
4. Распределения Alu повторов человека и SINE повторов мыши (B1, B2), но не LINE1 повторов, в X хромосомах коррелируют с распределениями CpG островков, белок-кодирующих генов и некоторых маркеров гетерохроматина.

Вспомогательными элементами при формировании факультативного гетерохроматина в ходе инактивации X хромосомы могут служить SINE-ассоциированные негативные элементы транскрипции.

Научная новизна. Впервые предложен принцип поиска в промоторах генов Alu повторов, участвующих в регуляции экспрессии генов на уровне транскрипции. Этот принцип подтвержден экспериментально на примере Alu повторов, локализованных в промоторах генов CAML и DNaseII. Впервые показано, что инсерция AluYa5 повтора в 16-й интрон гена АПФ коррелирует с уровнем некоторых аминокислот в плазме крови людей. На этом примере показана функциональная роль Alu повторов, локализованных в интронах. Впервые показано, что Alu повторы человека и SINE повторы мыши могут участвовать в регуляции генной экспрессии на уровне формирования факультативного гетерохроматина в ходе инактивации X хромосомы.

Теоретическое и практическое значение работы. Полученные в работе результаты важны для понимания функций повторяющихся элементов генома. Работа вносит определенный вклад в современные представления о механизмах регуляции экспрессии генов. В практическом отношении представляет интерес установленная корреляция между инсерционно-делеционным полиморфизмом в гене АПФ и уровнем фенилаланина и смеси лейцин+изолейцин в плазме крови людей. Коррекция уровня свободных аминокислот имеет значение при лечении некоторых наследственных заболеваний, в частности, аминокислотопатий. Результаты настоящего исследования могут быть использованы в курсах лекций по клеточной и молекулярной биологии для студентов медицинских и биологических специальностей высших учебных заведений.

Апробация работы. Основные положения работы были представлены на Международной молодежной научно-методической конференции «Проблемы молекулярной и клеточной биологии» (Томск, 2007), Международной конференции «Рецепция и внутриклеточная сигнализация» (Пушино, 2007), Международной конференции «II съезд Общества клеточной биологии совместно с юбилейной конференцией, посвященной 50-летию Института цитологии РАН» (Санкт-Петербург, 2007). Материалы диссертации докладывались и обсуждались на научных семинарах

лаборатории стабильности хромосом и клеточной инженерии Института цитологии РАН.

Объем и структура диссертации. Диссертация состоит из введения, обзора литературы, описания материалов и методов исследования, результатов и их обсуждения, выводов и списка цитируемой литературы, включающего 274 ссылки. Диссертация изложена на 151 странице и иллюстрирована 38 рисунками и 10 таблицами.

Список основных сокращений.

CAML – ген лиганда циклофилина, регулирующего кальциевый ответ, **DNaseII** – ген дезоксирибонуклеазы II, **LINE** – (long interspersed nucleotide elements) – длинные диспергированные нуклеотидные элементы, **L1** – повторы LINE1 семейства, **PcG** – белки группы Polycomb, **SINE** – (short interspersed nucleotide elements) – короткие диспергированные нуклеотидные элементы, **XIC** – (X inactivation center) – центр инактивации X хромосомы, **АПФ** – ангиотензинпревращающий фермент, **ПЦР** – полимеразная цепная реакция, **ССТФ** – сайты связывания транскрипционных факторов, **ЦРМ** – *цис*-регуляторный модуль.

МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

Компьютерный анализ. Из базы данных промоторов DBTSS (<http://dbtss.hgc.jp>) произвольным образом выбрали 3375 промоторов. Каждый промотор содержал 1000 н.п. до и 200 н.п. после точки инициации транскрипции. Нуклеотидные последовательности X хромосом человека и мыши взяли из базы данных Ensembl v.44 (www.ensembl.org). Alu повторы человека, SINE повторы мыши, LINE1 повторы и CpG островки идентифицировали в промоторах, в 16-м интроне гена АПФ и в X хромосомах при помощи программ RepeatMasker (www.repeatmasker.org) и CPGPLOT (www.ebi.ac.uk/emboss/cpgplot/). Весовые матрицы для поиска ССТФ Sp1, YY1 и ER α взяли из базы данных транскрипционных факторов TRANSFAC (www.gene-regulation.com/pub/databases.html/). Поиск ССТФ в промоторах осуществляли при помощи программы RSA-tools-patser (<http://rsat.ulb.ac.be/rsat/>).

Клонирование фрагментов промоторов генов DNase II и CAML в плазмидный вектор pGL3-Basic. При помощи полимеразной цепной реакции (ПЦР) амплифицировали различные фрагменты промоторов генов DNase II и CAML.

Последовательности олигонуклеотидных праймеров подбирали при помощи программы Vector NTI™ (InforMax, США). Для удобства клонирования в 5'-концы праймеров вводили сайты для эндонуклеаз рестрикции *Bgl*II, *Xho*I или *Hind*III. Для более точной амплификации мы использовали смесь двух полимераз - Platinum®Pfx DNA Polymerase и *Taq* DNA Polymerase (Invitrogen, США). Температуру отжига ($T_{отж}$ °C) подбирали экспериментально для каждой пары праймеров. В качестве матрицы использовали геномную ДНК, выделенную из лейкоцитарной фракции периферической крови человека. По окончании ПЦР реакционную смесь очищали от излишков праймеров, dNTPs и полимераз при помощи QIAquick PCR Purification Kit (Quiagen, Германия). Продукты ПЦР и вектор pGL3-Basic (Promega, США) обрабатывали эндонуклеазами рестрикции *Bgl*II, *Xho*I или *Hind*III. Векторную ДНК дополнительно обрабатывали щелочной фосфатазой Shrimp Alkaline Phosphatases (Fermentas, Литва). После рестрикции фрагменты ДНК разделяли при помощи электрофореза в 0,7 %-ом агарозном геле и выделяли из геля при помощи QIAquick Gel Extraction Kit (Quiagen, Германия). Затем проводили лигирование вектора и вставки в соотношении 1:3 при помощи ДНК-лигазы фага T4 (Invitrogen, США). Трансформацию бактериального штамма *E. coli* DH5 α продуктами лигазной реакции проводили по методу Мандела и Хиги [11] с применением хлористого кальция. Плазмидную ДНК из положительных трансформантов выделяли при помощи Plasmid Midi Kit (Quiagen, Германия). Наличие необходимых фрагментов промоторов в рекомбинантных плазмидах подтверждали секвенированием.

Трансфекция клеточных линий человека рекомбинантными плазмидами и измерение активности люциферазы. Клетки карциномы легкого человека линии A549 и клетки почки эмбриона человека линии HEK293, предоставленные банком клеточных культур Института цитологии РАН, культивировали в среде RPMI 1640 (Биолот, Россия), содержащей 10 % эмбриональной сыворотки телят (Gibco BRL, США), 1 mM L-глутамина, 100 ед/мл пенициллина и 100 мкг/мл стрептомицина. Клетки котрансфицировали 0,5 мкг ДНК конструкции и 5 нг плазмиды pRL-ТК при помощи реагента Lipofectamine™ 2000 (Invitrogen, США) в соответствии с рекомендациями фирмы-производителя. Через 1 сутки после трансфекции клетки лизировали при помощи буфера Passive Lysis Buffer (Promega, США). Активность люцифераз измеряли ручным люминометром TD 20/20 DLReady™ Luminometer

(Turner BioSystems, США) при помощи Dual-Luciferase® Reporter Assay system (Promega, США) в соответствии с инструкциями изготовителя. Векторы pGL3-Basic и pGL3-Control (Promega, США) использовали как отрицательный и положительный контроли соответственно. Трансфекцию и измерения активности люциферазы в лизатах проводили трижды в двойной повторности.

Определение I/D полиморфизма в гене АПФ. Геномную ДНК выделяли из лейкоцитарной фракции периферической крови. К пробе крови добавляли лизирующий буфер, содержащий 0,15 мг/мл протеиназы К, 0,1 М NaCl, 20 мМ ЭДТА, 1 % додецилсульфата натрия и 50 мМ трис-HCl. Полученную смесь инкубировали в течение ночи при 37 °С. ДНК выделяли из лизата стандартным фенол-хлороформным методом. Амплификацию фрагмента 16-го интрона гена проводили при помощи ПЦР. В качестве праймеров использовали следующие олигонуклеотидные последовательности: прямой праймер 5'- CTG GAG ACC ACT CCC ATC CTT TCT - 3', обратный праймер 5'- GAT GTG GCC ATC ACA TTC GTC AGA T - 3' [12]. Температура отжига праймеров составляла 60 °С. Материал, полученный в результате ПЦР, подвергали разделению при помощи гель-электрофореза в 8% полиакриламидном геле. По окончании разделения гель окрашивали при помощи бромистого этидия. Фрагменты визуализировали в проходящем ультрафиолетовом свете.

Определение уровня свободных аминокислот в плазме крови. Определение уровня свободных аминокислот в плазме крови проводили в клинико-диагностической лаборатории Санкт-Петербургской Медицинской Академии Последипломного Образования методом высокоэффективной жидкостной хроматографии.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Выбор генов для изучения функциональной роли Alu повторов, локализованных в промоторах. Консенсусная последовательность Alu повторов содержит *цис*-регуляторные элементы, которые взаимодействуют с различными эффекторами: гормонами, кальцием, транскрипционными факторами [7, 13]. Единичные изолированные ССТФ, как правило, не оказывают существенного влияния на транскрипцию большинства эукариотических генов. Важную роль в регуляции экспрессии генов играют так называемые *цис*-регуляторные модули

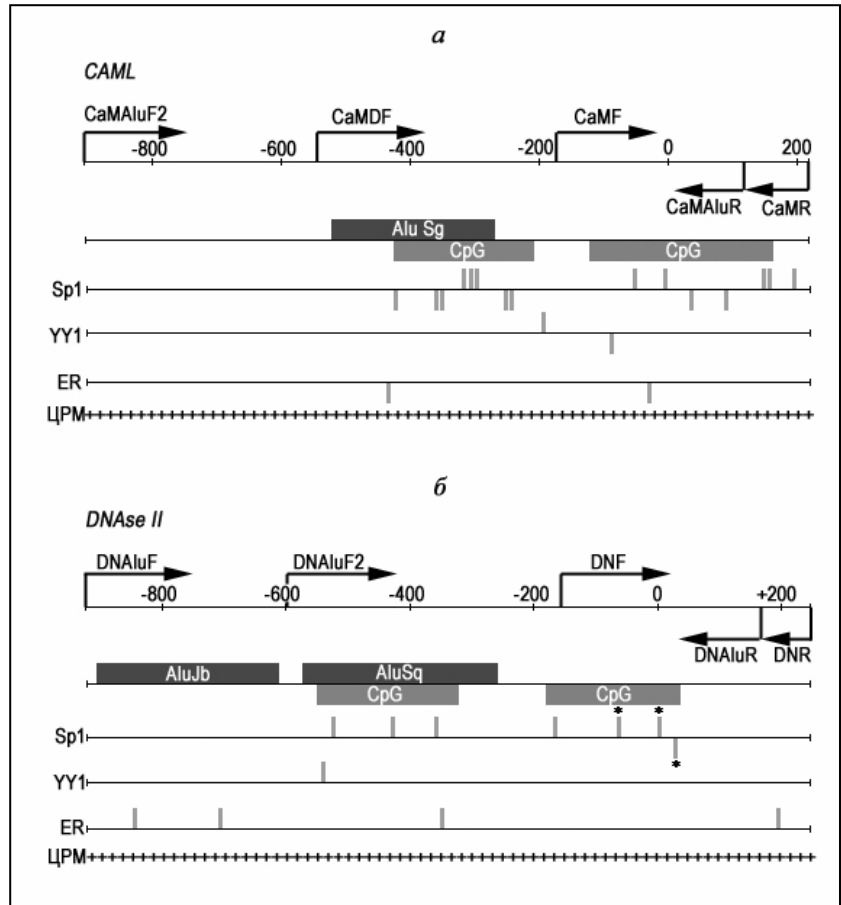
(ЦРМ). Они представляют собой последовательности ДНК длиной 100-1000 н.п., с которыми взаимодействует целый набор транскрипционных факторов. Различными группами исследователей [14-16] было показано, что важную роль в регуляции транскрипции в ходе раннего эмбрионального развития у *Drosophila melanogaster* играют кластеры сайтов связывания 3-6 транскрипционных факторов, плотность которых превышает некоторое критическое значение. Для изучения функциональной роли Alu повторов мы решили выбрать промоторы белок-кодирующих генов, которые содержат в своем проксимальном сегменте Alu элементы, перекрывающиеся с CpG островками и ЦРМ. ЦРМ мы считали те области в составе промоторов, которые, имея длину не более 700 н.п., содержали 7 или более ССТФ Sp1, ER α , YY1. При помощи методов компьютерного анализа мы идентифицировали перечисленные элементы в последовательностях 3375 промоторов, выбранных случайным образом.

Фракция Alu повторов в промоторах составила 7,5%, что ниже содержания Alu в интронах, в межгенных промежутках, а также в геноме в целом. Вероятно, в ходе эволюции выработались механизмы, препятствующие inserции Alu повторов в промоторы белок-кодирующих генов. Оказалось, что с Alu повторами перекрываются лишь 133 из 3044 идентифицированных в промоторах CpG островков. Из этих 133 CpG островков 48 (1,58%) находятся в проксимальном положении в промоторах. Большая часть Alu повторов в геноме метилирована [17] и, следовательно, путем спонтанной реакции дезаминирования в них происходит постепенное снижение фракции CpG-динуклеотидов [18]. Остается неясным, почему некоторые члены этого семейства повторов неметилированы и способствуют формированию CpG островков. Мы полагаем, что Alu повторы, ассоциированные с CpG островками, участвуют в регуляции генной экспрессии, так как могут нести в своем составе важные *цис*-регуляторные элементы для РНК-полимеразы II.

В результате компьютерного анализа мы идентифицировали несколько белок-кодирующих генов, промоторы которых полностью удовлетворяют нашим критериям. Например, ген выживаемости мотонейронов (SMN), ген человеческого гомолога белка Rad51 дрожжей (hRad51), ген субъединицы 4 пуриnergического рецептора (P2X), ген субъединицы IV третьего фактора инициации трансляции эукариот (EIFS4), ген, ассоциированный с многоочаговой лейкоэнцефалопатией (MLC1), ген поли(АДФ-рибоза) полимеразы 1 (ADPRT1), ген субъединицы В

сукцинатдегидрогеназного комплекса (SDHB). В качестве моделей для дальнейшего исследования функциональной роли Alu повторов в промоторах были выбраны промоторы двух генов: дезоксирибонуклеазы II (DNase II) и лиганда циклофилина, регулирующего кальциевый ответ (CAML) (рис. 1).

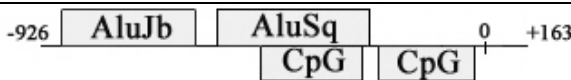
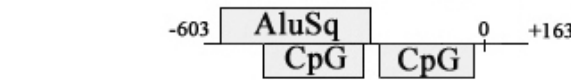

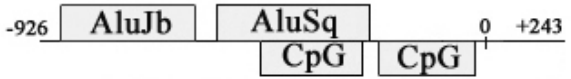


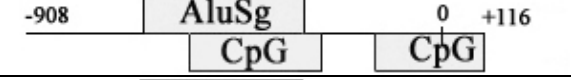
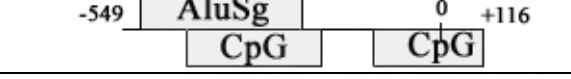

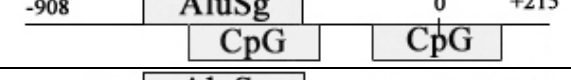
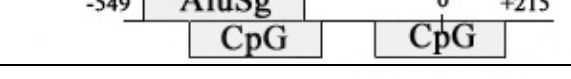
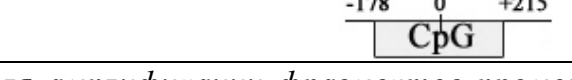
Рис. 1. Схема строения промоторов генов *CAML* (а) и *DNase II* (б) и расположения олигонуклеотидных праймеров, используемых в ПЦР. *Sp1*, *YY1*, *ER* – сайты связывания транскрипционных факторов *Sp1*, *YY1*, *ER*; ЦРМ – цис-регуляторный модуль; *CaMAluF2*, *CaMDF*, *CaMF*, *DNAIuF*, *DNAIuF2*, *DNF* – прямые олигонуклеотидные праймеры, *CaMAluR*, *CaMR*, *DNAIuR*, *DNR* – обратные олигонуклеотидные праймеры; *CpG* – *CpG* островки; *AluSg*, *AluJb*, *AluSq* – *Alu* повторы различных подсемейств. Ноль на оси соответствует точке инициации транскрипции. Звездочками отмечены сайты связывания *Sp1*, функциональная значимость которых была доказана экспериментально [19].



На первом этапе мы амплифицировали различные фрагменты этих промоторов при помощи ПЦР. Олигонуклеотидные праймеры для реакции выбирали таким образом, чтобы ее продукты различались по составу элементов: Alu повторов и CpG островков (Табл. 1). Продукты ПЦР субклонировали в полилинкер плазмиды pGL3-Basic, содержащей репортерный ген люциферазы *Photinus pyralis* (рис. 2). Полученными конструкциями трансфицировали клетки линии карциномы легкого человека (A549) и почки эмбриона человека (HEK293). Эти линии были выбраны по следующим причинам: по литературным данным в легких и почках наблюдается высокий уровень экспрессии генов DNase II и CAML, культивирование этих клеток не требует особых условий и сложных по составу сред, а эффективность трансфекции этих клеток относительно высокая. Для внутреннего контроля эффективности

трансфекции клеток мы проводили котрансфекцию полученных конструкций с вектором pRL-ТК, который содержит ген люциферазы *Renilla reniformis* и промотор гена тимидинкиназы вируса простого герпеса. Вектор pGL3-Control использовали в качестве положительного контроля трансфекции. Он отличается от вектора pGL3-Basic тем, что перед геном люциферазы в него встроен промотор вируса SV40 обезьян, а после гена – его энхансер.

Таблица 1. Состав амплифицируемых фрагментов промоторов

№ п/п ^а	Пара праймеров	Амплифицируемый фрагмент ^б	Длина, н.п.	Состав амплифицируемого фрагмента
1	DNA _l F-DNA _l R	-926/+163	1089	-926  +163
2	DNA _l F2-DNA _l R	-603/+163	766	-603  +163
3	DNF-DNA _l R	-161/+163	324	-161  +163
4	DNA _l F-DNR	-926/+243	1169	-926  +243
5	DNA _l F2-DNR	-603/+243	846	-603  +243
6	DNF-DNR	-161/+243	404	-161  +243
7	CaMA _l F2-CaMA _l R	-908/+116	1024	-908  +116
8	CaMDF-CaMA _l R	-549/+116	665	-549  +116
9	CaMF-CaMA _l R	-178/+116	294	-178  +116
10	CaMA _l F2-CaMR	-908/+215	1123	-908  +215
11	CaMDF-CaMR	-549/+215	764	-549  +215
12	CaMF-CaMR	-178/+215	393	-178  +215

^аПары праймеров №№1-6 используются для амплификации фрагментов промотора гена DNase II, №№7-12 – SAML, ^бПоложение амплифицируемого фрагмента указано относительно точки инициации транскрипции.

Результаты измерения активности люциферазы в лизатах клеток, трансфицированных различными конструкциями. В ходе измерений определяли

относительную люциферазную активность лизата – отношение интенсивностей люминесценции, сопровождающей две люциферазные реакции: с участием люциферазы *Photinus pyralis* и с участием люциферазы *Renilla reniformis*. Результаты измерения активности люциферазы в лизатах клеток линий A549 и HEK293, трансфицированных различными конструкциями, представлены на рис. 3. Относительная люциферазная активность для положительного и отрицательного контролей составила соответственно $2,0 \pm 0,1$ и 415 ± 25 (для клеток линии A549) или $2,6 \pm 0,7$ и 3272 ± 170 (для клеток линии HEK293). Клетки линии HEK293 трансфицировали только конструкциями –908/+116, –549/+116, –178/+116 (CAML) и –926/+243, –603/+243, –116/+243 (DNase II). Выбор данных конструкций был обусловлен результатами, полученными при измерении активности люциферазы в лизатах трансфицированных клеток линии A549.

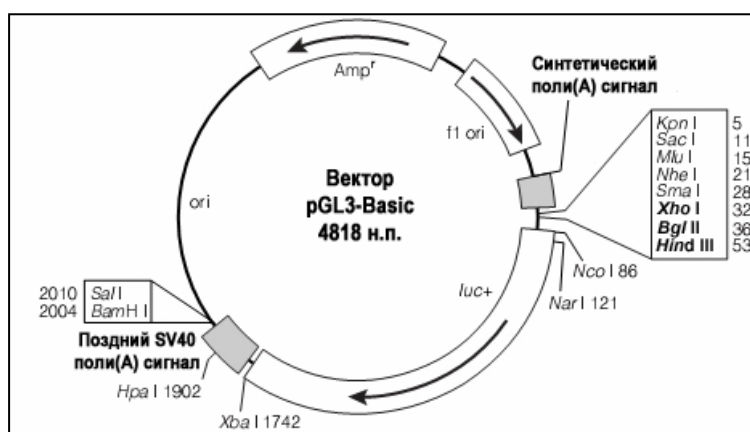


Рис. 2. Карта плазмиды pGL3-Basic: *luc+* - ген люциферазы *Photinus pyralis*, *Amp^r* - ген β-лактамазы, обуславливающий устойчивость бактерии-хозяина к ампициллину, *f1 ori* – *f1* ориджин репликации. Жирным ирифтом выделены эндонуклеазы рестрикции, по сайтам которых проводили клонирование фрагментов промоторов.

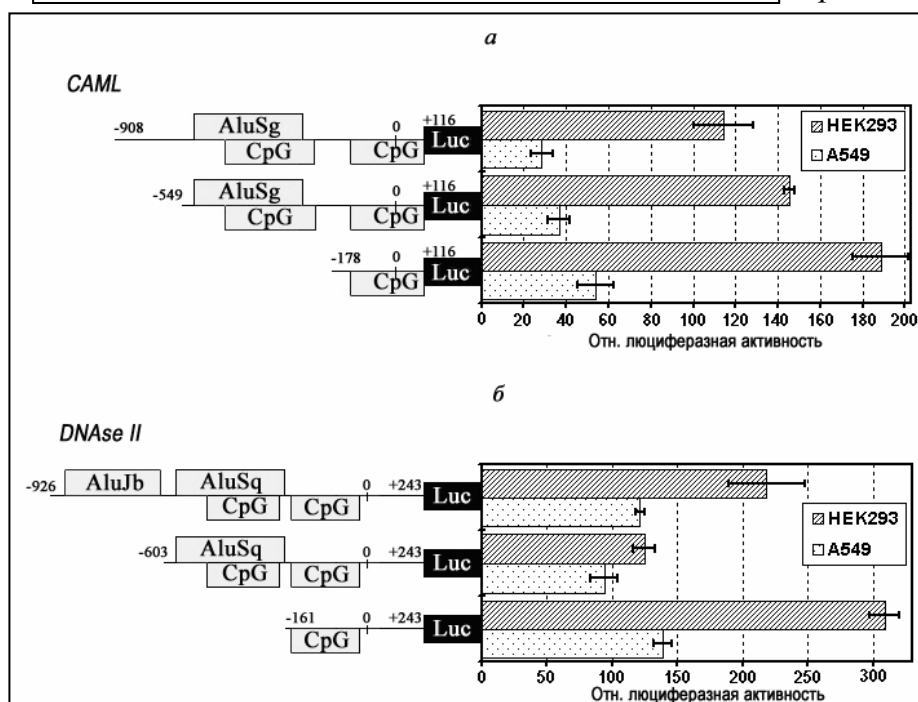


Рис. 3. Результаты измерения активности люцифераз в лизатах клеток линий A549 и HEK293, трансфицированных конструкциями, содержащими различные фрагменты промоторов генов CAML (а) и DNase II (б).

Заштрихованные столбики – результаты для клеток линии HEK293, светлые – для клеток линии A549.

Для конструкций с фрагментами промотора гена *SAML* (рис. 3, *а.*) самая высокая активность (для обеих клеточных линий) соответствовала конструкции –178/+116, которая не содержала в своем составе *Alu*. Активность люциферазы в лизатах клеток, трансфицированных конструкцией –549/+116, была на 32 (линия A549) и на 23 % (линия HEK293) ниже. Эта конструкция включает в себя фрагмент промотора, содержащий повтор *AluSg*. Активность люциферазы, соответствующая конструкции –908/+116, достоверно не отличалась от активности, соответствующей конструкции –549/+116, в клетках линии A549, но отличалась в клетках линии HEK293. Конструкции –908/+215, –549/+215, –178/+215 при трансфекции в клетки линии A549 показали активность на порядок меньшую, чем конструкции –178/+116, –549/+116 и –908/+116 (на рисунке не показано, см. стр. 97 диссертации). При этом активности, соответствующие этим конструкциям, достоверно не отличались друг от друга. Для конструкций с фрагментами промотора гена *DNAse II* (рис. 3, *б.*) самая высокая активность (для обеих клеточных линий) также соответствовала конструкции, не содержащей в своем составе *Alu* (–161/+243). Активность люциферазы в клетках, трансфицированных конструкцией –603/+243, содержащей *AluSq* повтор, была на 32 (линия A549) и на 60 % (линия HEK293) ниже. Люциферазная активность в лизатах клеток, трансфицированных конструкцией –926/+243, содержащей повторы *AluJb* и *AluSq*, была также ниже, но всего на 13 (линия A549) и на 32 % (линия HEK293). После трансфекции в клетки линии A549 конструкции –926/+163, –603/+163, –161/+163 показали активности люциферазы того же порядка, что и другие конструкции с фрагментами промотора гена *DNAse II*. Эти активности не отличались друг от друга, но достоверно отличались от активности, соответствующей конструкции –161/+243 (на рисунке не показано, см. стр. 97 диссертации).

Таким образом, исследования промотора гена *SAML* показали, что *AluSg*-повтор, находящийся в районе –526/–271, ослабляет экспрессию гена как в клетках линии A549, так и в клетках линии HEK293. Конструкции –908/+215, –549/+215, –178/+215 для клеток линии A549 показали активность на порядок меньшую, чем конструкции –178/+116, –549/+116 и –908/+116. Это может быть связано с тем, что на участке +116/+215 расположены два ССТФ *Sp1* или же сайты связывания других транскрипционных факторов, которые мы не рассматривали в настоящей работе.

Исследования промотора гена DNase II показали, что AluSq повтор, находящийся в проксимальном сегменте, ослабляет, а повтор AluJb, локализованный в дистальном сегменте, усиливает экспрессию гена. С AluJb повтором перекрываются два ССТФ ER α . Возможность функционирования Alu повтора как ER-зависимого транскрипционного энхансера была показана на примере гена BRCA-1 [20]. Помимо этого, воспользовавшись базой данных SAGEmap (www.ncbi.nlm.nih.gov/sage), мы определили, что специфический маркер транскриптов гена DNase II (AGCTGAGCTA) встречается в клетках MCF7, обработанных эстрадиолом, в 6 раз чаще. Наши результаты свидетельствуют о том, что Alu повторы в составе промоторов генов CAML и DNase II обладают аналогичной активностью в клетках линии A549 и HEK293. Возможно, в клетках других типов эти Alu повторы обладают прямо противоположной активностью или не имеют таковой вовсе. Тканеспецифическая активность некоторых сайтов для YY1 обнаружена на примере миозиновых генов мышей [21]. Экспрессия обоих генов осуществляется во всех тканях человека. Можно сделать предположение, что негативные Alu-ассоциированные элементы в промоторах генов CAML и DNase II не работают в тех тканях, где гены экспрессируются на высоком уровне: в лимфоцитах, тканях мозга и мочеполовой системы – ген CAML, в тканях простаты, легких, щитовидной железы – ген DNase II (<http://symatlas.gnf.org/SymAtlas/>).

В гене паратиреоидного гормона был обнаружен негативный кальций-чувствительный элемент 2-го типа (nCARE2) расположенный в Alu повторе, находящимся за 3,6 т.п.н. вверх по течению от точки инициации транскрипции [22]. Таким образом, Alu повторы могут выполнять регуляторную функцию, находясь за несколько тысяч пар нуклеотидов от точки инициации транскрипции. Мы обнаружили, что область -5000/0 гена DNase II содержит еще 3 Alu повтора, помимо тех, которые указаны на рис. 1. Область -5000/0 гена CAML содержит еще 6 полноразмерных и 7 частично делетированных Alu повторов. Вполне вероятно, что указанные последовательности также участвуют в регуляции экспрессии этих генов. Помимо этого, интроны указанных генов также содержат многочисленные Alu элементы.

Вероятнее всего, инсерция Alu повторов в промоторы генов CAML и DNase II произошла до дивергенции архантропов и высших обезьян, так как они содержатся в

гомологичных локусах шимпанзе (*Pan troglodytes*) и макаки-резуса (*Macaca mulatta*). При этом число ССТФ, ассоциированных с Alu повторами, неодинаково у разных видов. У человека ЦРМ полностью перекрываются с промоторами, в то время как у обезьян в промоторах уже можно обнаружить фрагменты, в которых плотность ССТФ меньше критического значения. Возможно, что данные гены регулируются у обезьян посредством Alu повторов иным образом или не регулируются вовсе. На данном примере видно, что в ходе эволюции может происходить формирование ЦРМ. Изначальное наличие в промоторах Alu повторов, перекрывающихся с ССТФ, ускоряет этот процесс.

Анализ 16-го интрона гена АПФ. В результате компьютерного анализа в составе нуклеотидной последовательности 16-го интрона гена АПФ нами был обнаружен ЦРМ, перекрывающийся с двумя Alu повторами и CpG островком (рис. 4.). Полиморфизм гена АПФ по типу инсерция/делеция (I/D) определяется наличием или отсутствием AluYa5 повтора в 16-м интроне.

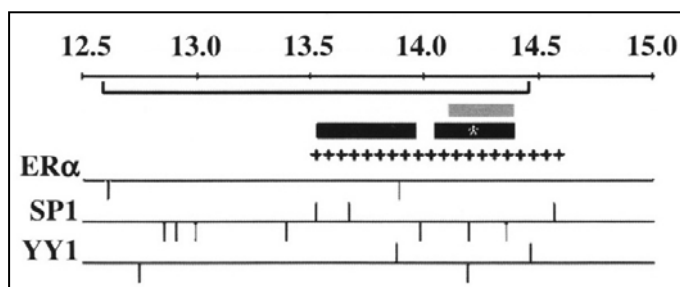


Рис. 4. ЦРМ в составе 16-го интрона гена АПФ. Звездочкой обозначен AluYa5-повтор, серым цветом – CpG островок, плюсами – ЦРМ. Цифрами обозначено расстояние в т.п.н. от точки инициации транскрипции.

Известно, что у инсерционных гомозигот (II) уровень АПФ в плазме крови вдвое ниже, чем у делеционных гомозигот (DD) [23]. На основании того, что Alu повтор локализован в интроне, многие исследователи считают маловероятной возможность его непосредственного участия в регуляции экспрессии гена. В большинстве случаев они склоняются к мысли, что Alu повтор находится в неравновесном сцеплении с каким-либо локусом количественных признаков. Однако поиски функциональных полиморфизмов пока что не увенчались успехом [24]. Мы полагаем, что в случае гена АПФ AluYa5 повтор играет роль транскрипционного сайленсера. Примеры негативной регуляции экспрессии генов транскрипционными факторами YY1 и ERα хорошо известны [25, 26], фактор Sp1 также может выступать в роли репрессора транскрипции [27].

Влияние ID-полиморфизма в гене АПФ на уровень фенилаланина и смеси лейцин+изолейцин в плазме крови людей. Проанализировав ди- и трипептиды,

отщепляемые АПФ от его основных субстратов, мы пришли к выводу, что в их состав входит неодинаковое количество различных аминокислотных остатков. Поскольку уровень фермента в крови у делеционных гомозигот превосходит таковой у инсерционных гомозигот почти в два раза, у нас возникло предположение о возможной корреляции между инсерцией *AluYa5* повтора в 16-й интрон и уровнем некоторых аминокислот в плазме крови людей. Нами были определены I/D полиморфизм в гене АПФ и уровень свободных аминокислот в плазме крови у 102 человек [28]. Статистическая обработка материалов проводилась при помощи t-критерия Стьюдента. Во всех случаях различия в уровнях свободных аминокислот между II и ID, а так же между DD и ID оказались недостоверными. У ID этот уровень практически не отличался от средних значений. У DD по отношению к II оказались повышены уровни Leu+Ile ($P = 0,03$) и Phe ($P = 0,004$).

Анализ X хромосом человека и мыши. Инактивация X хромосомы – это один из наиболее примечательных примеров эпигенетической регуляции генной экспрессии, посредством которой происходит полное подавление транскрипции большинства генов одной из двух хромосом в диплоидных соматических клетках самок млекопитающих. Она инициируется в так называемом «центре инактивации» (XIC) и распространяется в обе стороны *in cis* [29]. Самым важным геном этого района является *Xist*. Усиленная транскрипция гена *Xist* начинается на ранних стадиях эмбриогенеза. Образующаяся нетранслируемая *Xist* РНК специфически покрывает инактивированную X хромосому в течение клеточного цикла. Это приводит к многочисленным последовательным модификациям гистонов, которые способствуют установлению и поддержанию транскрипционно неактивного гетерохроматина [30, 31]. Еще одним признаком гетерохроматина является метилирование CpG островков, локализованных в промоторах [32]. Согласно принятой модели, в X хромосоме имеются специализированные вспомогательные элементы («way stations» или «boosters»), которые способствуют процессу распространения инактивации [33]. Сначала *Xist* РНК связывается с вспомогательными элементами, расположенными около XIC, индуцируя тем самым локальные конформационные изменения. Вследствие этого становится возможным связывание с белками, способствующими гетерохроматинизации. РНК-белковые комплексы привлекают более удаленные вспомогательные элементы, происходит

кооперативное распространение инактивации. Была выдвинута гипотеза, утверждающая, что подобными «way station» являются LINE повторы, в частности, LINE1 (L1) повторы мыши и человека [34].

При помощи методов компьютерного анализа мы установили, а затем сравнили распределения генов, SINE и L1 повторов, CpG островков и некоторых маркеров гетерохроматина в X хромосомах человека и мыши (рис. 5.). Полученные нами результаты свидетельствуют о том, что распределение SINE повторов (Alu и B1, B2-B4), но не L1 повторов, сходно с распределениями CpG островков и таких важных признаков факультативного гетерохроматина в X хромосоме, как Xist РНК и H3-K27m3 (триметилирование гистона H3 по лизину в 27-м положении). Известно, что распределение другого маркера гетерохроматина – убиквитинилированного гистона H2A по лизину в 119-м положении – также совпадает у мыши с распределением ген-богатых сегментов, а также с распределением Xist РНК [35]. Исходя из вышесказанного, мы полагаем, что именно SINE повторы млекопитающих играют роль вспомогательных элементов при распространении инактивации вдоль X хромосомы.

Важную роль в инактивации X хромосомы играют белки группы Polycomb (PcG), которые могут образовывать два белковых комплекса - Polycomb repressive complex 1 и 2. Белки EED и Ring1A/B этой группы осуществляют метилирование гистона H3 по лизину в 27-м положении и убиквитинилирование гистона H2A соответственно [36, 37]. Транскрипционный фактор YY1 единственный из всей группы PcG способен специфически связываться с последовательностью ДНК. Мы считаем, что взаимодействие комплексов PcG с SINE повторами может осуществляться за счет связывания с ССТФ YY1. Известно, что ССТФ YY1 присутствуют в консенсусной последовательности Alu. Кроме того, непосредственное связывание YY1 с Alu было доказано нами экспериментально [38].

Мы обнаружили (рис. 6.), что распределение Alu повторов (но не L1 повторов) строго коррелирует с распределением ССТФ YY1, то есть большое число Alu повторов, локализованных в X хромосоме, могут связывать YY1. Мотив CCAATSTTGNC, который является консенсусной последовательностью ССТФ YY1 у человека, не обнаруживает корреляции с SINE повторами мыши. При этом, консенсусная последовательность B2 элементов у мыши содержит мотив

GCCATCTCACC, который, вероятно, и является сайтом связывания YY1 у мышей, что, однако, требует экспериментального подтверждения.

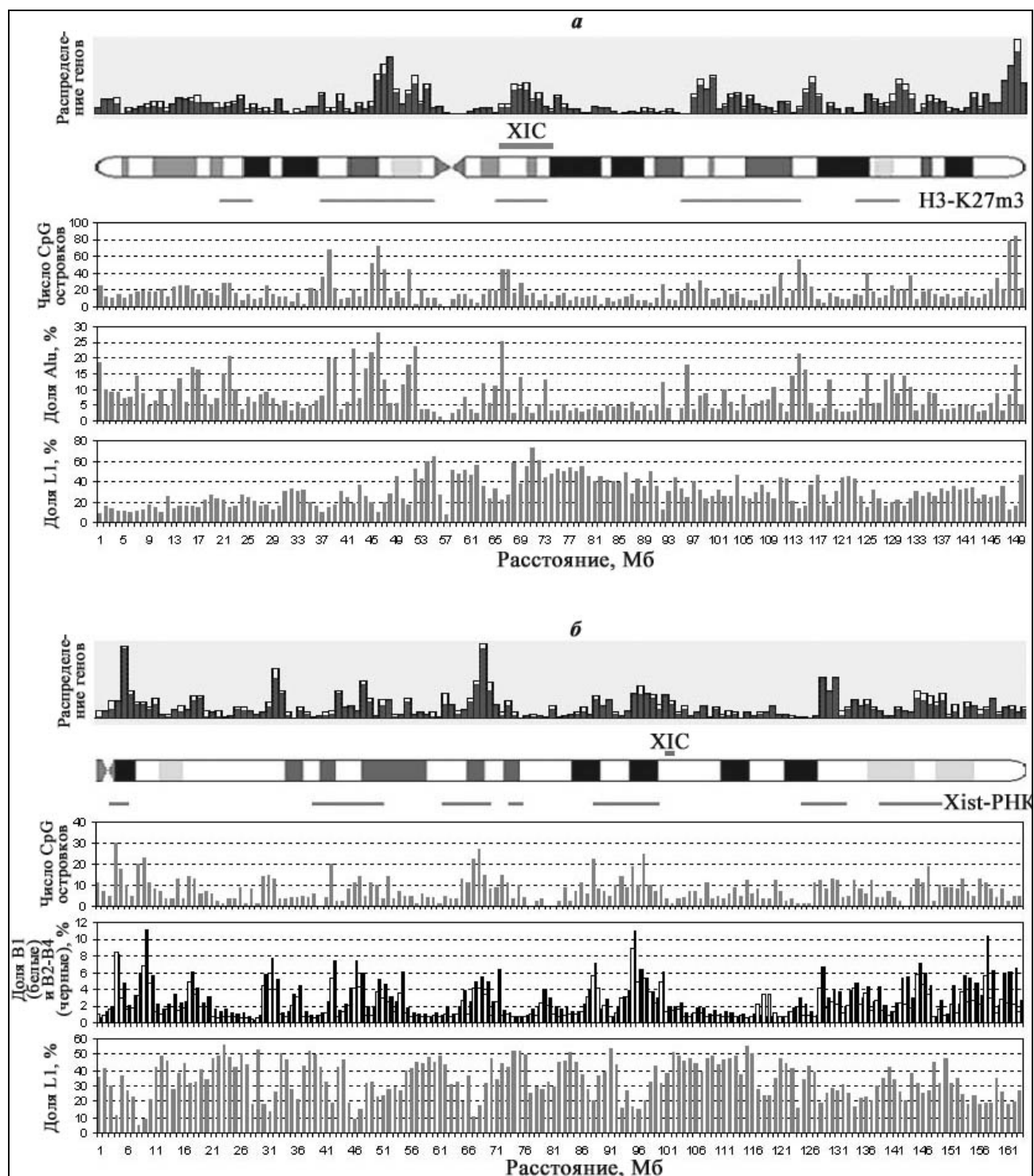


Рис. 5. (а) Распределение генов, *Alu* повторов, *CpG* островков и *LINE* повторов (*L1*) в *X* хромосоме человека. Распределение *H3-K27m3* взято из [39]. (б) Распределение генов, *SINE* повторов (*B1*, *B2-B4*), *CpG* островков и *L1* в *X* хромосоме мыши. Распределение *Xist PRK* взято из [35].

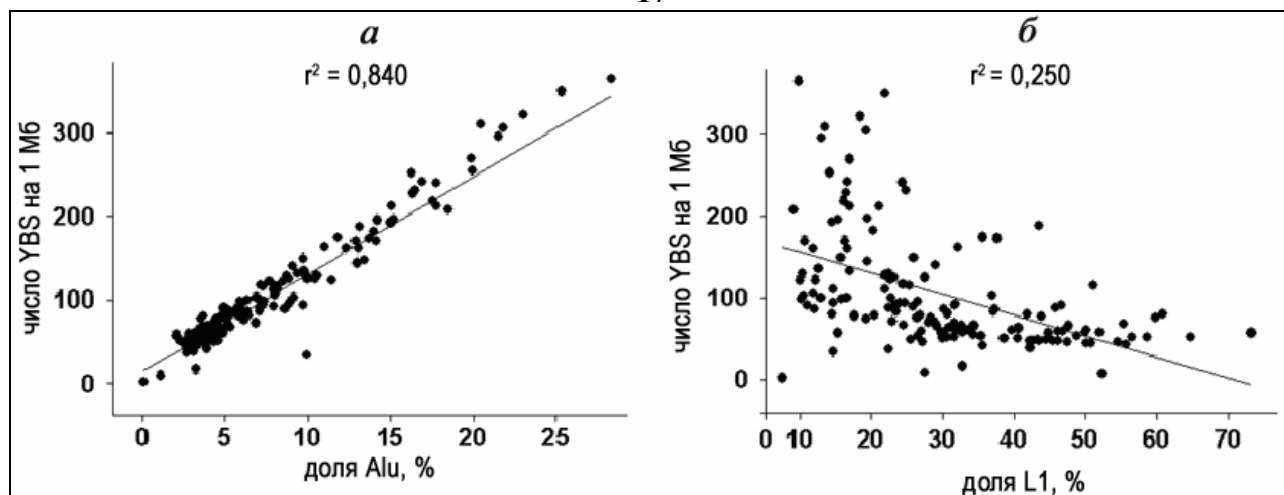


Рис. 6. Корреляция между потенциальными сайтами связывания YY1 (YBS) и плотностью Alu повторов (а) и L1 (б) в X хромосоме человека.

ВЫВОДЫ

1. Транскрипция генов CAML и DNase II в клетках линий HEK293 и A549 регулируется Alu повторами, локализованными в их промоторах.
2. Инсерция AluYa5-повтора в 16-й интрон гена ангиотензинпревращающего фермента приводит к формированию в его составе *цис*-регуляторного модуля и коррелирует с уровнем Phe и Leu+Ile в плазме крови людей.
3. Глобальное распределение Alu повторов человека и SINE повторов мыши (B1, B2) в X хромосомах коррелирует с распределением CpG островков и белок-кодирующих генов.
4. SINE-ассоциированные негативные элементы транскрипции могут служить вспомогательными элементами при формировании факультативного гетерохроматина в ходе инактивации X хромосомы.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Oei S.L., Babich V.S., Kazakov V.I., **Usmanova N.M.**, Kropotov A.V., Tomilin N.V. 2004. Clusters of regulatory signals for RNA polymerase II transcription associated with Alu family repeats and CpG islands in human promoters. *Genomics*. 83: 873-882.
2. Казаков В.И., Кадурина Т.И., **Усманова Н.М.**, Томилин Н.В. 2003. Влияние инсерционно-делеционного полиморфизма в гене ангиотензин-превращающего фермента на уровень свободных аминокислот в сыворотке крови больных с дисплазиями соединительной ткани. *Генетика*. 39 (8): 1136-1140.
Имеется перевод: Kazakov V.I., Kadurina T.I., **Usmanova N.M.**, Tomilin N.V. 2003. Insertion-deletion polymorphism of the angiotensin converting enzyme gene and its relationship to serum levels of free amino acids in the patients with connective tissue dysplasias. *Russian Journal of Genetics*. 39: 955-959.
3. **Усманова Н.М.**, Казаков В.И., Томилин Н.В. 2008. Ретропозоны Alu-семейства из *cis*-регуляторных модулей промоторов генов DNase II и SAML влияют на генную экспрессию в клетках A549 и HEK293. *Цитология*. 50 (3): 249-255.
4. **Усманова Н.М.**, Казаков В.И., Томилин Н.В. 2008. SINE элементы в геномах млекопитающих могут служить вспомогательными элементами при формировании факультативного гетерохроматина. *Цитология*. 50 (3): 256-260.
5. **Усманова Н.М.**, Казаков В.И. 2007. Роль Alu повторов в регуляции экспрессии гена дезоксирибонуклеазы II человека. Сборник материалов международной молодёжной научно-методической конференции «Проблемы молекулярной и клеточной биологии», Томск, Май 9–12. 1: 172-173.
6. **Усманова Н.М.**, Казаков В.И., Томилин Н.В. 2007. Особенности строения промотора гена SAML человека и роль Alu повтора в регуляции экспрессии этого гена. Материалы международной конференции «Рецепция и внутриклеточная сигнализация», Пущино, Июнь 5–7. 1: 44-45.
7. **Усманова Н.М.**, Казаков В.И., Томилин Н.В. 2007. О роли ретротранспозонов Alu семейства в негативной регуляции транскрипции в клетках человека. Тезисы докладов и сообщений, представленных на II съезд Общества клеточной биологии совместно с юбилейной конференцией, посвященной 50-летию Института цитологии РАН, Санкт-Петербург, Октябрь 16–19, *Цитология*. 49(6): 800.

СПИСОК ЦИТИРУЕМОЙ ЛИТЕРАТУРЫ

1. Lander E.S. et al. 2001. *Nature*. 409: 860-921.
2. Makalowski W. 2003. *Science*. 300: 1246-1247.
3. Slotkin R.K., Martienssen R. 2007. *Nat. Rev. Genet.* 8: 272-285.
4. Grover D. et al. 2005. *Genetica*. 124: 273-289.
5. Häsler J., Strub K. 2006. *Nucleic Acids Res.* 34: 5491-5497.
6. Paulsen M., Ferguson-Smith A.C. 2001. *J. Pathol.* 195: 97-110.
7. Polak P., Domany E. 2006. *BMC Genomics*. 7: 133.
8. Häsler J., Strub K. 2006. *Nucleic Acids Res.* 34: 2374-2385.
9. Häsler J. et al. 2007. *Cell. Mol. Life Sci.* 64: 1793-1800.
10. Villard J. 2004. *Swiss. Med. Wkly.* 134: 571-579.
11. Mandel M., Higa A. 1970. *J. Mol. Biol.* 53: 159-62.
12. Rigat B. et al. 1992. *Nucleic Acids Res.* 20: 1433.
13. Shankar R. et al. 2004. *BMC Evol. Biol.* 4: 37.
14. Berman B.P. et al. 2002. *Proc. Natl. Acad. Sci. USA.* 99: 757-762.
15. Halfon M.S. et al. 2002. *Genome Res.* 12: 1019-1028.
16. Rajewsky N. et al. 2002. *BMC Bioinformatics*. 3: 30.
17. Weisenberger D.J. et al. 2005. *Nucleic Acids Res.* 33: 6823-6836.
18. Xing J. et al. 2004. *J. Mol. Biol.* 344: 675-682.
19. Chou S.F. et al. 2003. *Eur. J. Biochem.* 270: 1855-1862.
20. Norris J. et al. 1995. *J. Biol. Chem.* 270: 22777-22782.
21. Caretti G. et al. 2004. *Genes Dev.* 18: 2627-2638.
22. Hamdi H.K. et al. 2000. *J. Mol. Biol.* 299: 931-939.
23. Rigat B. et al. 1990. *J. Clin. Invest.* 86: 1343-1346.
24. Sayed-Tabatabaei F.A. et al. 2006. *Circ. Res.* 98: 1123-1133.
25. Gordon S. et al. 2006. *Oncogene*. 25: 1125-1142.
26. Heldring N. et al. 2007. *Physiol. Rev.* 87: 905-931.
27. Won J. et al. 2002. *J. Biol. Chem.* 277: 38230-38238.
28. Казаков В.И. и др. 2003. *Генетика*. 39 (8): 1136-1140.
29. Duthie S.M. et al. 1999. *Hum. Mol. Genet.* 8: 195-204.
30. Chow J.C. et al. 2005. *Annu. Rev. Genomics Hum. Genet.* 6: 69-92.
31. Duret L. et al. 2006. *Science*. 312: 1653-1655.
32. Norris D.P. et al. 1991. *Mamm. Genome*. 1: 78-83.
33. Gartler S.M., Riggs A.D. 1983. *Annu. Rev. Genet.* 17: 155-190.
34. Lyon M.F. 1998. *Cytogenet. Cell. Genet.* 80: 133-137.
35. Smith K.P. et al. 2004. *Chromosoma*. 113: 324-335.
36. Ng K. et al. 2007. *EMBO Rep.* 8: 34-39.
37. Moss T.J., Wallrath L.L. 2007. *Mutat. Res.* 618: 163-174.
38. Oei S.L. et al. 2004. *Genomics*. 83: 873-882.
39. Chadwick B.P., Willard H.F. 2004. *Proc. Natl. Acad. Sci. USA.* 101: 17450-17455.