

**Федеральное государственное бюджетное учреждение науки  
ИНСТИТУТ БИООРГАНИЧЕСКОЙ ХИМИИ  
им. академиков М.М. Шемякина и Ю.А. Овчинникова  
Российской академии наук**

*УДК 577.214.6*  
На правах рукописи

Акопов Сергей Борисович

**СТРУКТУРНО-ФУНКЦИОНАЛЬНЫЙ АНАЛИЗ ЭНХАНСЕРНЫХ И  
ИНСУЛЯТОРНЫХ СИСТЕМ РЕГУЛЯЦИИ ТРАНСКРИПЦИИ**

Специальность 03.01.03 – молекулярная биология

**ДИССЕРТАЦИЯ**

на соискание ученой степени  
доктора биологических наук

Научный консультант:  
академик РАН Е.Д.Свердлов

Москва – 2014

<b>СОДЕРЖАНИЕ</b>	<b>2</b>
<b>Список сокращений</b>	<b>5</b>
<b>1. Введение</b>	<b>8</b>
<b>2. Обзор литературы</b>	<b>13</b>
<b>Регуляторные элементы генома</b>	<b>13</b>
<b>2.1. Инсуляторы</b>	<b>13</b>
2.1.1. <i>Функции инсуляторов</i>	14
2.1.2. <i>Белки, обеспечивающие действие инсуляторов</i>	20
2.1.3. <i>Механизмы блокирования энхансеров</i>	23
2.1.3.1. <i>Модель ловушки (promoter decoy)</i>	23
2.1.3.2. <i>Модель отслеживания (tracking)</i>	24
2.1.4. <i>Нейтрализация эффекта положения</i>	26
2.1.5. <i>Инсуляторы и S/MAR-элементы</i>	27
2.1.6. <i>Регуляция функционирования инсуляторов</i>	29
2.1.7. <i>Локус-контролирующие последовательности</i>	31
<b>2.2. Белок CTCF и его роль в крупномасштабной регуляции активности генома позвоночных</b>	<b>32</b>
2.2.1. <i>CTCF как многофункциональный регулятор</i>	33
2.2.1.1. <i>CTCF и гены, ассоциированные с раком</i>	33
2.2.1.2. <i>CTCF и развитие: регулятор основных регуляторов?</i>	34
2.2.2. <i>CTCF и функционирование генома</i>	36
2.2.2.1. <i>Геномное распределение сайтов связывания CTCF</i>	36
2.2.2.2. <i>CTCF опосредует взаимодействие между регуляторными сайтами на больших расстояниях</i>	38
2.2.3. <i>Взаимодействие CTCF с другими белками</i>	38
2.2.3.1. <i>CTCF и когезиновый комплекс</i>	40
2.2.4. <i>CTCF и повторяющиеся элементы генома</i>	41
2.2.5. <i>Регуляция гена CTCF</i>	41
2.2.5.1. <i>Возможные механизмы регуляторной активности CTCF</i>	42
<b>2.3. Ретроэлементы</b>	<b>44</b>
2.3.1. <i>Краткая история исследования ретровирусов и ретроэлементов</i>	44
2.3.2. <i>Жизненный цикл и строение ретровирусов, синтез LTR</i>	45
2.3.3. <i>Возникновение, структура и классификация ретроэлементов</i>	47
2.3.4. <i>Структура эндогенных ретровирусов</i>	52
2.3.5. <i>Семейство HERV-K(HML-2)</i>	53
2.3.6. <i>Функциональные элементы в составе LTR</i>	54
2.3.7. <i>Биологическая роль LTR</i>	58
2.3.7.1. <i>Экспрессия вирусных генов</i>	59
2.3.7.2. <i>Влияние LTR на экспрессию клеточных генов</i>	60
2.3.7.3. <i>Защита от повторного заражения</i>	62
2.3.7.4. <i>Придание пластичности геному</i>	62
2.3.8. <i>Белковые факторы транскрипции, способные связываться с LTR</i>	63
<b>3. Материалы и методы</b>	<b>64</b>
3.1. <i>Материалы</i>	64
3.2. <i>Методы</i>	66
3.2.1. <i>Стандартные методики</i>	67
3.2.2. <i>Культивирование клеток</i>	67
3.2.3. <i>Трансфекция клеток</i>	68
3.2.4. <i>Трансфекция клеток электропорацией</i>	68
3.2.5. <i>Позитивно-негативная селекция последовательностей потенциальных инсуляторов</i>	69
3.2.6. <i>Получение вирусных частиц и инфицирование клеток</i>	69

3.2.7. Определение титра вирусных частиц	70
3.2.8. Селекция клеток на среде с генетицином G 418	70
3.2.9. Получение библиотеки фрагментов ДНК локуса <i>FXYD5-COX7A</i> хромосомы 19 человека	70
3.2.10. Радиоактивное мечение праймеров	71
3.2.11. Очистка меченой ДНК	71
3.2.12. Определение активности люциферазы	72
3.2.13. Фиксация и окрашивание клеток с помощью <i>Coomassie Blue</i>	73
3.2.14. Картирование последовательностей на геноме	73
3.2.15. Метод двумерного EMSA	73
3.2.16. Иммунопреципитация хроматина	74
3.2.17. Двумерный электрофорез с ультрафиолетовой сшивкой	74
3.2.18. Приготовление ДНК-аффинного носителя	75
3.2.19. Фракционирование ядерного экстракта на ДНК-аффинной колонке	77
3.2.20. Фракционирование ядерного экстракта методом аффинной элюции	77
3.2.21. Обработка гелей, гидролиз трипсином и экстракция пептидов	78
3.2.22. Анализ триптических пептидов при помощи масс-спектрометрии	79
<b>4. Результаты и обсуждение</b>	<b>80</b>
4.1. Стратегия идентификации энхансер-подобных элементов в протяженных областях сложных геномов	80
4.1.2. Экспериментальная идентификация и картирование энхансер-подобных элементов в локусе <i>FXYD5-COX7A1</i> хромосомы 19 человека	80
4.1.3. Создание векторных конструкций для селекции энхансер-подобных элементов	81
4.1.4. Клонирование и селекция энхансер-подобных элементов	82
4.1.5. Анализ отобранных потенциальных энхансерных фрагментов.	89
4.1.6. Способность отобранных фрагментов связываться с клеточными белками	91
4.1.7. Способность отобранных фрагментов активировать минимальный промотор с репортерным геном	93
4.1.8. Анализ расположения потенциальных энхансеров в геноме. Построение карты расположения энхансеров	94
4.1.9. Функциональный анализ активности потенциального энхансера "U2AF1L4"	97
4.2. Стратегия экспериментального поиска инсуляторов в протяженных областях сложных геномов	100
4.2.1. Система для экспериментального поиска инсуляторов	101
4.2.2. Анализ расположения инсуляторов в геноме	105
4.2.3. Функциональный анализ инсуляторов	108
4.3. Энхансер-блокирующая активность CTCF-связывающих последовательностей	112
4.4. Идентификация и картирование CTCF-связывающих последовательностей в глобиновом локусе птиц	119
4.5. Анализ функциональной архитектуры LTR семейства <i>HERV-K(HML-2)</i> и его взаимодействий с регуляторными компонентами клетки	126
4.5.1. Характеристика LTR <i>HERV-K</i>	126
4.5.2. Промоторная активность LTR	130
4.5.3. Негативный регуляторный элемент (НРЭ)	135
4.5.4. Энхансерная активность LTR	136
4.5.5. Клеточные белки, специфически связывающиеся с LTR <i>HERV-K</i>	138
4.5.6. Связывание белков другими участками LTR	147
4.5.7. Связывание белков с негативным регуляторным элементом (NRE)	149

<i>4.5.8.Идентификация белков, потенциально ответственных за энхансерную активность LTR HERV-K</i>	<i>151</i>
<i>4.5.9.Выделение и идентификация белков ERLBF1, 2 и 3, специфически связывающихся с LTR HERV-K</i>	<i>153</i>
<b>5. Выводы</b>	<b>166</b>
<b>6. Список цитированной литературы</b>	<b>168</b>
<b>7.Приложения</b>	<b>201</b>

## Список сокращений

- 3C (Chromosome Conformation Capture) – фиксация конформации хромосомы
- Abd-B – гомеозисный ген Abdominal-B
- apoB – аполипопротеин B
- BEAD (Blocking Element Alpha/Delta) – элемент, обладающий активностью блокирования энхансера
- BEAF (Boundary Element Associated Factor) – белковый фактор, связывающийся с последовательностью инсулятора
- BIV (Bovine Immunodeficiency Virus) – вирус иммунодефицита коров
- BX-C – гомеозисные гены комплекса *Bithorax*
- ChIP (Chromatin Immunoprecipitation) - иммунопреципитация хроматина
- CMV – цитомегаловирус
- CNS (Conserved Noncoding Sequence) – консервативная некодирующая последовательность
- CTCF (CCCTC-binding factor) - (фактор, связывающийся с последовательностью CCCTC)
- cHS4 (chicken  $\beta$ -globin DNase I Hypersensitive Site 4) – участок гена  $\beta$ -глобина кур, сверхчувствительный к расщеплению ДНКазой I, инсулятор
- DMR (Differentiated Methylation Region) – дифференциально-метилованная область
- DPE (Downstream Promoter Element) – участок ТАТА-несодержащего промотора, связывающийся с основным фактором транскрипции TFIID
- EGFP (Enhanced Green Fluorescent Protein) F64L/S65T – мутант зеленого флуоресцентного белка
- EMSA (Electrophoretic Mobility Shift Assay) - метод сдвига электрофоретической подвижности
- ERLBF - (Endogenous Retrovirus LTR Binding Factor) – эндогенный ретровирусный LTR-связывающий фактор
- ERV (Endogenous Retrovirus) – эндогенный ретровирус
- Fab (Frontabdominal) – инсулятор, расположенный между двумя доменами *iab*
- FISH (Fluorescent In Situ Hybridization) -гибридизация *in situ* с флуоресцентно меченым зондом
- FRT-сайты – сайты рекомбинации с участием FLP-рекомбиназы
- G418 - генетицин
- GANC – ганцикловир
- GDG – глутаминдегидрогеназа
- HEPES – N-2-гидроксиэтилпиперазин-N'-2-этилсульфоновая кислота
- HERV (Human Endogenous RetroVirus) - эндогенный ретровирус человека
- HFV (Human Foamy Virus) – спумавирус человека
- HIV (Human Immunodeficiency Virus) – вирус приобретенного иммунодефицита человека (ВИЧ)
- HML (Human Endogenous MMTV-like) – человеческие эндогенные MMTV-подобные ретровирусы – одно из названий HERV класса II

HRE (Hormone-Responsive Element) – последовательность ДНК, взаимодействующая с рецепторами глюкокортикоидных гормонов

HS-сайты, области (Hypersensitive sites) – последовательности ДНК, характеризующиеся повышенной чувствительностью к нуклеазам

HSV-tk (Herpes Simplex Virus type 1 Thymidine Kinase) – тимидинкиназа вируса простого герпеса типа 1

элементы)

HTDV (Human Teratocarcinoma-Derived Virus) – VLP, образуемые в тератокарциномах представителем семейства HERV-K(НМЛ-2)

HTLV (Human T-cell Leukemia Virus) – вирус Т-клеточного лейкоза человека

Iab (Infraabdominal) – регуляторный домен комплекса ВХ-С, отвечающий за экспрессию гена Abd-B в передней части брюшка у *Drosophila*

IAP (Intracisternal type A Particle) – ERV грызунов, образующие внутриклеточные VLP

IPTG - изопропил-бета-D-1-тиогалактопиранозид

IRES (Internal Ribosome Entry Site) – внутренний сайт посадки рибосом

JSRV (Jaagsiekte Sheep Retrovirus) – ретровирус овец, вызывающий карциному легких

LAD (Lamina Associated Domains) – участок связывания ДНК с ядерной ламинной

LAS (Loop Attachment Site) - участок прикрепления петель ДНК

LCR (Locus Control Region) - комплексный активатор транскрипции

LINE (Long Interspersed Element) - длинный диспергированный повтор

LTR (Long Terminal Repeat) - длинный концевой повтор

Мср (miscadestral pigmentation) – инсулятор, расположенный между двумя доменами iab

MIR (Mammalian Interspersed Repeat) – одно из семейств SINE

MMTV (Mouse Mammary Tumor Virus) – вирус, вызывающий рак молочной железы у мышей

Mod(mdg4) (modifier of mdg4)– ген *Drosophila melanogaster* или продукт этого гена, взаимодействующий с инсулятором gypsy

NPTII – неомицинфосфотрансфераза II

PBS (Phosphate Buffered Saline) - фосфатно-солевой буфер

PBS (Primer Binding Site) – сайт связывания тРНК-затравки

PGK - фосфоглицераткиназа

PMSF – фенилметилсульфотрифтормид

PS – парасегмент у *Drosophila melanogaster*

PsG (Polycomb Group) – гены, кодирующие белки-репрессоры гомеозисных генов

PTS (promoter targeting sequence) – последовательность в геноме *Drosophila*, модулирующая активность инсулятора Fab-8

R (Redundant region) – прямой повтор на концах ретровирусной геномной РНК

Rec (Regulator of Expression encoded by *Corf*) – (он же – cORF) дополнительный белок, кодируемый HERV-K(НМЛ-2), аналог Rev белка HIV

RO (Repeat Organizer) – область спейсера между генами рРНК, состоящая из повторов последовательностей в 100 и 35 п.н.

RTVL-H (Retrovirus-Like element H) – альтернативное название представителей семейства HERV-H

S/MAR (Scaffold/Matrix Attachment Region) - участок связывания ДНК с ядерным матриксом

SBP (scs Binding Protein) – белок, связывающийся с последовательностью scs

scs (specialized chromatin structure) – специализированные структуры хроматина

SDS - додецилсульфат натрия

SV40 (Semian Virus 40) – обезьяний вирус 40

Su(Hw) (Suppressor of Hairy-wing) – белок-супрессор

TAF (Transcription Activation Factor) – фактор активации транскрипции

TBE - трис-боратный буфер с добавлением ЭДТА

TBP (TATA-binding protein) – ТАТА-связывающий белок

TCR (T-cell receptor) – рецептор Т-клеток

TE - трис-ЭДТА буфер

Trl (Trithorax-like) – белок группы TrxG

TrxG – trithorax group – гены, кодирующие белки – активаторы гомеозисных генов

UAS (Upstream Activating Sequence) – дистальная активирующая последовательность

VLP (Virus-Like Particles) – вирус-подобные частицы

x-gal - 5-бromo-4-хлоро-3-индолил-бета-D-галактопиранозид

zw5 (zeste-white 5) – ген, кодирующий белок SBP

БСА - бычий сывороточный альбумин

Да – атомная единица массы

ДДС – додецилсульфат натрия

ДТТ – дитиотрейтол

кДа – килодальтон

НРЭ – негативный регуляторный элемент

ПААГ – полиакриламидный гель

ПИК – преинициаторный комплекс

ПЦР – полимеразная цепная реакция

п.о. - пар оснований

т.п.о. - тысяч пар оснований

ТХУ – трихлоруксусная кислота

ЭДТА - этилендиаминтетрауксусная кислота

ЯМ - ядерный матрикс

ЯЭ – ядерный экстракт

## ВВЕДЕНИЕ

Геном человека содержит множество функционально значимых элементов, таких как промоторы, энхансеры, сайленсеры, инсуляторы, сайты связывания гормонов и транскрипционных факторов. Однако определение одной только нуклеотидной последовательности генома человека не даёт исчерпывающей информации о расположении генов, их функциональном назначении и регуляции их экспрессии. Недавние исследования генома человека позволили идентифицировать большое число транскриптов и потенциальных регуляторных элементов, при этом регуляторных элементов в геноме существенно больше, чем генов, которые они контролируют. Кроме того регуляторные элементы нередко располагаются на больших расстояниях от генов, которые они регулируют. Регуляторные элементы часто подвергаются более быстрым эволюционным превращениям, поэтому их предсказание и поиск невозможен с использованием одних только методов биоинформатики, поскольку в первичной структуре ДНК большинства из них не выявлено высокой степени идентичности с уже известными регуляторами. Одно из актуальных направлений современной функциональной геномики - создание экспериментальных методов идентификации регуляторных элементов, а также разработка подходов к их функциональному анализу. В геномах млекопитающих существует примерно 20-25 тысяч кодирующих белки генов, что составляет один процент геномной ДНК (Lander et al., 2001), но транскрибируется гораздо большая часть генома (Carninci, 2006; Rosenbloom et al., 2009; Djebali et al., 2012). По теоретическим прогнозам в геномах эукариот функционируют сотни тысяч *цис*-регуляторных элементов. Консорциумом ENCODE в геноме человека идентифицированы ~400000 потенциальных энхансеров, и, возможно, это число может увеличиться до миллиона (Bernstein et al., 2012). Таким образом, значительная часть нашего генома, 25% и более, занята регуляторной информацией, что позволяет предположить, что типичный человеческий ген может регулироваться десятками энхансеров. Недавно появившиеся работы позволяют, таким образом,



количественно охарактеризовать наиболее значимые регуляторные элементы генома человека (de Laat et al., 2013).

В среднем регуляторный «арсенал» в геномах млекопитающих содержит:

около 20000 генов;

около  $10^6$  энхансеров (потенциальных регуляторных последовательностей);

545 “генных пустынь”;

до четырех энхансеров связываются с активным геном в клетке одного типа; средний размер петли, образующейся при взаимодействии энхансера и промотора составляет 120 т.п.о.;

самое большое расстояние энхансер-промоторного взаимодействия (SOX9, синдром Робена) составляет 1300 т.п.о.

Более простые существа, по-видимому, обладают меньшим числом энхансеров; например, *Drosophila* содержит 50,000-100,000 энхансеров (Arnold et al., 2013; Levine et al., 2014). Но и здесь количество регуляторных последовательностей значительно превосходит количество генов. Подчеркнем, что подобные оценки весьма приблизительны, так как в основном базируются на результатах полногеномного анализа, недостаточно проверенных при помощи прямых функциональных тестов.

Регуляторные области с похожим регуляторным потенциалом часто имеют лишь незначительное структурное сходство при сравнении как внутри одного генома так и между разными геномами. Это связано, по-видимому, с тем, что транскрипционный аппарат медленнее меняется в эволюции, чем те *цис*-регуляторные последовательности, которые он «обрабатывает» для осуществления регулируемой транскрипции. Проблема с пониманием механизмов, лежащих в основе эволюции *цис*-регуляторных элементов, в частности, консервации функций без консервации последовательности, это часть общей проблемы понимания функций регуляторных систем.

В свете приведенных выше данных очевидна необходимость дальнейшего изучения обнаруженных, а также поиска новых последовательностей, выполняющих регуляторные функции, таких, как промоторы, энхансеры,

инсуляторы, или участвующих в структурной организации ядра, таких как фрагменты, связывающиеся с ядерным матриксом (S/MAR-элементы).

Геном человека, также как и других многоклеточных, содержит громадное количество остатков ретроэлементов, многие из которых в эволюции приобрели важные для генома регуляторные функции. В частности, в геноме человека обнаруживают более полутора тысяч одиночных длинных концевых повторов (Long Terminal Repeat – LTR) эндогенных ретровирусов, которые представляют собой последовательности, содержащие несколько взаимодействующих регуляторных элементов - ТАТА-бокс, участки связывания ядерных факторов и сигнал полиаденилирования. Описано влияние LTR на регуляцию экспрессии близлежащих генов, ее уровень и тканеспецифичность. Внедрение LTR возле или внутри гена может вызывать альтернативный сплайсинг, терминацию транскрипции. LTR выявлены вблизи генов, кодирующих регуляторные белки, и могут изменять уровень их экспрессии, и, т.о., косвенно влиять на процессы развития различных патологий, включая злокачественную трансформацию клеток или аутоиммунные заболевания, а также могут являться факторами эволюции, придавая геному хозяина дополнительную изменчивость. По этой причине изучение эндогенных ретровирусов с их LTR позволило получить ряд важных новых данных о механизмах регуляции на уровне полного генома человека.

Методы масштабного поиска регуляторных элементов, которые используют в настоящее время, можно разделить на две группы: структурные и функциональные. Во многих из этих подходов используется один и тот же ключевой этап – получение высокообогащенных клонотек фрагментов ДНК, содержащих соответствующие регуляторные элементы. Структурные методы, например иммунопреципитация хроматина, основаны на структурных взаимодействиях ДНК с регуляторными белками или субклеточными структурами, а также на структурных особенностях и модификациях ДНК и хроматина (метилование ДНК, метилирование и ацетилирование гистонов). Применение этих методов ограничено нашими знаниями о транскрипционных факторах и их сайтах связывания с ДНК. Поэтому, определенное преимущество

получают функциональные подходы, использующие «репортерные гены», т.е. методы, основанные на анализе функциональной активности последовательностей ДНК. Эти методы позволяют идентифицировать фрагменты генома, обладающие специфической активностью.

Международный консорциум ENCODE (the ENCyclopedia Of DNA Elements, <http://www.genome.gov/10005107>) (ENCODE consortium 2004) разработал проект, предполагающий картирование всех *цис*-регуляторных элементов генома. На первом этапе проекта предполагалось идентифицировать и картировать эти элементы в сравнительно короткой (~1%) области генома человека. Впоследствии были разработаны и применены полногеномные методы (Rosenbloom et al., 2009; Djebali et al., 2012). В настоящее время благодаря усовершенствованным технологиям накоплены огромные массивы данных, которые нуждаются в более детальной проверке в функциональных тестах. К сожалению, картирование огромного числа функциональных элементов, выполненное с применением методов, основанных на секвенировании нового поколения, сопряжено с большим количеством ошибок. Так, например, применение ДНК микрочипов, в принципе, позволяет проводить полногеномный анализ экспрессии генов и расположения регуляторных элементов, но при этом сопровождается низкой воспроизводимостью результатов и другими недостатками (Shields, 2006).

Картирование такого большого количества функциональных последовательностей на уровне целого генома с учетом всевозможных типов клеток многоклеточных организмов связано с рядом технических трудностей. Полногеномные методы исследования, на наш взгляд, следует сочетать с исчерпывающим функциональным анализом отдельных протяженных сегментов генома с последующей интеграцией полученных данных в полногеномную карту регуляторных элементов

В представленной работе, будут описаны подходы, основанные на использовании именно таких методов экспериментального поиска и картирования регуляторных элементов в протяженных полигенных участках генома на основе

их функциональной активности. В качестве объекта исследования использовали хромосому 19 человека и глобиновый локус кур.

## **ОБЗОР ЛИТЕРАТУРЫ**

### **Регуляторные элементы генома**

#### **2.1. Инсуляторы**

Существование различных типов тканей высших эукариот определяется экспрессией генов в соответствующих типах клеток. Регулироваться могут различные стадии экспрессии генов, в том числе инициация и элонгация транскрипции, процессинг РНК, транспорт и трансляция матричных РНК и т.д.

Транскрипция регулируется при помощи целого ряда механизмов. Существуют механизмы, обеспечивающие конститутивную экспрессию генов домашнего хозяйства, специфические пути регуляции, обеспечивающие повышение или снижение экспрессии в ответ на внешние сигналы и, наконец, специальный класс регуляторных элементов, ответственный за полную инактивацию генов в ряде тканей.

Инициация транскрипции гена обеспечивается последовательностью корового промотора и связанного с ним основного транскрипционного комплекса. Экспрессия, инициированная только промотором, чаще всего протекает на минимальном уровне, для изменения которого необходимы дополнительные факторы. В частности, уровень экспрессии может быть повышен за счет функционирования энхансера, понижен или выключен посредством сайленсеров. Так как энхансеры способны активировать гены, расположенные от них на расстояниях до сотен тысяч пар нуклеотидов, должны существовать механизмы, препятствующие активации всех генов, доступных для определенного энхансера.

Транскрипция генов зависит и от действия различных сайленсеров, в том числе конденсированного хроматина. Конденсация хроматина также может распространяться на большие расстояния, поэтому должны существовать и механизмы, ограждающие экспрессирующиеся гены или локусы генов от влияний со стороны окружающего хроматина для поддержания специфического уровня экспрессии.

Последовательности ДНК, способные, находясь между промотором и энхансером, блокировать активирующее действие последнего *in vivo*, а также защищать экспрессию трансгена от эффекта положения, блокируя распространение конденсированного (неактивного) хроматина, принято называть инсуляторами (Cai and Levine, 1995; Kuhn et al., 2003; Gallagher et al., 2010; Grandchamp et al., 2011). Более чем за 25 лет исследований инсуляторы были выявлены в геномах различных организмов - *Saccharomyces cerevisia*, *Drosophila*, *Sea urchin*, курицы и млекопитающих. Среди идентифицированных элементов такие, как инсулятор *gypsy* и пограничные *scs/scs'*-элементы локуса генов теплового шока *hsp 70 Drosophila*, *ars*- и *sns*-элементы морского ежа, инсулятор 5'HS4, расположенный в локусе  $\beta$ -глобиновых генов кур, BEAD-элемент локуса  $\alpha/\delta$  рецептора Т-клеток (TCR) и DMR-элемент локуса *Igf2/H19* человека и др. (West et al., 2002; D'Apolito et al., 2009; Wang et al., 2011; Cena et al., 2013).

Проведенные полногеномные исследования сайтов связывания инсуляторного белка CTCF показали, что геном человека содержит более 30000 последовательностей, которые, возможно, играют роль инсуляторов (Kim et al., 2007; Xie et al., 2007). Подобно энхансерам, инсуляторы выявляются предпочтительно в межгенном хроматине, хотя встречаются и внутри генов (в интронах) (Mukhopadhyay et al., 2004; Akopov et al., 2006; Barkess and West, 2012). Следует отметить, что термин «инсулятор» применяют для обозначения как энхансер-блокирующих элементов, так и пограничных элементов генома, разделяющих участки хроматина с различной структурой или активностью. Отдельные инсуляторные элементы обладают обеими активностями, но, поскольку иногда эти активности могут быть разделены, можно предположить, что за них ответственны различные (возможно, находящиеся по соседству) участки инсулятора (Burgess-Beusse et al., 2002; Recillas-Targa et al., 2002; Singh and Srivastava, 2008; Chai et al., 2013).

### 2.1.1. Функции инсуляторов

Исследование изменений уровня экспрессии гена *yellow* у *Drosophila*, вызванного вставкой ретротранспозона *gypsy* в регуляторную область этого гена, явилось одним из первых аргументов в пользу существования элементов, способных блокировать энхансеры. Вставка *gypsy* в 5'-область гена *yellow* между дистальными энхансерами *wing* и *body cuticle* и проксимальным энхансером *larval* приводила к блокированию действия дистальных энхансеров на промотор, но не влияла на активность проксимального энхансера (Geyer et al., 1990; Geyer and Corces, 1992).

Позднее были выявлены инсуляторные последовательности, ограничивающие локус генов теплового шока в геноме *Drosophila melanogaster*. Их назвали специализированными структурами хроматина (*scs*<sup>2</sup>/*scs* – specialized chromatin structures). В трансгенных мухах *scs* или *scs*<sup>2</sup>-элемент, клонированный между промотором и энхансером, подавлял активацию транскрипции репортерного гена (Kellum and Schedl, 1992). Впоследствии анализ блокирования энхансера стал основным тестом на наличие инсуляторной активности.

При помощи этого теста было выявлено несколько инсуляторов у других беспозвоночных и позвоночных. К ним относятся *sns*-инсуляторы (*silencing nucleoprotein structure*), расположенные в кластере ранних генов гистонов морского ежа (Palla et al., 1997; D'Apolito et al., 2009); RO-область межгенного спейсера генов рРНК у *Xenopus* (Robinett et al., 1997), BEAD-элемент (*blocking element alpha/delta-1*) локуса  $\alpha/\delta$  рецептора Т-клеток (TCR) человека, DMR-элемент (*Differentially Methylated Region*) локуса *Igf2/H19* человека (Zhong and Krangel, 1997), и, наконец, 5'- и 3'-границы локуса  $\beta$ -глобиновых генов кур (Zhan et al., 2001).

$\beta$ -глобиновый локус состоит из пяти глобиновых генов  $\epsilon$ ,  $\gamma^G$ ,  $\gamma^A$ ,  $\delta$  и  $\beta$ , расположенных на хромосоме в последовательности их экспрессии при онтогенезе. В 5'-области от гена  $\epsilon$  расположены 7 сайтов гиперчувствительности к ДНК-азе I (HS). Пять ближайших к гену  $\epsilon$  HS функционируют как локус-контролирующая последовательность (LCR, *locus control region*), совмещающая функции инсулятора и энхансера. Далее за глобиновыми генами расположен

единичный 3'HS. 5'- и 3'HS функционируют как инсуляторы (Harju et al., 2002; Bender et al., 2012). С 5'-стороны от глобиновых генов за областью конденсированного хроматина расположен ген фолатного рецептора (FR), а с 3'-стороны – ген обонятельного рецептора (OR). Все три группы генов характеризуются различными программами экспрессии (Bell et al., 1999; Amouyal, 2010; Barrow et al., 2014).

Первым инсулятором, обнаруженным в дрожжах *S. cerevisiae*, был UAS<sub>upg</sub>-элемент (upstream activation site), представляющий собой участок промотора двух генов *TEF1* и *TEF2*. Эти гены относятся к большому семейству генов, кодирующих компоненты трансляционной машины. Инсулятор обеспечивает устойчивость репортерного гена к сайленсингу, вызванному вставкой экспрессирующей конструкции в НМ-локус (локус генов спаривания гомоталлических штаммов дрожжей), хроматин которого находится в конденсированном состоянии (Bi and Broach, 1999; Cena et al., 2013).

Перечисленные выше инсуляторы, кроме ретротранспозона *gypsy* и UAS<sub>upg</sub>, были выявлены на границах отдельных генов или генных локусов, что соответствует их предполагаемой роли в поддержании хромосомных доменов и защите их от влияния окружающего хроматина (Bell et al., 2001; Molto et al., 2009).

Инсуляторы, ограничивая влияние окружающего конденсированного хроматина на экспрессию генов в локусе, могут служить границами отдельных функциональных доменов. Scs'/scs-инсуляторы *Drosophila* расположены на границах пуфа локуса генов теплового шока *hsp70* и фланкирующих его областей конденсированного хроматина (Kellum and Schedl, 1992; Ghirlando et al., 2012; Kirkland et al., 2012). В 5'-области локуса β-глобиновых генов кур находится регуляторная область DCR (dominant control region), содержащая четыре участка гиперчувствительности к ДНК-азе I (HS). Три из них, ближайшие к β-глобиновым генам, входят в состав активатора транскрипции LCR (locus control region), ответственного за эритроид-специфичную экспрессию β-глобиновых генов. Содержащий четвертый HS участок DCR длиной 1,2 т.п.н. является β-глобиновым



инсулятором, и, в отличие от первых трех, является гиперчувствительным во всех типах тканей (Chung et al., 1997; Furlan-Magaril et al., 2010). С 5'-конца от  $\beta$ -глобинового инсулятора находится область конденсированного хроматина длиной 16 т.п.н., которая на протяжении всего эритропоэза находится в деацетилованном и метилированном состоянии. Таким образом, куриный инсулятор 5'HS4 расположен на границе гетерохроматина и активного хроматина, и функция его заключается в блокировании влияния гетерохроматина на экспрессию глобиновых генов (Prioleau et al., 1999; Litt et al., 2001; Okamura et al., 2013).

Гистоновые гены морского ежа организованы в кластеры, содержащие по одной копии гена каждого из пяти гистонов. В каждом кластере транскрипция гистона H2A регулируется энхансер-подобным модулятором, расположенным на 5'-конце гена H2A. На 3'-конце этого гена расположена последовательность *sns*, роль которой состоит в блокировании действия модулятора на промоторы других генов кластера (Palla et al., 1997; Cavalieri et al., 2009).

Тандемные повторы генов рибосомных РНК эукариот разделены межгенными спейсерами. Значительная часть спейсера у *Xenopus laevis* содержит повторы длиной 60 и 81 п.н. Эта область является энхансером, активирующим транскрипцию генов 40S рРНК, расположенных с 3'-конца. С 5'-стороны от энхансера располагается участок ДНК, названный организатором повторов RO (Repeat Organizer), который состоит из повторов длиной 35 и 100 п.н. RO является инсулятором, ограничивающим действие энхансера 60/81 на промоторы генов рРНК, расположенных с 5'-стороны от него (Robinett et al., 1997).

Локус TCR  $\alpha/\delta$  человека содержит несколько сегментов, которые кодируют две разные цепи рецепторов Т-клеток. В границах локуса расположены два энхансера  $E_\delta$  и  $E_\alpha$ , которые активируют V(D)G-рекомбинацию между генами соответствующих сегментов на разных стадиях созревания Т-клеток. Перестройка сегментов  $\alpha$ -генов сменяет перестройку сегментов  $\delta$ -генов. Гены  $J_\alpha$  и  $C_\alpha$  расположены возле энхансера  $E_\delta$ , но перестройка  $\alpha$ -генов не активируется его влиянием, поскольку между ними расположен инсулятор BEAD-1. Таким

образом, BEAD-1 обеспечивает независимую экспрессию генных сегментов TCR- $\alpha$  и TCR- $\delta$  (Zhong and Krangel, 1997; Bell et al., 2001; Ramezani et al., 2008).

Особый класс инсуляторов был обнаружен среди S/MAR-элементов. S/MAR-элементы (Scaffold/Matrix Attachment Region) – это последовательности ДНК, способные связываться с компонентами ядерного белкового матрикса или скэффолда *in vitro*. S/MAR-элементы преимущественно локализируются в области центромер, в интронах и межгенных участках. Их функции *in vivo* не установлены, однако наиболее признанным является предположение о том, что S/MAR-элементы служат для закрепления хроматиновых петель интерфазных и митотических хромосом на белковом матриксе или скэффолде (Chernov et al., 2004). Часто S/MAR-элементы обнаруживаются на концах генных локусов, поэтому была выдвинута гипотеза, что они могут проявлять свойства инсуляторов (Glazkov, 1995; Benabdellah et al., 2014). Для подтверждения этой гипотезы S/MAR-элементы были проверены на способность блокировать энхансер и защищать от эффекта положения. Первым выявленным S/MAR-элементом, проявляющим свойства инсулятора, стал 5'-MAR гена лизоцима цыпленка. Среди S/MAR-элементов человека были исследованы 3'MAR AroB, MAR-элемент локуса  $\alpha 1$ -антитрипсина, 5'-SAR гена интерферона и другие S/MARs (Namciu et al., 1998; Zhan et al., 2001; Sass et al., 2005). Было показано, что инсулятор *gypsy* также содержит последовательности, связывающиеся с ядерным матриксом в клетках *Drosophila*, мыши и человека (Nabirochkin et al., 1998).

Для проявления их энхансер-блокирующей активности инсуляторы должны располагаться между энхансером и промотором гена, иными словами, функционирование инсулятора зависит от взаимного расположения энхансера, промотора и самого инсулятора (Kellum and Schedl, 1992; Chung et al., 1993). Инсуляторы, подобно энхансерам, способны функционировать, располагаясь на различных расстояниях от промотора (Dorsett, 1993). Активность инсуляторов может зависеть (Bi and Broach, 1999; Gruzdeva et al., 2005) или не зависеть (Zhou et al., 1996; Yannaki et al., 2002; Belozarov et al., 2003) от их ориентации относительно промотора. Кроме того, в случае, когда между энхансером и

промотором находится более одного инсулятора, их энхансер-блокирующая активность может снижаться (Mongelard and Corces, 2001; Kuhn and Geyer, 2003; Maksimenko et al., 2008; Maksimenko and Georgiev, 2014), при этом взаимная нейтрализация инсуляторов может зависеть и от их ориентации друг относительно друга (Kyrchanova et al., 2007).

В условиях транзientной трансфекции использование одной копии инсулятора для блокирования энхансера оказывается недостаточным в силу двунаправленного действия энхансера в кольцевой плазмиде и для полного блокирования энхансера необходимо, чтобы он был фланкирован копиями инсулятора с обеих сторон (Recillas-Targa et al., 1999; Parnell and Geyer, 2000; Maksimenko and Georgiev, 2014). Линеаризация плазмид позволяет преодолеть двунаправленное действие энхансера и, использование одного инсулятора оказывается достаточным для блокирования энхансера в таких конструкциях (Recillas-Targa et al., 1999). Эти данные подтверждают более ранние исследования, в которых показано, что инсулятор не инактивирует ни энхансер, ни промотор, а лишь препятствует их взаимодействию друг с другом. Энхансер, ограниченный с одной стороны инсулятором, способен активировать промоторы, расположенные с другой стороны. Этот феномен справедлив и для промоторов: будучи ограниченными инсулятором с одной стороны, они могут быть активированы энхансерами, расположенными с другой стороны. То есть, инсулятор является барьером, придающим четкую направленность действию энхансера (Cai and Levine, 1995; Ghirlando et al., 2012).

Таким образом, инсуляторы подразделяют геном на домены, содержащие экспрессирующиеся при сходных условиях гены, а также, вместе с генами, промоторами и энхансерами образуют многоуровневую сеть регуляции транскрипционной активности. Многообразие функций инсуляторов свидетельствует об их важности для правильного функционирования генома, а также о необходимости их изучения для более глубокого понимания механизмов, регулирующих экспрессию генов.

### 2.1.2. Белки, обеспечивающие действие инсуляторов

Для исследованных инсуляторов не было обнаружено гомологии в последовательностях их ДНК. У некоторых инсуляторов основная часть активности была сосредоточена в участке, который был назван коровым элементом. Чаще всего коровый элемент содержит один или несколько гиперчувствительных к действию ДНК-азы I сайтов и служит участком связывания специфического белка или группы белков. Влияние белков на свойства инсулятора было подробно изучено для инсулятора в составе ретротранспозона *gypsy*.

Инсулятор *gypsy* представляет собой последовательность в 350 п.н., расположенную на 5'-конце транскрибируемого, но нетранслируемого участка ретротранспозона *gypsy*. Инсуляторная активность сосредоточена в участке, который содержит 12 повторов последовательности PuPuTTGCATACCPu, разделенных AT-богатыми последовательностями. С каждым из 12 повторов с различной степенью сродства при помощи цинковых пальцев связывается белок Su(Hw) (Suppressor of Hairy-wing). С-концевой домен белка Su(Hw) и участок "лейциновой молнии" взаимодействуют с С-концевым участком белка Mod(*mdg4*), продуктом гена *modifier of mdg4*. Около 20 изоформ этого белка, образующихся в результате альтернативного сплайсинга, выявляются в клетке. Основная форма белка необходима для белок-белковых взаимодействий и представляет собой продукт трансляции транскрипта длиной 2,2 т.п.н., содержащий N-концевой домен ВТВ из 115 аминокислотных остатков. Домен ВТВ обеспечивает димеризацию и олигомеризацию белка Mod(*mdg4*)<sub>2.2</sub>, что делает возможным его связывание с Su(Hw). Mod(*mdg4*)<sub>2.2</sub>, в отличие от других ВТВ-содержащих белков, не способен связываться с ДНК, однако выявляется примерно в 500 участках политенной хромосомы. Примерно в 200 сайтах его локализация совпадает с присутствием белка Su(Hw). В остальных 300 сайтах Mod(*mdg4*)<sub>2.2</sub>, по-видимому, взаимодействует с другими белками, фиксирующими его на хромосоме. Эти области не содержат копий

ретротранспозона *gypsy*, но, очевидно, являются эндогенными инсуляторами, сходными с *gypsy* (Gdula et al., 1996; Ghosh et al., 2001; Schwartz et al., 2012).

Мутации в гене *mod(mdg4)*, инактивирующие его белковый продукт, вызывают репрессию всех энхансеров регуляторной области репортерных генов *yellow* и *white* независимо от их позиции относительно инсулятора *gypsy*. Следовательно, *gypsy* в отсутствие Mod(mdg4) обладает активностью сайленсера (Cai and Levine, 1997). Таким образом, белок Mod(mdg4) придает определенную направленность действию сайленсера (Gdula et al., 1996). В результате неполной сборки белкового комплекса инсулятора *scs'*/*scs*-элементы также проявляли свойства сайленсера (Parnell and Geyer, 2000; Golovnin et al., 2013).

Участок гиперчувствительности к ДНК-азе I *scs*-элемента включает последовательность ДНК в 24 п.н., содержащую палиндром ТТСГСГСГАА, и является сайтом связывания белка SBP (*scs* binding protein), кодируемого геном *zeste-white* (*zw5*). Поскольку белок SBP отвечает лишь за часть активности *scs*-инсулятора, выдвинуто предположение, что он способен образовывать комплексы с другими еще не идентифицированными белками (Gaszner et al., 1999; Hens et al., 2011).

Коровый элемент *scs'*-инсулятора длиной 515 п.н. содержит два HS-сайта. Они разделены АТ-богатым спейсером длиной 200 п.н. Каждый из сайтов содержит несколько мотивов СГАТА, которые ответственны за большую часть активности *scs'* и служат сайтом узнавания для двух белков BEAF-32А и 32В (*boundary element associated factor*). Белки эти различаются только N-концевыми доменами, которые связываются со специфическими последовательностями ДНК. С-концевой домен обеспечивает белок-белковые взаимодействия. Так как белки различаются своими ДНК-связывающими доменами, то различные их комбинации в составе тримера позволяют ему связывать разные последовательности ДНК, отличные от *scs'*/*scs*-элементов (Hart et al., 1997; Gurudatta et al., 2012). С последовательностью *scs'* связывается гетеротример BEAF-(32В)<sub>2</sub>32А. Белки BEAF-32А и 32В были выявлены во многих локусах политенных хромосом, часто на границе пуфов. Были идентифицированы еще две

последовательности СВЕ76 и СВЕ28, содержащие повторы мотива СGАТA и обладающие свойствами инсуляторов. Таким образом, можно выдвинуть предположение о наличии целого класса инсуляторов, содержащих СGАТA-мотивы, которые служат сайтами связывания белков ВЕАF-32А и 32В (Cuvier et al., 1998; Cuartero et al., 2014).

Активность  $\beta$ -глобинового инсулятора кур сосредоточена в основном в коровом элементе длиной 250 п.н., который содержит сайт 5'HS4, и зависит от числа его копий. В частности, две его копии по активности соответствуют полноразмерному инсулятору длиной 1,2 т.п.н. Последовательность корового элемента отличается высоким содержанием остатков G и C и повышенной плотностью CpG-динуклеотидов. Фрагмент длиной 500 п.н., включающий коровый элемент, находится в деметилированном состоянии (Chung et al., 1997). Фрагмент коровой ДНК, содержащий 5'HS4-сайт, является участком связывания белка CTCF (Bell et al., 1999; Spencer et al., 2011) (подробно про CTCF см. ниже).

UAS<sub>trg</sub> элементы промоторов генов TEF содержат несколько сайтов связывания белков Rap1 (Bi and Broach, 1999; Chakraborty et al., 2011).

Промотор *eve* (*even-skipped*) *Drosophila* содержит последовательность GAGA, которая обладает инсуляторной активностью. Важным обстоятельством в функционировании GAGA в качестве инсулятора является связывание с белком Trl (Trithorax-Like protein) группы TrxG (Ohtsuki and Levine, 1998; Erokhin et al., 2011; Fujioka et al., 2013).

Таким образом, важным фактором, обеспечивающим инсуляторную активность всех рассмотренных элементов является их взаимодействие с группами специфических белков и присутствие в них соответствующих сайтов связывания этих белков. Наличие разных инсуляторов, взаимодействующих с одним белком (или одной группой белков), позволяет выявлять новые инсуляторные элементы по связыванию с известными белками-инсуляторами или предсказывать новые инсуляторы методом компьютерного анализа последовательностей. Этот метод был успешно применен к белкам ВЕАF-32А, -32В и CTCF (Cuvier et al., 1998; Mukhopadhyay et al., 2004).

### 2.1.3. Механизмы блокирования энхансеров

Исследование активности инсуляторов в различных системах дало основание для создания ряда моделей, описывающих механизм действия этих регуляторных элементов.

#### 2.1.3.1. Модель ловушки (*promoter decoy*)

Согласно модели «ловушки» или «приманки», белковые комплексы промотора и энхансера взаимодействуют напрямую друг с другом за счет выпетливания фрагмента ДНК, расположенного между ними. Белковый комплекс, связывающийся с инсулятором, имитирует транскрипционный аппарат промотора, поэтому энхансер будет взаимодействовать с инсулятором. Таким образом, инсулятор является конкурентом, ингибирующим процесс активации транскрипции и выступает в роли «ловушки» для энхансера (Zhan et al., 2001; Chopra et al., 2009; Core and Lis, 2009). Примером такого механизма является функционирование ретротранспозона *gypsy Drosophila*, (Geyer et al., 1986). Его последовательность содержит 12 сайтов связывания белка Su(Hw), образующего комплекс с двумя другими белками CP190 и Mod(mdg4) (Pai et al., 2004). Этот комплекс, в свою очередь, посредством дополнительных белков взаимодействует с белками ядерной ламины (Capelson and Corces, 2005). В результате несколько ретротранспозонов *gypsy*, расположенных в разных участках хромосомы, сближаются, образуя кластеры или инсуляторные тельца, которые располагаются на периферии ядра (Gerasimova et al., 2000). В процессе блокирования энхансера расположение инсуляторных телец по периферии ядра не является обязательным (Xu et al., 2004), в отличие от Su(Hw)-зависимого процесса образования петель хроматина. Последовательности инсуляторов в инсуляторных тельцах взаимодействуют друг с другом посредством белок-белковых взаимодействий, и энхансер с промотором оказываются в разных петлях хроматина, что, по всей видимости, и является причиной неактивного состояния гена.

Похожая ситуация наблюдается и при образовании петлевого домена в локусе генов белков теплового шока Hsp70 *Drosophila*, фланкированного элементами *scs* и *scs'*. Элементы *scs'* и *scs*, в отличие от *gypsy*, взаимодействуют с изоформами А и В белка BEAF32, а также Zw5. Взаимодействуя друг с другом *in vivo*, эти белки участвуют в формировании петельно-доменной структуры хромосом (Blanton et al., 2003).

Роль инсуляторного белка у позвоночных выполняет CTCF, многофункциональный фактор, ортолог которого (dCTCF) был обнаружен и у *Drosophila*. Интересно, что, как и белок Su(Hw), dCTCF, также взаимодействует с белком CP190, и также образует инсуляторные тельца (Gerasimova et al., 2007). Эти данные указывают на существование у *Drosophila* нескольких типов инсуляторов, отличающихся друг от друга белковыми компонентами комплексов. Подобно инсуляторным белкам *Drosophila*, одной из функций белка CTCF позвоночных является установление и поддержание функциональных доменов хроматина за счет взаимодействия с белками ядерного матрикса (Yusufzai et al., 2004; Kurukuti et al., 2006; Hou et al., 2008; Naumova et al., 2013 ).

#### 2.1.3.2. Модель отслеживания (*tracking*)

Механизм, при котором инсуляторы вовлекаются во взаимодействие с энхансер-связывающими белками с образованием нового белкового комплекса, напоминающего по структуре комплекс транскрипционных факторов, связанных с промотором, по видимому, является не единственным, т.к. энхансер, заблокированный с одной стороны инсулятором, способен активировать промоторы, расположенные с другой стороны (Cai and Levine, 1995). Для изоляции энхансера необходимо два инсулятора, фланкирующих либо энхансер, либо промотор гена (Cai and Shen, 2001; Muravyova et al., 2001; Farrell et al., 2002; Singh and Srivastava, 2008). Однако в экспериментах с использованием линейризованных плазмид было продемонстрировано, что для проявления инсулирующего эффекта достаточно одного инсулятора, расположенного между энхансером и промотором (Recillas-Targa et al., 1999). На этом основании была



сформулирована модель «отслеживания» (tracking), согласно которой комплекс активаторных белков, собранный на энхансере, перемещается вдоль ДНК, пока не достигнет целевого промотора. Комплекс останавливается, если встречается с инсулятором. Инсулятор выступает в роли физического барьера, препятствующего передвижению комплекса и, таким образом, нарушает взаимодействие с промотором. В качестве аргументов в пользу этой модели можно привести данные, полученные при транзientной трансфекции конструкций, содержащих  $\beta$ -глобиновый инсулятор (Recillas-Targa et al., 1999; Zhan et al., 2001; Kolovos et al., 2012). В качестве компонента подобного комплекса может выступать РНК-полимераза II (West and Fraser, 2005). Взаимодействуя с энхансером в составе преинициаторного комплекса, фермент транскрибирует промежуточную ДНК до встречи с промотором (Ashe et al., 1997; Kong et al., 1997). Активация глобинового гена  $\epsilon$ , входящего в локус  $\beta$ -глобиновых генов человека, и расположенного в 5'-области локуса, осуществляется подобным же образом. Локус-контролирующая последовательность (LCR), расположенная с 5'-стороны от  $\epsilon$ -глобинового гена, содержит пять сайтов гиперчувствительности к ДНК-азе I, один из которых (HS2) является энхансером. Энхансер HS2 способен активировать транскрипцию глобиновых генов в эритроидных клетках, располагаясь на значительном расстоянии (10-50 тыс. п.о.) от генов локуса, при этом активация сопровождается синтезом коротких некодирующих полиаденилированных РНК с ДНК, расположенной между энхансером и промотором (Tuan et al., 1992; Ling et al., 2005). Инсерция CTCF-зависимого инсулятора cHS4 между энхансером и промотором приводит к значительному снижению активности промотора, уменьшению количества коротких мРНК, синтезируемых с 3'-стороны от инсулятора, и увеличению количества ТВР и РНК-полимеразы II в области инсулятора. Методом 3С была показана колокализация последовательностей энхансера и инсулятора, что подтверждает существование механизма отслеживания в данной системе, конечной стадией которого является образование петли из промежуточной ДНК (Zhu et al., 2007; Kolovos et al., 2012). Любопытно,

что CTCF способен взаимодействовать с большой субъединицей РНК-полимеразы II посредством С-концевого домена как *in vitro*, так и *in vivo*. Кроме того, полногеномный анализ расположения сайтов связывания CTCF и РНК-полимеразы II показал, что значительная часть их колокализуется. Вероятно, эти последовательности представляют собой особый тип CTCF-зависимых инсуляторов, функционирующих подобно куриному инсулятору cHS4, защищая гены от влияния энхансеров, работающих по tracking-механизму (Chernukhin et al., 2007; Shrimali et al., 2012).

#### 2.1.4. Нейтрализация эффекта положения

Впервые способность защищать интегрированный в геном репортерный ген как от негативного, так и позитивного влияния эффекта положения было обнаружено при изучении *scs* и *scs'*-элементов *Drosophila* (Bell et al., 2001). При трансформации клеток *Drosophila* конструкцией, содержащей ген *white*, ответственный за проявление цвета глаз, были получены мухи с глазами различной пигментации, что являлось проявлением зависимости экспрессии трансгена от эффекта положения. В ситуации, когда ген *white* в конструкции был окружен последовательностями *scs* и *scs'*, разнообразие в окраске глаз пропадало, и эффект положения подавлялся (Kellum and Schedl, 1991; Chetverina et al., 2008).

Дальнейшие эксперименты выявили сходные свойства и у инсулятора *gypsy*, который, как выяснилось, был способен защищать от эффекта положения как трансген, так и точку начала репликации (Lu and Tower, 1997). Нейтрализация эффекта положения была продемонстрирована и для инсулятора 5'HS4 (Chung et al., 1993; Pikaart et al., 1998; Seo et al., 2010). Было показано, что элемент 5'HS4 способен защищать стабильно интегрированный ген *IL2R* (interleukin 2 receptor), имеющий высокую вариабельность экспрессии, от репрессорного влияния хроматина. Не фланкированный инсуляторами репортерный ген *IL2R* терял транскрипционную активность: методом флуоресцентной цитометрии было показано значительное снижение экспрессии *IL2R* и выявлено, что репортерный ген при этом терял эпигенетические маркеры, характерные для активного

хроматина, такие как доступность для нуклеаз, гипометилирование ДНК и гиперацетилирование гистонов. В то же время ген, окруженный инсуляторами 5'HS4, оставался доступным для нуклеаз и содержал гиперацетилированные гистоны. Способность блокировать эффект положения была продемонстрирована также для А-элемента гена лизоцима цыпленка и 3' MAR-элемента гена аполипопротеина (ApoB) человека (Namciu et al., 1998; Brasset and Vaury, 2005).

Кроме того, было показано, что выключение стабильно интегрированного репортерного гена не является результатом инактивации транскрипционных факторов - длительная инкубация клеток с репрессированным интегрированным геном с добавлением ингибиторов деацетилаз гистонов приводит к восстановлению работы гена (Chen et al., 1994; Kumar et al., 2014). Такие реактивированные ингибиторами деацетилаз гены остаются активными даже в случае трансфекции ими клеточной популяции, в которой имеется репрессированный стабильно интегрированный репортерный ген, поскольку в данном случае не задействованы транскрипционные факторы.

Элемент UAS<sub>rgg</sub> (upstream activation site) *S. cerevisiae* является промотором двух генов *TEF1* и *TEF2*, которые относятся к большому семейству генов, кодирующих компоненты трансляционного аппарата. Устойчивость репортерного гена к сайленсингу, вызываемому вставкой экспрессирующей конструкции в HM-локус (локус генов спаривания гомоталлических штаммов дрожжей), хроматин которого находится в конденсированном состоянии, обеспечивал инсулятор (Vi and Broach, 1999).

Большинство инсуляторов, кроме ретротранспозона *gypsy* и UAS<sub>rgg</sub>, были выявлены на границах генов или генных локусов, что соответствует их предполагаемой роли в поддержании хромосомных доменов и защите их от влияния окружающего хроматина (Bell et al., 2001).

### 2.1.5. Инсуляторы и S/MAR-элементы

Доменная организация обусловлена ассоциацией хроматина с нуклеопротеиновыми комплексами, которые в свою очередь «заякоряют»

связанные участки хромосом в ядерном матриксе. Кроме структурной, петельным доменам приписывают и функциональную роль, которая заключается в том, что активность генов, находящихся в одном домене, не зависит от регуляторных элементов других доменов. Важную роль в образовании петельных доменов хроматина играют участки прикрепления ДНК к внутреннему каркасу ядра (ядерному матриксу), называемые scaffold/matrix attachment regions (S/MARs). S/MARs – нуклеотидные последовательности размером от 200 до 1000 п.о., проявляющие повышенное сродство к ядерному матриксу и образующие, за счет связывания их с матриксом, основания петельных доменов (Chernov et al., 2004). В большинстве S/MARs обнаруживаются А+Т-богатые районы различной протяженности. Предположительно, именно в этих областях происходит взаимодействие S/MARs с ядерным матриксом (Акоров et al., 1997).

Было показано, что наличие S/MAR-элементов по концам интегрированного гена усиливает его транскрипцию и делает ее мало зависимой от эффекта положения (Chernov et al., 2004), т.е. элементы обладают активностью, подобной инсуляторам. Показано, что в отдельных случаях активности S/MAR-элементов и инсуляторов локализуются в одном фрагменте ДНК. Например, инсулятор *gypsy* способен взаимодействовать с ядерным матриксом не только *Drosophila*, но и мышей и человека (Nabirochkin et al., 1998).

Принимая во внимание эту особенность инсулятора *gypsy*, была предложена гипотеза о формировании несколькими инсуляторами и связанными с ними белками «телец», ассоциированных с ядерной ламиной (Gerasimova et al., 2000). В пользу этой гипотезы свидетельствуют несколько доводов. Последовательность, в норме локализованная в центре ядра, начинала перемещаться к его периферии, когда внутри нее оказывался *gypsy*. Более того, интеграция *gypsy* обуславливала сближение последовательностей, в норме расположенных на хромосоме на определенном расстоянии. Наконец, белки Su(Hw) и Mod(mdg4), связывающиеся с *gypsy*, преимущественно локализованы на периферии ядра, где было обнаружено около 25 мест локализации этих белков (Gerasimova et al., 2000).

Впрочем, для инсуляторов *scs*, *scs'* и 5'HS4 не было выявлено матрикс-связывающей активности. Более того, не все S/MAR-элементы определяют границы доменов хроматина, в отдельных случаях S/MARs выявляли и внутри экспрессирующихся генов (Zhan et al., 2001). Следовательно, вопрос о связи S/MAR-элементов и инсуляторов остается открытым и требует идентификации и сравнительного анализа большего числа этих элементов.

#### 2.1.6. Регуляция функционирования инсуляторов

Активность инсуляторов, как и прочих *цис*-регуляторных элементов, определяется специфическими белками, взаимодействующими с ДНК. Влияние этих белков на активность инсуляторов свидетельствует о том, что действие таких элементов не абсолютно, а, напротив, отличается избирательностью и поддается регулированию. Поэтому функционирование того или иного инсулятора зависит от многих факторов, например, от пост-трансляционных модификаций инсуляторных белков, от белок-белковых взаимодействий, а также от доступности последовательности инсулятора в составе хроматина для этих белков, что, в свою очередь, определяется как ковалентными модификациями гистонов, так и самой ДНК (Bushey et al., 2008; Majocchi et al., 2013).

В частности, блокирование энхансера происходит с разной эффективностью в зависимости от места интеграции в геном человеческих клеток K-562 конструкции, содержащей 5'HS4-инсулятор кур и репортерный ген  *$\beta$ -geo* (кодирует белок с активностями  $\beta$ -галактозидазы и неомицинофосфотрансферазы) с его регуляторной последовательностью (Walters et al., 1999). Следовательно, инсулятор может подвергаться влиянию окружающего хроматина (Zhan et al., 2001; Gurudatta et al., 2012). Эффективность блокирования энхансера инсулятором *gypsy* у *Drosophila* зависит от силы энхансера (Scott et al., 1999). Приведенные примеры свидетельствуют о том, что действие инсулятора не абсолютно, и в зависимости от условий инсулятор может оказывать различное по силе действие, что позволяет говорить о величине активности или силе того или иного инсулятора. У позвоночных регуляция активности инсуляторов в первую очередь

зависит от белка CTCF. В настоящее время описано несколько типов регуляции функционирования инсуляторов.

Например, связывание белка CTCF позвоночных может быть заблокировано с помощью метилирования его сайта связывания, то есть, метилирование ДНК инсулятора препятствует связыванию с ним белков, сопряженных с его активностью, и инсулятор оказывается инактивированным (Filipova et al., 2002).

Локус *Igf2/H19* мыши и человека включает сайты связывания белка CTCF (4 и 7 сайтов, соответственно). Район, содержащий эти сайты и обладающий активностью инсулятора, назван DMR (Differently Methylated Region). На материнском аллеле DMR выполняет функцию блокирования энхансера, активирующего транскрипцию гена *Igf2*, поэтому *Igf2* не экспрессируется с материнского аллеля. На отцовской хромосоме DMR находится в метилированном состоянии, что препятствует связыванию белка CTCF с областью DMR, и приводит к инактивации инсулятора. В результате *Igf2* экспрессируется с отцовского аллеля (Bell et al., 2001; Engel et al., 2008).

Блокирование энхансера может быть нарушено и из-за модификации связывающихся с инсуляторами белков с помощью различных кофакторов. Показано, что у позвоночных активность элементов, связывающих CTCF, может модулироваться расположенными рядом сайтами связывания рецепторов тиреоидного гормона (TR, thyroid hormone receptor). Подобная форма регуляции может быть реализована при регуляции конститутивных инсуляторов в различных тканях во время развития и дифференцировки. Активность белков и связывающих их инсуляторов может быть изменена в результате посттрансляционной модификации, которая может приводить как к стабилизации контакта белков с сайтами их связывания в составе инсулятора, так и к полной потере их функциональной активности (West and Fraser, 2005).

Эксперименты с трансгенными *Drosophila* указывают на существование и другого пути регуляции функционирования инсуляторов. Две копии инсулятора *gypsy* нейтрализуют энхансер-блокирующую активность друг друга (Muravyova et al., 2001). В то же время, введение в плазмидную конструкцию двух инсуляторов

5'HS4 повышает степень блокирования энхансера (Bell and Felsenfeld, 1999; Ramezani et al., 2008).

Таким образом, различные последовательности в разных тест-системах могут проявлять или утрачивать инсуляторную активность в зависимости от используемых комбинаций окружающих регуляторных элементов, их видо- и тканеспецифичности и т.д.

### *2.1.7. Локус-контролирующие последовательности*

Под локус-контролирующими областями (LCR, locus control region) подразумевают последовательности, способные усиливать экспрессию связанного с ними гена или локуса генов и поддерживать ее на уровне, соответствующем определенным физиологическим условиям для данной ткани. Было показано, что наличие LCR является необходимым и достаточным для полной, независимой от геномного окружения экспрессии трансгена; это позволяет выдвинуть предположение о важной роли LCR в установлении «точки отсчета» эухроматиновых доменов (Bulger et al., 2003; Ho et al., 2008). Функцией усиления транскрипции обладают также энхансеры, однако, в отличие от LCR, усиленная за счет энхансера экспрессия трансгена все же существенно зависит от сайта интеграции, т. е. продолжает испытывать влияние хромосомного окружения. Согласно этим представлениям классический LCR должен содержать в своем составе энхансер и инсулятор, наличие которых действительно было показано для большинства LCR (Mutskov et al., 2002; Gaussin et al., 2011).

Впервые LCR был идентифицирован в  $\beta$ -глобиновом локусе человека при исследовании структуры глобиновых генов у больных  $\beta$ -талласемией. Делеционный анализ выявил наличие гиперчувствительного к ДНК-азе I сайта, фланкирующего в 5'-области  $\beta$ -глобиновые гены, известного теперь как  $\beta$ -глобиновый LCR, и являющегося критическим для нормальной экспрессии этого локуса. В экспериментах по получению трансгенных мышей была точно определена его функциональная активность (Grosveld et al., 1987; Li et al., 1999). Позже было идентифицировано и описано множество LCR-последовательностей

для целого ряда генов и локусов генов позвоночных и всех их отличала способность поддерживать экспрессию генов на физиологическом уровне как в естественных условиях, так и в составе трансгенных конструкций. Эти данные дают основания предположить, что регуляция экспрессии генов, определяющая тканеспецифичность и нормальное эмбриональное развитие является комплексной и определяется не только предназначенными для каждого отдельного гена *cis*-регуляторными элементами, но влиянием дальнодействующей регуляции на более высоком уровне, определяющим независимую экспрессию целых локусов генов (Dixon et al., 2012).

## **2.2. Белок CTCF и его роль в крупномасштабной регуляции активности генома позвоночных**

CTCF (также известный как CCCTC-связывающий фактор) - повсеместно экспрессирующийся ядерный белок позвоночных с многочисленными функциями. CTCF был впервые обнаружен как белок, специфически узнающий последовательность из трех регулярно расположенных повторов CCCTC, расположенную в 200 п.н. с 5'-стороны от участка начала транскрипции гена *MYC* курицы (Lobanenkov and Gudvin, 1989; Lobanenkov et al., 1990; Holwerda and de Laat, 2013) и связывающийся с лизоцимным сайленсером кур (Baniahmad et al., 1990, Burcin et al., 1997). Белок CTCF содержит три домена, одним из которых является ДНК-связывающий домен с 11 цинковыми пальцами.

CTCF является эволюционно консервативным "от дрозофилы до человека" (Gray and Coates, 2005; Moon et al., 2005; Hore et al., 2008; Ni et al., 2012), 83-85% аминокислотных остатков полноразмерного белка идентичны у человека, кролика, курицы и лягушки. Идентичность достигает 100 % в областях, содержащих цинковые пальцы (Burke et al., 2002).

Ген CTCF экспрессируется в различных тканях. В геноме человека он находится на хромосоме 16q22.1 в области потери гетерозиготности и является предполагаемым опухолевым супрессором рака молочной и предстательной желез (Filippova et al., 1998; Green et al., 2009). Для CTCF характерно



динамическое распределение по клеточным компартментам и зависимость этого распределения от клеточного цикла. В интерфазе CTCF локализуется в ядре и в основном исключен из ядрышка. Во время митоза, особенно в анафазе и метафазе, CTCF ассоциирован с centrosомой (Zhang et al., 2004). CTCF фосфорилируется протеинкиназой CK2 *in vivo* (Klenova et al., 2001), а также поли(АДФ)-рибозилируется (см. обзор (Caiafa and Zlatanova, 2009)). Кроме того, CTCF способен к активации поли(АДФ-рибоза)-полимеразы 1 (Guastafierro et al., 2008). Недавно были опубликованы данные о сумоилировании (MacPherson et al., 2009; Kitchen and Schoenherr, 2010; Fiorentino and Giordano, 2011) и десумоилировании CTCF (Wang et al., 2012).

Используя различные комбинации цинковых пальцев, CTCF связывается с различными участками цис-регуляторных последовательностей (Renda et al., 2007). База данных сайтов связывания CTCF содержит более 10 млн сайтов (Ziebarth et al., 2012). Связывание CTCF с целевым участком ДНК может приводить как к активации, так и к репрессии транскрипции гена (Filippova, 2008; Ong and Corces, 2014).

Связывание CTCF с ДНК чувствительно к метилированию, за исключением тех случаев, когда сайт связывания не содержит CpG динуклеотидов. Связывание с CTCF может защищать эти сайты от метилирования (см. обзоры (Filippova, 2008; Wang et al., 2012)), что предполагает возможность участия CTCF в эпигенетической регуляции (Bell and Felsenfeld, 2000; Hark et al., 2000; Chao et al., 2002; Kemp et al., 2014) и в инактивации X-хромосомы (Filippova et al., 2005; Donohoe et al., 2007; Chadwick, 2008; Spencer et al., 2011; Atchison, 2014). Связывание CTCF с ДНК может регулироваться положением нуклеосом (Kanduri et al., 2002; Fu et al., 2008; Ohlsson et al., 2009).

### *2.2.1. CTCF как многофункциональный регулятор*

#### *2.2.1.1. CTCF и гены, ассоциированные с раком*

Расположение гена CTCF человека в области потери гетерозиготности при раке молочной и предстательной желез (Filippova et al., 1998) позволило

предположить, что ген CTCF может являться кандидатным опухолевым супрессором. Эту гипотезу подтверждали опухоль-специфичная перегруппировка экзонов *CTCF* (Filippova et al., 1998) и наличие нескольких связанных с раком ген-специфических мутаций, препятствующих связыванию CTCF (Filippova et al., 2002). Однако исследование уровня мРНК CTCF при карциноме молочной железы не показало существенных и связанных с опухолью изменений экспрессии, и поэтому CTCF вряд ли может служить геном-супрессором (Rakha et al., 2004).

Кроме того, было показано, что при раке молочной железы уровень белка CTCF был повышен в опухоли по сравнению с нормальными аналогами, что также свидетельствовало в пользу того, что CTCF не является геном-супрессором опухоли (Docquier et al., 2005). В то же время авторы этой работы выдвинули гипотезу о том, что индукция CTCF может быть связана с устойчивостью клеток рака молочной железы к апоптозу. Это предположение было подтверждено демонстрацией того, что оверэкспрессия CTCF частично защищает клетки от апоптоза, индуцированного проапоптотическим белком Bax (Docquier et al., 2005). Таким образом, вышеприведенные эксперименты указывают на то, что CTCF ведет себя скорее как регулятор других опухолеассоциированных генов, чем как классический опухолевый супрессор. Это соответствует наблюдениям, что CTCF связывается с различными регуляторными генами, среди них известный онкоген *MYC* (Lobanenkov et al., 1990) ген-супрессор опухоли *RBI* (De La Rosa-Velazquez et al., 2007) и другие гены (см. (Filippova, 2008; Gomes and Espinosa, 2010; Batlle-Lopez et al., 2013; Zampieri et al., 2014)).

#### *2.2.1.2. CTCF и развитие: регулятор основных регуляторов?*

Полный нокаут гена CTCF вызывает раннюю эмбриональную гибель мышей, что указывает на его участие в раннем развитии (Fedoriw et al., 2004; Splinter et al., 2006; Herold et al., 2012). Однако частичная делеция гена CTCF в эмбриональных клетках печени (Splinter et al., 2006) и тимоцитах (Heath et al., 2008) не вызывала смерти клеток, хотя затрагивала их рост и регуляцию. При

нокдауне гена CTCF с использованием антисмысловых конструкций наблюдалось ингибирование дифференцировки клеток K-562 (Torrano et al., 2005; Manavathi et al., 2011) и апоптотическая гибель клеток в клеточных линиях рака молочной железы (Docquier et al., 2005). Снижение экспрессии CTCF, обусловленное siRNA, приводило к снижению экспрессии генов МНС II класса (Majumder et al., 2006). Приведенные выше данные свидетельствуют о многофункциональной природе CTCF, и в то же самое время показывают, что CTCF не существенен для роста клеток в культуре или пролиферации опухолевых клеток.

CTCF регулирует экспрессию гена *Paxb*, кодирующего высококонсервативный транскрипционный фактор, который имеет важное значение в развитии центральной нервной системы, включая развитие глаза (Osumi et al., 2008). Нокдаун CTCF у трансгенных мышей усиливает транскрипцию *Paxb*, тогда как суперэкспрессия CTCF подавляет его транскрипцию, возможно за счет разобщения промотора *Paxb* и его энхансера (Li et al., 2004; Fang et al., 2014).

Истощение материнского CTCF в ооцитах мыши приводит к нарушению регуляции транскрипции многих генов, мейотическим дефектам в яйцеклетке и эмбрионе. Авторы пришли к заключению, что CTCF играет существенную роль в раннем эмбриональном развитии (Wan et al., 2008; Moore et al., 2012).

Приведенные выше данные кажутся противоречивыми, однако следует принимать во внимание существенное различие между генным нокаутом, ведущим к полному исчезновению CTCF из всех клеток организма и условным генным нокаутом или нокдауном транскрипта в соматической ткани. В последних случаях может существовать депо стабильных белков, которое, вероятно, имеется в случае CTCF. Это депо может быть достаточным для поддержания нескольких клеточных делений, например, в случае материнского CTCF в яйцеклетке (Wan et al., 2008; Singh et al., 2012). К сожалению, данные, позволяющие оценить содержание белка CTCF в различных клетках, в настоящее время отсутствуют.

В целом приведенные результаты указывают на существенную роль CTCF в развитии млекопитающих. По-видимому, функции CTCF включают (но не

ограничиваются) регуляцию других регуляторов. Как CTCF осуществляет эту регуляцию, остается пока неизвестным.

## *2.2.2. CTCF и функционирование генома*

### *2.2.2.1. Геномное распределение сайтов связывания CTCF*

Публикация последовательности геномов человека и других многоклеточных открыли возможность для анализа распределения CTCF-связывающих участков в пределах геномов. При помощи иммунопреципитации хроматина в геноме мыши было идентифицировано более 200 CTCF-связывающих фрагментов ДНК. Значительная часть этих фрагментов проявляла инсулятор-подобные свойства (Mukhopadhyay et al., 2004; Prickett et al., 2013). Связывающие CTCF участки были найдены в межгенной ДНК, в регуляторных областях генов и в интронах. Число идентифицированных сайтов было, однако, недостаточным, чтобы сделать достоверное заключение об их распределении в геноме. Была предпринята попытка картирования всех связывающих участков CTCF в области человеческого генома длиной 1 млн п.н. (Vetchinova et al., 2006). Идентифицированы десять участков связывания, что позволило оценить общее количество связывающих CTCF участков в человеческом геноме приблизительно в 30000. Связывающие CTCF участки были найдены внутри интронов, а также в повторяющихся элементах, в особенности Alu. Методом иммунопреципитации хроматина с последующей идентификацией при помощи микрочипов (ChIP-on-chip) было идентифицировано 13804 CTCF связывающих участка в геноме человеческих фибробластов IMR90 (Kim et al., 2007). Их распределение коррелировало с положением генов в геноме. 46% сайтов были найдены в межгенных последовательностях, 20% около точек начала транскрипции, и 34% внутри генов, главным образом в интронах. Примечательно, что более 67% CTCF связывающих участков, занятых белками в клетках IMR90, были также заняты белком в клетках U937, что предполагает конститутивный характер связывания CTCF большинством его сайтов (Kim et al., 2007).

Сходные результаты были получены в работе (Barski et al., 2007), где распределение CTCF-связывающих участков в геноме изучали методом иммунопреципитации хроматина в комбинации с масштабным параллельным секвенированием (ChIP-seq). Были идентифицированы более чем 20000 связывающих CTCF участков в CD4+ Т-клетках. Приблизительно 40% сайтов располагались в межгенных областях, 30% внутри генов и 30% вблизи сайтов начала транскрипции. Более подробный анализ был выполнен позже для трех типов клеток (CD4+ Т-клетки, Jurkat и HeLa) (Cuddapah et al., 2009; Taslim et al., 2012) и показал, что приблизительно 49-56% сайтов связывания CTCF расположены в межгенном пространстве, 33% - в интронах, и 3-4% - в экзонах. Наибольшей тканеспецифичностью отличались сайты связывания CTCF, расположенные вблизи промоторов. Авторы продемонстрировали, что CTCF предпочтительно занимает границы неактивных областей хроматина, обогащенных триметилированным по K27 гистон H3 (Cuddapah et al., 2009).

Чен и др. (Chen et al., 2008) использовали ChIP -Seq подход для идентификации и картирования сайтов связывания CTCF и других транскрипционных факторов в полном геноме мыши. Число сайтов связывания CTCF в геноме было оценено как 40000, что несколько превышает предыдущие оценки (см. выше), и значительно (в 2-50 раз) превышает число сайтов связывания других изученных транскрипционных факторов. Было также показано (Barski et al., 2007), что CTCF предпочтительно связывается с областями, содержащими метилированный гистон H3. CTCF-связывающие области были обогащены H3K4 (все статусы метилирования) и H3K9me1, но не H3K9me2 и H3K9me3.

Кроме того, было обнаружено обогащение CTCF-связывающих участков вариантом гистона H2A.Z, что согласовалось с ранее полученными данными о предпочтительном расположении H2A.Z на границах доменов хроматина дрожжей (Meneghini et al., 2003).

Сайты гиперчувствительности к ДНК-азе I также во многих случаях связывают CTCF (Xi et al., 2007). База данных сайтов связывания CTCF, CTCFBSDB

(<http://insulatordb.uthsc.edu/>), содержит почти 15 миллионов CTCF-связывающих последовательностей в 10 видах (Ziebarth et al., 2012).

#### 2.2.2.2. CTCF опосредует взаимодействие между регуляторными сайтами на больших расстояниях

Применение различных вариантов техники Chromosome Conformation Capture (3C) (Simonis et al., 2007) позволило идентифицировать петли хроматина, образующиеся с участием CTCF (Williams and Flavell, 2008; Zlatanova and Caiafa, 2009). Внутри ответственного за импринтинг локуса *Igf2/H19* мыши петли с участием CTCF могут образовываться между областью контроля импринтинга (ICR) и дифференциально метилированной областью (DMR) (Murrell et al., 2004; Kurukuti et al., 2006; Yoon et al., 2007; Nativio et al., 2011). Кроме того, CTCF опосредует взаимодействие ICR одного аллеля локуса *Igf2/H19* на хромосоме 7 с аллелем локуса *Wsb1/Nf1* на хромосоме 11 (Ling et al., 2006; Thomas et al., 2011).

CTCF может участвовать в образовании петли несколькими способами: а) образуя димеры или олигомеры, способные взаимодействовать с двумя или более различными областями ДНК; б) взаимодействуя с другими белками, способными связываться с ДНК или белками, и в) только за счет взаимодействий CTCF-ДНК в том случае, если одна молекула CTCF может связывать отдаленные области ДНК.

#### 2.2.3. Взаимодействие CTCF с другими белками

Данные исследований *in vitro* и двугибридного анализа свидетельствуют о возможности образования димеров и даже мультимеров CTCF (Pant et al., 2004; Yusufzai et al., 2004; Rousseau et al., 2013), однако существование ди- и олигомеров CTCF *in vivo* пока не доказано.

Методом аффинного фракционирования ядерного экстракта на колонке с иммобилизованным CTCF выявило его связывание с фактором транскрипции YB-1, связывающим Y-box, однако физиологическая роль такого взаимодействия пока не вполне понятна (Chernukhin et al., 2000).

Другим примером CTCF-связывающего белка является содержащий цинковые пальцы фактор транскрипции Kaiso, принадлежащий семейству POZ. Взаимодействия между Kaiso и CTCF были выявлены с помощью двугибридной системы и коиммунопреципитации (Defossez et al., 2005). Позднее было показано, что фактор Kaiso узнает не занятые CTCF последовательности в тех случаях, когда связывание CTCF теряется из-за метилирования ДНК (De La Rosa-Velazquez et al., 2007). Недавно было продемонстрировано, что CTCF может связываться с фактором транскрипции YY1 и трансактивировать *Tsix* (антисмысловый регулятор гена *Xist*) и, таким образом, играет ключевую роль в инактивации X-хромосомы (Donohoe et al., 2007; Spencer et al., 2011; Atchison, 2014). Это первый пример доказанного функционально значимого взаимодействия CTCF с другим белком. Предполагается, что взаимодействие CTCF-YY1 может быть важно и для регуляции других областей контроля импринтинга (ICR) (Kim, 2008; Zhang et al., 2011; Atchison, 2014).

Также по данным исследований *in vitro* и двугибридной системы, CTCF взаимодействует с белковым компонентом Suz12 раннеэмбрионального репрессивного комплекса PRC2, который отвечает за метилирование лизина 27 гистона H3, что в результате приводит к репрессии хроматина. Возможно, этот механизм отвечает за подавление материнских промоторов гена *Igf2* (Li et al., 2008).

CTCF может взаимодействовать и активировать аутомодификацию поли-ADP-рибозилполимеразы 1, которая, в свою очередь, может влиять на активность цитозин(С5)-ДНК-метилтрансферазы 1 (DNMT1) и, следовательно, на структуру хроматина (Reale et al., 2005; Guastafierro et al., 2008; Zampieri et al., 2011).

Было показано, что в клетках HeLa большая субъединица РНК-полимеразы II (Pol II LS) ко-иммунопреципитирует с CTCF, и что эпитопы CTCF и Pol II LS колокализуются с бета-глобиновым инсулятором *in vivo* (Chernukhin et al., 2007; Kang and Lieberman, 2011; Li et al., 2014).

Иммунопреципитация *in vitro* и двугибридный анализ показали, что сходная с SNF2 хромодоменная геликаза CHD8 взаимодействует через свою С-концевую

область с цинк-фингерным доменом CTCF. По данным иммунопреципитации хроматина CHD8 обнаруживается в некоторых сайтах связывания CTCF *in vivo*. Авторы предположили возможное участие CHD8 в CTCF-зависимом инсулировании (Ishihara et al., 2006; Ong and Corces, 2014).

Была показана колокализация CTCF *in vivo* с нуклеофосмином, многофункциональным кислым ядрышковым белком, который участвует в регуляции клеточного цикла и пролиферации (Okuwaki, 2008; Ong and Corces, 2014). Так как нуклеофосмин является известным компонентом ядерного матрикса (Mattern et al., 1996), было высказано предположение, что взаимодействие между этим белком и CTCF приводит к появлению CTCF во фракции матрикса (Yusufzai et al., 2004; Ouboussad et al., 2013).

Интересно, что большинство известных сейчас партнеров CTCF являются либо транскрипционными регуляторами (YB-1, Kaiso, YY1), либо связаны с регуляцией других клеточных функций. Однако стоит отметить, что в большинстве случаев данные о CTCF-белковых взаимодействиях были получены косвенным или *in vitro* методами, такими как chromatin pull-down и двугибридный анализ. Прямой *in vivo* анализ CTCF взаимодействий ограничен из-за отсутствия адекватных методов.

#### 2.2.3.1. CTCF и когезиновый комплекс

Когезин представляют собой консервативный мульти-белковый комплекс, основная функция которого состоит в том, чтобы удерживать вместе сестринские хроматиды в ходе фаз S и G2 клеточного цикла для обеспечения надлежащего расхождения хромосом. Роль когезина в регуляции генов рассмотрена в обзорах (Onn et al., 2008; Peters et al., 2008). При изучении геномного распределения сайтов связывания когезина в хроматине человека и мыши было установлено, что в 60-70% случаев когезин и CTCF колокализованы, и нокдаун CTCF нарушает расположение когезина. Важно отметить, что обратное неверно, и уменьшение количества когезина в клетке не влияет существенно на распределение CTCF в хроматине. Нокдаун когезина в некоторой степени ингибирует активность



инсулятора бета-глобинового локуса кур при транзientной трансфекции. Авторы пришли к выводу, что локализация когезина во многом определяется CTCF, а когезин вовлечен в инсуляторную функцию (Parelho et al., 2008; Stedman et al., 2008; Wendt et al., 2008; DeMare et al., 2013; Prickett et al., 2013; Ong and Corces, 2014). Предполагается, что CTCF взаимодействует с когезиновым комплексом через субъединицу Scc3/SA (Rubio et al., 2008; Nasmyth, 2011).

#### 2.2.4. CTCF и повторяющиеся элементы генома

Сайты связывания CTCF были обнаружены в повторяющихся элементах, таких как (GT)<sub>22</sub>(GA)<sub>15</sub> микросателлита A9 во втором интроне гена *HLA-DRB1(\*):0401* (Arnold et al., 2000), в тринуклеотидных повторах, связывание CTCF с которыми влияло на их стабильность (Libby et al., 2008; Brouwer et al., 2013; Darrow and Chadwick, 2014), и в Alu-элементах (Vetchinova et al., 2006; Ruiz-Narvaez and Campos, 2008; Schmidt et al., 2012; Prendergast et al., 2014; Schwalie et al., 2014). Также было показано, что повторяющиеся последовательности B2 значительно перепредставлены в CTCF-связывающих районах генома мыши (Bourque et al., 2008).

Интересно, что некоторые элементы Alu могут функционировать как инсуляторы (Willoughby et al., 2000; Ruiz-Narvaez and Campos, 2008). Alu элементы способны к транспозиции, и их перемещение в пределах генома вместе с сайтами связывания CTCF может привести к реорганизации структуры домена, аналогично тому, как мы ранее предположили для S/MARs-содержащих ретроэлементов (Chernov et al., 2004).

#### 2.2.5. Регуляция гена CTCF

Несмотря на очевидно важные функции CTCF в регуляции клеточных процессов, очень мало известно о регуляции самого гена CTCF. Транскрипция гена CTCF умеренно (в 2-3 раза) индуцировалась в эпителиальных клетках роговицы кролика эпидермальным фактором роста, а в Rb-клетках человека активация сывороткой приводила к подавлению транскрипции *Paxb* (Li and Lu,

2005; Li et al., 2006; Fang et al., 2014). Уровень экспрессии CTCF у мышей понижается с возрастом (Fu et al., 2008). Обработка клеток JAr хориокарциномы человека литием приводила к увеличению в 2-3 раза содержания мРНК CTCF (Roberts et al., 2007).

Промотор CTCF не содержит ТАТА бокса и содержит CpG островок, сайт связывания высококонсервативного фактора транскрипции YY1 и потенциальные сайты связывания для GATA-1 и p53 (Klenova et al., 1998; Pugacheva et al., 2006; Whitfield et al., 2012). В пределах промотора гена CTCF курицы были найдены несколько элементов, характерных для генов, регулируемых в клеточном цикле (Klenova et al., 1998).

#### *2.2.5.1. Возможные механизмы регуляторной активности CTCF*

CTCF может влиять на транскрипцию генов несколькими возможными способами:

1) Регуляция активности инсулятора. Связывание CTCF регулирует доступность промотора для энхансера либо через связывание CTCF с сайтом связывания, расположенным между энхансером и промотором, либо при освобождении конститутивно занятого сайта связывания из-за модификации (например, фосфорилирования) CTCF или модификации (метилования) его сайта связывания. По-видимому, этот вариант используется при регуляции транскрипции локуса *H19-Igf2* (Nativio et al., 2011), а также в регуляции промотора *Paxb* (Li and Lu, 2005; Engel et al., 2008). В последнем случае регуляция основана на способности CTCF подвергаться фосфорилированию (Klenova et al., ; Deng et al., 2012 ), poly-ADP-рибозилированию (Yu et al., 2004; Caiafa and Zlatanova, 2009; Zampieri et al., 2011; Beneke, 2012) или сумоилированию (Yu et al., 2004; Caiafa and Zlatanova, 2009; MacPherson et al., 2009; Kitchen and Schoenherr, 2010; Fiorentino and Giordano, 2011; Zampieri et al., 2011; Beneke, 2012).

2) Регуляция на уровне полного генома. CTCF может служить якорным белком, связывание которого с его сайтом позволяет ДНК функциональному

(например, транскрипционному) комплексу связаться с ДНК. Большое число молекул CTCF и его сайтов связывания делает возможным регуляцию многих процессов с участием его ДНК-связывающего домена и способности взаимодействия с другими белками. CTCF работает как конститутивный демаркационный инструмент генома/хроматина. В зависимости от геномного контекста и наличия различных факторов, появляющихся в ответ на внешние или внутренние клеточные сигналы, одиночный CTCF рекрутирует определенный набор белков для выполнения определенной функции. Таким образом, колокализация CTCF с РНК-полимеразой II, а также его способность связывать различные транскрипционные факторы, может играть определенную роль в регуляции транскрипции.

3) Формирование функционального домена и инсуляторная функция. Наконец, способность CTCF к формированию ди- или мультимеров и взаимодействовать с другими ДНК-связывающими белками предполагает его возможное участие в изгибании ДНК в петлю и направлении энхансеров на соответствующие промоторы. Кроме того, две молекулы CTCF в кооперации с другими белками могут изолировать энхансер-промоторную систему от внешнего регуляторного вмешательства.

Хотя все эти возможности кажутся вполне разумными, они безусловно нуждаются в дополнительной экспериментальной проверке.

CTCF является ярким примером многофункционального регуляторного фактора. Он участвует в активации и репрессии активности генов, инсуляции хроматина, формировании хроматиновых петель, инактивации X-хромосомы, позиционировании когезинового комплекса и т.д. CTCF регулирует ряд регуляторных генов, включая онкогены и супрессоры опухолевого роста, а также гены, регулирующие развитие и дифференцировку. CTCF является объектом нескольких важных пост-трансляционных модификаций. CTCF не является существенным для пролиферации отдельной клетки, но очень важен для развития многоклеточных организмов. Количество в геноме сайтов связывания CTCF превышает это количество для большинства других транскрипционных факторов.

В то же время, механизмы, лежащие в основе функционирования CTCF, остаются малоизученными. В частности функционирование CTCF как компонента инсуляторного комплекса и его участие в формировании петельных доменов предполагает белок-белковые взаимодействия с его участием, но количество обнаруженных белков-партнеров, взаимодействующих с CTCF, невелико. Таким образом, поиск белков и/или белковых комплексов, взаимодействующих с CTCF *in vivo*, должно внести существенный вклад в углубление понимания его роли в клеточных процессах.

## **2.3. Ретроэлементы**

### *2.3.1. Краткая история исследования ретровирусов и ретроэлементов*

Впервые ретровирусы (Avian Leukosis Virus - ALV) обнаружили в 1908 году датские ученые В. Эллерман и О. Бэнг, изучавшие эритробластоз кур, а через три года Фрэнсис Раус выделил вирус саркомы кур (Rous Sarcoma Virus - RSV). Первый ретровирус млекопитающих - вирус, ассоциированный с раком молочных желез мышей (Mouse Mammary Tumor Virus - MMTV) был открыт Дж. Биттнером в 1936 году, а существование эндогенных ретровирусов (ERV) было показано примерно двадцать лет спустя у цыплят и мышей (Lindeskog et al., 1999). Позже многочисленные последовательности ERV были обнаружены в геномах других позвоночных (Martin et al., 1997).

На рубеже 70-х годов Темин и Балтимор открывают явление обратной транскрипции (Baltimore, 1970; Temin and Mizutani, 1970), а вскоре группа исследователей обнаруживает длинные концевые повторы (LTR) (Hsu et al., 1978; Shank et al., 1978). После этого увеличивается число работ, посвященных поискам ретровирусов человека. Были обнаружены вирус Т-клеточного лейкоза (Human T-cell Leukemia Virus – HTLV) (Poiesz et al., 1980), эндогенные ретровирусы человека (HERV) (Bonner et al., 1982; Martin et al., 1981) и вирус приобретенного иммунодефицита человека (Human Immunodeficiency Virus - HIV) (Barre-Sinoussi et al., 1983).

Хотя Магер и Фримен указывали на наличие в человеческом геноме родственных ретровирусам последовательностей еще в начале 70-х годов (Mager and Freeman, 1995), первый представитель эндогенных ретровирусов человека семейства К (HERV-K10) был обнаружен Оно с соавторами при гибридизации фрагмента гена *pol* из генома цитоплазматических вирусоподобных частиц (Intracisternal A-Particle – IAP) сирийского хомячка с ДНК генома человека (Оно, 1986). Позже было показано, что наблюдаемые в электронный микроскоп вирусоподобные частицы в клеточных линиях тератокарцином кодируются представителем того же семейства эндогенных ретровирусов человека HERV-K (HML-2) (Boller et al., 1993).

### 2.3.2. Жизненный цикл и строение ретровирусов, синтез LTR

Вопреки многообразию вызываемых ими заболеваний и различным видам хозяев, практически все ретровирусы имеют схожую структуру вириона, геномную организацию и репликативный цикл. По своей структуре эндогенные ретровирусы подобны провирусам экзогенных ретровирусов, от которых, вероятно, и произошли (Temin, 1980; Sverdlov, 1998; Weiss and Stoye, 2013). Несмотря на то, что в результате мутаций в ходе эволюции они потеряли способность к репродукции, некоторые из них, вероятно, кодируют вирусоподобные частицы, например представители семейства HERV-K (HML-2) продуцируют HTDV (Human Teratocarcinoma-derived Virus) (Lower et al., 1993; Lower et al., 1996). Структурно эти частицы походят на вирионы экзогенных ретровирусов.

Геном ретровирусов состоит из двух молекул плюс-РНК размером 7-10 тыс. нт, которые связаны между собой нековалентно в 5'-области. Перенос генетической информации из одноцепочечной РНК в двухцепочечную ДНК осуществляется ферментами – обратной транскриптазой (Reverse Transcriptase - RT), интегразой (Integrase - IN) и протеазой (Protease – PR) – продуктами гена *pol* (Polymerase), необходимыми также для ковалентного соединения вирусной и геномной ДНК, то есть для образования провируса (Leis et al., 1988; Urnovitz and

Murphy, 1996). На расстоянии 100-500 нт от 5'-конца генома ретровируса локализована нуклеотидная последовательность (Primer Binding Site – PBS), с которой связываются 18 3'-концевых нуклеотидов молекулы тРНК, ассоциированной с геномной РНК ретровируса. Эта молекула играет главную роль в репликации, поскольку является праймером для начала синтеза ДНК обратной транскриптазой (Temin, 1981).

По концам геномной РНК расположены нетранслируемые последовательности, обозначаемые RU5 и U3R, состоящие из прямого концевой повтора (Redundant - R), который ограничивает 5'-специфичную (5'-Unique, U5) и 3'-специфичную (3'-Unique, U3) области, на соответствующих концах генома. Порядок распределения генов *gag-pol-env* одинаков для всех ретровирусов. Перечисленные гены кодируют белки нуклеокапсида, ферменты и белки оболочки, соответственно. При этом у *сложных* ретровирусов, таких, как HTLV, HIV и др., набор генов дополнен регулируемыми вирусную экспрессию генами. Наличие последовательностей, гомологичных таким генам, было описано для провирусов HERV-K(HML-2) (Magin et al., 1999).

В жизненном цикле ретровирусов выделяют две стадии. Первая стадия проходит в отсутствие экспрессии вирусных генов и осуществляется вирусными белками, входящими в состав вириона. При попадании в клетку происходит частичное высвобождение вирусной РНК, после этого обратная транскриптаза в капсиде синтезирует двунитевую ДНК-копию ретровируса, которая интегрируется в геном клетки хозяина при помощи интегразы. Провирус, синтезированный подобным образом, содержит вирусные гены *gag-pol-env*, и, вследствие специфических особенностей механизма обратной транскрипции и интеграции, содержит на концах LTR, отсутствующие в геномной РНК и фланкирован 4-6 нуклеотидными прямыми повторами хозяйской ДНК (Temin, 1981; Varmus, 1982; Horie et al., 2010).

Второй этап репликативного цикла выполняется системами репликации, транскрипции и модификации клетки-хозяина. Встроенный провирус подвергается репликации и далее передается по наследству, как фрагмент ДНК

хромосомы хозяина. Наличие LTR, которые содержат такие регуляторные последовательности, как промоторные и энхансерные элементы (U3-область), сигнал и сайт полиаденилирования (R-область), активный провирус способен транскрибироваться. После этого его РНК претерпевает посттранскрипционные модификации и сплайсинг и далее транслируется с образованием вирусных белков-предшественников. Несплайсированная вирусная РНК укладывается в нуклеокапсид, который затем отделяется от клетки, образуя вирусную частицу (Lower et al., 1996; Sverdlov, 1998).

Жизненный цикл эндогенных ретровирусов отличается от такового экзогенных ретровирусов тем, что у ERV и ретротранспозонов отсутствует внеклеточная фаза, для которой необходимы белки, кодируемые геном *env*. Несмотря на то, что полноразмерные HERV содержат ген *env*, их инфекционность не выявлена до настоящего времени. Имеются ретровирусы, находящиеся как в экзогенной, так и в эндогенной форме в организме хозяина, например MMTV и JSRV (Jaagsiekte Sheep Retrovirus) (Lower et al., 1996; Qin et al., 1999).

### *2.3.3. Возникновение, структура и классификация ретроэлементов*

Общим свойством, объединяющим все ретроэлементы, является ретротранспозиция, т.е. способность к умножению числа копий ДНК в геноме хозяина при помощи обратной транскрипции и реинтеграции. Наиболее важная черта этого явления - образование коротких прямых повторов хозяйской ДНК в сайте интеграции. Большая часть ретроэлементов в геноме транскрипционно не активна, а умножение числа копий в геноме происходит, по некоторым предположениям, благодаря активности «мастер-генов» (Deininger et al., 1992).

В настоящее время основными признаются две гипотезы, объясняющие возникновение экзогенных ретровирусов и ретроэлементов. В соответствии с одной из них, клетки зародышевого пути предков современных животных были инфицированы экзогенными предшественниками ERV на разных этапах эволюции. В клетках хозяина они реплицировались вместе с хозяйскими генами, претерпевая внутригеномную амплификацию и ретротранспозицию (Lower et al.,

1996). У особей, которые развивались из зараженных гамет, провирусы реплицировались вместе с генами во всех клетках организма хозяина и наследовались в соответствии с законами Менделя. Большая часть ERV лишилась ряда свойств экзогенных ретровирусов, например, способности образовывать инфекционные вирусные частицы, в процессе эволюции по причине мутаций, (Sverdlov, 1998).

Протовирусная гипотеза, сформулированная Теминым, допускает поэтапную эволюцию ретроэлементов от простых к более сложным по составу. В соответствии с этим воззрением, ретроэлементы в геноме эволюировали от прототипа гена *pol*. Обретение новых генов, кодирующих структурные белки, благодаря рекомбинациям, а также регуляторных последовательностей, привело к образованию существующего многообразия ретроэлементов, в том числе экзогенных ретровирусов (Temin, 1971; Lindeskog et al., 1999).

**Таблица 1.** Характеристика ретроэлементов из (Fuchs et al., 2011; Lower et al., 1996) с дополнением (Bannert and Kurth, 2004).

Элемент	Структурные различия	Пример	Число копий на геном; (%)
Ретроген (псевдоген)	Ретроэлемент с (без) промотором без гена RT*	Фосфоглицераткиназа человека	1-10; суммарно <1%
SINE-ретропозон	Ретроэлемент с внутренним промотором без гена RT	Alu SINE-R	10 <sup>6</sup> ; 13% 10 <sup>3</sup> ; незначительная
LINE-ретропозон	Ретроэлемент с внутренним промотором и гена RT	LINE-1	8x10 <sup>5</sup> ; 21%
Ретротранс-позон	Ретроэлемент с геном RT и LTR	RTVL-H - дефектные HERV-H элементы	~1000
Эндогенный ретровирус	Ретроэлемент с геном RT, LTR и <i>env</i>	HERV-H, HERV-K	~100 ~50
LTR	Одиночный LTR	LTR HERV-K LTR HERV-H	10 <sup>3</sup> -10 <sup>4</sup> ; ~0,2% ~1000

\*RT – ген, кодирующий обратную транскриптазу.



В соответствии с этими гипотезами, можно допустить либо потерю, либо обретение гена *env* в процессе эволюции, что повлекло превращение ретровирусов в ретротранспозоны или образование экзогенных ретровирусов. Необходимо подчеркнуть, что гипотезы эти не являются противоречащими или несовместимыми - эволюция ретроэлементов шла, вероятно, по обоим путям (Leib-Mosch and Seifarth, 1995).

Принято подразделять ретроэлементы на несколько групп в соответствии с их структурными различиями и на основании положения на эволюционном «древе» (табл. 1) (Leib-Mosch and Seifarth, 1995; Lower et al., 1996; Andersson et al., 1998; Fuchs et al., 2011; Benachenhou et al., 2013).

*Псевдогены.* Появляются в результате обратной транскрипции мРНК хозяина с последующей реинтеграцией кДНК в геном. Псевдогены, которые активно транскрибируются благодаря наличию промоторных последовательностей, были названы *ретрогенами*. Они не обладают активностью обратной транскриптазы и амплифицируются с небольшим количеством копий. Онкогены в составе провирусов и «мастер-гены» Alu-повторов, произошедшие от гена 7SL РНК предков человека, некоторые исследователи также относят к ретрогенам (Mighell et al., 1997).

*Ретропозоны.* Отличительной чертой ретропозонов является отсутствие гена *env* и длинных концевых повторов. Ретропозоны подразделяют на LINE (Long Interspersed Element) в случае присутствия гена, гомологичного гену *pol*, и SINE (Short Interspersed Element) – в случае его отсутствия (Lower et al., 1996). В геноме они амплифицированы до очень большого количества копий; значительная часть из них содержит множественные мутации и делеции. Alu-повторы, например, относятся к SINE, и являются наиболее многочисленной группой ретроэлементов в геноме человека с числом копий порядка  $10^6$  (Mighell et al., 1997; Lander et al., 2001). Вследствие того, что ретропозоны не имеют механизма восстановления промотора после транскрипции, аналогичного синтезу LTR, они транскрибируются с внутренних промоторов, LINE - РНК полимеразой

II, a SINE - РНК полимеразой III (Leib-Mosch and Seifarth, 1995; Gualtieri et al., 2013).

*Ретротранспозоны* обнаружены в геномах большинства организмов – от простейших до человека. Они отличаются от ретровирусов отсутствием гена *env*. Ретротранспозоны представлены большим числом копий в геноме, многие из которых дефектны. Наличие генов *gag* и *pol* и LTR, а также гомологичность последовательностей провирусов и ретротранспозонов, указывают на то, что последние могут быть предшественниками ретровирусов или, напротив, ретровирусами с делецией гена *env*.

*Эндогенные ретровирусы (ERV)*. Ретровирусы представляют собой специализированные мобильные элементы, приспособленные к стремительному распространению в популяции хозяина (Lower et al., 1996; Hayes et al., 2013). За счет способности к ретротранспозиции в геноме постоянно возрастает число копий провируса и одиночных LTR, однако новых инфекций не происходит. Поскольку в геноме человека около 8% приходится на последовательности, гомологичные последовательностям ретровирусов (Lander et al., 2001), подобная внутриклеточная ретротранспозиция весьма продуктивна. Подобно “классическим” провирусам экзогенных ретровирусов, полноразмерные HERV так же содержат гены *gag*, *pol* и *env*, несущие информацию о синтезе белков, необходимых для сборки инфекционных частиц и фланкированы длинными концевыми повторами, регулирующими их экспрессию.

*Одиночные LTR* формируются вследствие рекомбинации между LTR, фланкирующими ретровирус, и делеции внутренних последовательностей. Они являются следами интеграций ретровирусов (Sverdlov, 2000; Subramanian et al., 2011). Рекомбинация между двумя одиночными LTR или LTR двух различных провирусов является теоретически возможной, и это может привести к инверсии или потере участка хромосомы хозяина. Впрочем, фиксировать такие события можно лишь при сохранении обоих аллелей в организме хозяина. В качестве примера можно привести ген рецептора гормона роста человека, третий экзон которого либо фланкирован фрагментами LTR семейства HERV-P, либо

отсутствует (частоты встречаемости аллелей 75% и 25%, соответственно). Во втором варианте в интроне между 2-м и 4-м экзонами присутствует продукт гомологичной рекомбинации – одиночный LTR (Pantel et al., 2000).

Возможно, в процессе эволюции сохранились лишь те интеграции, которые не причиняли вреда организму хозяина или играли определенную положительную для него роль. Наличие активных ретроэлементов в геноме в течение продолжительного времени влечет за собой необходимость формирования механизмов коадаптации, уменьшающих отрицательные эффекты транспозиции, например, путем специфической интеграции в некодирующие, нетранслируемые области, интроны и межгенные пространства. Другой путь - это реинтеграция в уже существующий ретроэлемент или в область, соседствующую с ним (Sverdlov, 1998; Subramanian et al., 2011). Например, локус гаптоглобина содержит три автономно интегрированные HERV-I элемента (Leib-Mosch and Seifarth, 1995). Два автономных LTR семейства HERV-K(HML-6) интегрированы в геномы двух его представителей HML-6.4 и HML-6.17 (Medstrand et al., 1997). Описаны явления интеграции провирусов одного семейства в последовательность провируса другого семейства в составе ДНК хозяина. В частности, в 5'-LTR одного HERV-K(HML-2) элемента на 19 хромосоме расположен целый HERV-H, а на хромосоме 15 выявлен LTR HuERs-P3 со вставкой LTR ERV9 (Lapuk et al., 1999). Описан гибрид HERV-H/HERV-E, который содержит две пары генов *gag* и *pol* и ген *env* от HERV-E. Он присутствует в нескольких копиях в геноме человека. Для него было продемонстрировано наличие сплайсированных гибридных транскриптов (Lindeskog et al., 1998). Ген плеiotропина (*PTN*) содержит гибридный провирус, в котором между генами *gag* и *pol*, подобных последовательностям представителей семейства HERV-E, встроены гены *pol* и *env* элемента HERV-I. Внедрение данного элемента явилось причиной образования промотора внутри гена *PTN*, который активен в некоторых эмбриональных и опухолевых тканях. Вместе с тем, транскрибируется гибридный продукт, содержащий три 5'-нетранслируемых вирусных экзона и открытую рамку считывания *PTN* (Schulte and Wellstein, 1998).

Следовательно, взаимные рекомбинации и интеграции разнообразных одиночных LTR и HERV довольно часты. Такого типа события защищают организм хозяина от негативных последствий инсерций, и, вместе с тем, повышают эволюционный ресурс генома, “перемешивая” регуляторные последовательности, входящие в состав подвижных элементов. Например, в гибриде HERV-K/HERV-H расположены разделенные спейсером в 400 пн консервативная последовательность HERV-K, связывающая человеческие ядерные белковые факторы, и GC/GT-бокс, обеспечивающий транскрипционную активность LTR HERV-H (Lapuk et al., 1999).

#### *2.3.4. Структура эндогенных ретровирусов*

Около двух десятков различных семейств эндогенных ретровирусов, обнаруженных в геноме человека, подразделяются на три класса. Это подразделение основано на их сходстве с экзогенными ретровирусами млекопитающих типа С (класс I), с ретровирусами млекопитающих типов А, В, D и ретровирусами птиц типа С (класс II) и со спумавирусами (класс III) (Tristem, 2000). Систематика HERV довольно сложна, одновременно существуют различные системы классификации (Lower et al., 1996; Andersson et al., 1999). В основу классификационного подразделения положен тип молекулы тРНК, служащей праймером при репликации ретровируса. Одновременно, в названии семейства HERV указывается однобуквенный код аминокислоты, соответствующей тРНК. Как и MMTV, все эндогенные ретровирусы человека *класса II*, для которых определена последовательность сайта связывания праймера, имеют PBS, комплементарные лизиновой тРНК, и, соответственно, их обозначают HERV-K. Позднее, для них была предложена система классификации, рассматривающая еще и гомологию последовательностей HERV, наряду со специфичностью PBS (Andersson et al., 1999).

#### *2.3.5. Семейство HERV-K (HML-2)*

Способность HERV к образованию инфекционных частиц была утрачена в результате того, что большая часть последовательностей подверглась точечным мутациям и делециям. Многие элементы HERV-K(HML-2), в отличие от представителей большинства других семейств, транскрипционно активны и содержат небольшое число мутаций, нарушающих открытые рамки считывания (Zsiros et al., 1998). Они содержат последовательности, кодирующие функциональные ферменты, вирусные частицы и аутоантигены, что свидетельствует о наличии у некоторых из них свойств ретровирусов (Medstrand and Mager, 1998; Schmitt et al., 2013). Кодированные HERV-K ретровирусные белки, активно экспрессируются в различных клеточных линиях. Вирусоподобные частицы HERV-K HTDV, найденные в эмбриональных тканях, плаценте и тератокарциномах, содержат функциональную обратную транскриптазу (Lower et al., 1996; Seifarth et al., 1998; Ruprecht et al., 2008; Hohn et al., 2013). Имеются отдельные данные о случаях ретротранспозиций представителей этого семейства, произошедшие в геноме предков человека за последние пять миллионов лет (Medstrand and Mager, 1998; Lebedev et al., 2000; Costas, 2001). Обнаружение в геноме тринадцати человекспецифичных провирусов HERV-K (Hughes and Coffin, 2004) и более сотни одиночных LTR указывает на вероятность подобных событий (Buzdin et al., 2003). Следовательно, в настоящее время семейство HERV-K можно считать одним из наиболее биологически активных семейств HERV.

Описываемое нами семейство довольно велико: около 30 полноразмерных элементов присутствуют в гаплоидном геноме. Для сравнения, семейство HERV-R представлено всего одним экземпляром на гаплоидный геном.

Образование вирусоподобных частиц – HTDV становится возможным благодаря комплементации нескольких копий провирусов с различными мутациями (Tonjes et al., 1999; Hohn et al., 2013). Существование открытых рамок считывания для всех вирусных генов было подтверждено анализом кДНК из линий клеток тератокарцином, производящих HTDV (Lower et al., 1996; Lee and Bieniasz, 2007). Оказалось, что комплементация допустима не только между провирусами одного семейства HERV, но также между экзо- и эндогенными

ретровирусами. Так, например, протеаза, кодируемая HERV-K10 является активным эндогенным ферментом, способна к процессингу белков вируса иммунодефицита человека (HIV-1), и, в то же время, является устойчивой к ингибиторам, используемым для подавления активности протеазы HIV (Towler et al., 1998; Kraus et al., 2011). Вероятно, HIV использует эндогенную протеазу в случае, когда его собственный фермент ингибирован. Показано также, что для переноса несплайсированной вирусной РНК в цитоплазму, элемент RcRE (Res-responsive element) транскриптов HERV-K(HML-2) может эффективно взаимодействовать не только с Res-белком, но и с его аналогами – Rev (regulator of expression of viral proteins) и Rex, кодируемыми HIV и HTDV, соответственно (Magin-Lachmann et al., 2001; Hanke et al., 2012).

Семейство HERV-K(HML-2) может являться переходным звеном в эволюции от простых ретровирусов к сложным. В пользу этого предположения свидетельствует обнаружение у представителей семейства транспортного комплекса RcRE/Res, участвующего в регуляции экспрессии (Magin et al., 1999). На сегодняшний день семейство HERV-K(HML-2) является единственным из семейств HERV, у представителей которого обнаружен ген регуляторного белка, характерный для сложных ретровирусов (Bannert and Kurth, 2004; Morozov et al., 2013).

Одной из проблем ксенотрансплантации, то есть пересадки человеку органов животного, может быть наличие в геномах млекопитающих множества эндогенных ретровирусов, которые могут стать активными в клетках человека благодаря комплементации. Кроме того, рекомбинация эндогенных ретровирусов животного и человека может привести к возникновению нового, возможно патогенного, типа вируса (Wilson et al., 2003; Shin et al., 2013).

### *2.3.6. Функциональные элементы в составе LTR*

Контроль экспрессии ретровирусных последовательностей обеспечивается регуляторными элементами, входящими в состав LTR, в особенности в его U3-область на 5'-конце провируса. Обычно в состав LTR входит весь ансамбль

регуляторных элементов, обязательных для транскрипции провирусной ДНК РНК-полимеразой II хозяина: промотор, энхансер, сигналы терминации и полиаденилирования РНК. Помимо этого, некоторые LTR содержат элементы, отзывающиеся на сигналы, поступающие в клетку, например, глюкокортикоидные гормоны (Hormone Responsive Element - HRE) (Ono, 1986; Sverdlov, 1998; Manghera and Douville, 2013). Отдельные представители семейств ретровирусов, в том числе HERV-K(HML-2) и HTLV-I, кодируют элемент, который взаимодействует с транспортным белком, обеспечивающим перенос несплайсированной РНК в цитоплазму (Magin et al., 1999).

Два главных элемента, контролирующих экспрессию, расположены в U3-области LTR ретровирусов: промотор находится непосредственно перед сайтом инициации транскрипции на границе U3/R, а энхансер содержит ряд сайтов связывания транскрипционных факторов. Также в U3-области на 5'-конце находится сигнальная последовательность (*att*), которая обеспечивает успешную интеграцию вирусной ДНК в геном. В начале R области располагается сайт инициации транскрипции (чаще всего GC) (Strazzullo et al., 1998). У представителей семейства HERV-K(HML-2) R-область выполняет важную роль при переносе цепи в процессе обратной транскрипции, также в ней находится сигнал полиаденилирования. Следом за сайтом полиаденилирования начинается область U5, которая содержит последовательности, необходимые для запуска обратной транскрипции, на 3'-конце которой находится сигнальная последовательность (*att*). Вирусный геном теряет по две пары нуклеотидов на каждом конце при интеграции в геном хозяина, которые восстанавливаются во время следующего цикла репликации за счет *att*, расположенных в U3 и U5 областях.

При сравнении последовательностей большого числа длинных концевых повторов HERV-K была выявлена неравномерность распределения частот мутаций вдоль LTR. В структуре LTR имеется несколько высококонсервативных участков, локализованных в районе границ U3/R и R/U5, потенциального энхансера, HRE, промотора и сигнала полиаденилирования. Кроме того, имеются

два консервативных участка, которые содержат предполагаемые сайты связывания транскрипционных факторов. Эти данные неявно указывают на способность LTR к регуляции транскрипции (Lavrentieva et al., 1998).

Регуляция экспрессии экзогенных ретровирусов изучена в деталях (Kingsman and Kingsman, 1996; Einfeld et al., 1997; Qin et al., 1999; Zhu et al., 2000; Katoh and Kurata, 2013; Katoh, 2014), однако регуляция экспрессии HERV исследована значительно хуже. Для ERV-9 установлено, что транскрипция активируется на его промоторе фактором транскрипции Sp1 в комплексе с неизвестным белком (La Mantia et al., 1992; Strazzullo et al., 1994). Представители белков семейства Sp1 инициируют также транскрипцию с промотора HERV-H (Sjottem et al., 1996; Katoh and Kurata, 2013). Вероятный участок связывания Sp1 находится и в U3-области LTR HERV-K (Lavrentieva et al., 1998; Fuchs et al., 2011). Для представителей семейства HERV-K методом репортерного гена были идентифицированы энхансер на 5'-конце U3-области и минимальный промотор на ее 3'-конце (Domansky et al., 2000; Santoni et al., 2012). Исследование энхансера в составе области U3 LTR HERV-K методом EMSA (Electrophoretic Mobility Shift Assay) выявил присутствие участков связывания ядерных белков клетки-хозяина (Акоров et al., 1998; Knossl et al., 1999). Оказалось, что одним из транскрипционных факторов, взаимодействующих с энхансером, является YY1, консенсусная последовательность которого расположена с 64 по 80 п.н. LTR (Knossl et al., 1999). Были выявлены три новых, не охарактеризованных ранее, белка, названных ERLBF-1, -2 и -3 (Endogenous Retrovirus LTR Binding Factor), которые в виде комплекса связываются с последовательностью LTR между 49 и 97 п.н. (Акоров et al., 1998).

**Таблица 2.** Факторы транскрипции клетки-хозяина, взаимодействующие с LTR HERV.

Таксон	Факторы, взаимодействующие с LTR	Источник
HERV-H	Белки семейства Sp1; Myb	(Sjottem et al., 1996; de



		Parseval et al., 1999)
HERV-K(HML-2)	ERLBF-1,-2,-3; YY1; для Sp-1 и AP-2; потенциальные сайты связывания; YY1	(Akopov et al., 1998; Lavrentieva et al., 1998; Knossl et al., 1999; Fuchs et al., 2011);
ERV9	Sp1	(La Mantia et al., 1992)

Имеются результаты, указывающие на наличие у некоторых LTR двунаправленной промоторной активности. Так, представители семейства HERV-N имеют LTR, способные инициировать транскрипцию в обратном направлении (Feuchter and Mager, 1990; Levin and Moran, 2011). Такими же особенностями обладают LTR IAP мыши, HERV-K и HIV (Peeters et al., 1996; Domansky et al., 2000). При функциональной диссекции LTR IAP мыши была продемонстрирована необходимость 5'-концевого фрагмента U3-района для промоторной активности в "обратной" ориентации. В случае делеции 5'-концевого фрагмента U3-района промоторная активность LTR HERV-K в "обратной" ориентации снижалась в несколько раз, но в отличие от IAP, инициация транскрипции с LTR HERV-K могла происходить в области, отличной от предполагаемой промоторной области на границе U3/R (Christy and Huang, 1988; Mayer et al., 1993; Domansky et al., 2000; Levin and Moran, 2011). Таким образом, LTR, которые подверглись делеции канонических регуляторных областей, могут инициировать экспрессию с альтернативных промоторов.

Следует отметить, что U5-район и функционирующие в нем регуляторные элементы мало изучены в сравнении с U3-районом LTR HERV-K, содержащим канонические энхансер и промотор. Вместе с тем, имеются отдельные указания на наличие регуляторных элементов в U5-районе LTR некоторых экзогенных ретровирусов и родственного ретровирусам HFV (Human Foamy Virus, см. табл. 3).

**Таблица 3.** Регуляторные элементы расположенные в U5-районе LTR ретровирусов.

Вирус	Функции	Ссылка
BIV	U5-район и нетранслируемый лидер подавляют транскрипцию во всех исследованных линиях клеток	(Fong et al., 1995)
HTLV-I	Негативный регулятор, содержащий сайт связывания Sp1 в U5-районе	(Okumura et al., 1996)
HFV	Негативный регулятор, снижающий уровень экспрессии на порядок в U5-районе	(Yang et al., 1997)
HIV-I	Позитивный регуляторный элемент, взаимодействующий с факторами транскрипции AP-1 и CREB-1/ATF-1,-2	(Rabbi et al., 1997; Pereira et al., 2000)

Имеются данные, указывающие на существование в составе U5-района LTR HERV-K негативного регулятора транскрипции, проявляющего активность в линии клеток NT2/D1 (Domansky et al., 2000).

### 2.3.7. Биологическая роль LTR

Анализ распределения LTR HERV-K на хромосоме 19 (Lebedev et al., 1995; Vinogradova et al., 1997; Lavrentieva et al., 1998) показал, что не обнаруживается корреляции между сближенностью LTR на карте хромосомы и степенью идентичности их последовательностей. Вместе с тем, позиции LTR на хромосоме 19 коррелируют с расположением генов. Несмотря на то, что трактовать подобные результаты как проявление регуляторной активности между соседствующими на физической карте LTR и генами не вполне корректно, ряд совпадений кажутся весьма интересными. Более существенным является многократная конкордантность позиций LTR и генов белков, содержащих мотивы «цинковых пальцев» (Vinogradova et al., 1997; Thomas and Schneider, 2011). Кроме этого LTR были выявлены вблизи генов, причастных к пути передачи сигнала Jak-Stat (Sverdlov, 1998). Такое многократное совпадение позиций LTR и генов с аналогичными или сцепленными функциями может указывать на их эволюционное родство или участие LTR в регуляции их экспрессии.

В широкий набор биологических функций HERV входит активация и репрессия клеточных генов, экспрессия вирусных белков, перестройки генных локусов при ретропозициях HERV, образование вирусных частиц, происходящие с участием длинных концевых повторов (Leib-Mosch et al., 1992; Di Cristofano et al., 1995b; Hohenadl et al., 1996; Lower et al., 1996; Pantel et al., 2000; Sverdlov, 2000; Hancks and Kazazian, 2012; Friedli et al., 2014; Turelli et al., 2014).

### 2.3.7.1. Экспрессия вирусных генов

Многие провирусы HERV-K содержат открытые рамки считывания всех вирусных белков. Это отличает их от эндогенных ретровирусов человека других семейств, которые в большинстве своем дефектны. Экспрессия генов *gag*, *pol*, *env* осуществляется в различных линиях клеток, преимущественно в линиях тератокарцином, в которых описано образование вирусных частиц HTDV (Kurth et al., 1983; Boller et al., 1993; Lower et al., 1996; Kessler et al., 2014). Наличие в этих линиях клеток молекул процессированного белка Gag свидетельствует о присутствии функциональной протеазы. В лимфоцитах здоровых доноров был обнаружен транскрипт, гомологичный на 98% последовательности гена *pol* HERV-K10 (Medstrand et al., 1992). В экстрактах вирусоподобных частиц, выделенных из супернатанта клеток линий тератокарцином, наблюдалась активность обратной транскриптазы. Позднее при помощи различных методов было выявлено присутствие вирусных белков в плаценте на определенном сроке беременности, и в некоторых герминогенных опухолях (тератокарциномах, семиномах), миелоидных лейкомиях и глиомах (Lower et al., 1996; Harris, 1998; Seifarth et al., 1998). Наличие транскриптов ERV в плаценте человека и животных свидетельствует о влиянии белков эндогенных ретровирусов на имплантацию зиготы и формирование плаценты в процессе беременности (Harris, 1998). Показано, что экспрессия белка Env на мембране трофобластов из плаценты в состоянии оказывать воздействие на межклеточные мембранные взаимодействия и помогает в формировании синцития. Также описана иммуносупрессивная активность, которой обладают очищенный белок Env, а также пептиды,

соответствующие его высококонсервативному мембранному домену (Cianciolo et al., 1985; Fei et al., 2013).

Активно изучается участие HERV в различных патологиях человека. Экспрессия HERV часто связана с воспалительными заболеваниями нервной системы, аутоиммунными заболеваниями и иммуноглобулинопатиями (Urnovitz and Murphy, 1996; Katoh and Kurata, 2013; Katoh, 2014). Часто экспрессия HERV приводит к синтезу и иммунной представленности ретровирусных белков, сходных по аминокислотному составу с клеточными белками (молекулярная мимикрия). В некоторых работах по белкам, кодируемым ERV, высказывается предположение о том, что аутоиммунная реакция вызывается молекулярной мимикрией клеточных белков (Hohenadl et al., 1996; Tugnet et al., 2013). Возможно также, что иммунная система испытывает воздействие со стороны суперантигенов, которые кодируются ретровирусными последовательностями (Indraccolo et al., 1995; Katoh and Kurata, 2013). К примеру, подобные суперантигены кодируются вирусом MMTV, а их экспрессия вызывает активацию Т-лимфоцитов и эксплозию цитокинов, что порождает неспецифический иммунный ответ, токсический шок и, в некоторых случаях, смерть.

Можно предположить, что HERV-K, вследствие родства с MMTV, также могут кодировать эндогенные суперантигены человека (Conrad et al., 1997; Katoh, 2014). Примером кодирующей функции LTR является суперантиген MMTV - U3 область длиной 1197 п.н. вмещает его открытую рамку считывания (Urnovitz and Murphy, 1996).

#### *2.3.7.2. Влияние LTR на экспрессию клеточных генов*

Реинтеграция ERV в область экзона может вызвать сдвиг рамки считывания, что приводит к утрате функции одного из двух аллелей гена. Интеграция LTR в регуляторную область гена клетки хозяина может присоединить к клеточным элементам этого гена регуляторные элементы LTR и, в результате, изменить его экспрессию в ту или иную сторону, поменять ее

тканеспецифичность или привести к ранней терминации транскрипции (Mager et al., 1999; Katoh and Kurata, 2013).

Подобные факты были выявлены в ряде исследований. Например, транскрипция гена *Hox-2.4* мыши инициируется 3'-LTR IAP (Kongsuwan et al., 1989), а ген *H-plk*, кодирующий содержащий цинковые пальцы белок, транскрибируется с 5'-LTR провируса HERV-R (Kato et al., 1987; Kato et al., 1990; Abrarova et al., 2010). Показано, что LTR ERV-9 регулирует экспрессию гена *ZNF80*, кодирующего высококонсервативный у приматов белок (Di Cristofano et al., 1995a; Cohen et al., 2009). Участвующие в альтернативном сплайсинге *ZNF177* последовательности HERV-H (Baban et al., 1996) супрессируют трансляцию гибридного транскрипта из локуса *PLA2L*, который содержит два домена, гомологичных фосфолипазе  $A_2$  (Kowalski and Mager, 1998; Gogvadze and Buzdin, 2009). В 19-м интроне гена лептинового рецептора *obr* человека расположен LTR HERV-K, который предоставляет альтернативный последний экзон и сигнал полиаденилирования. Вследствие этого, у человека присутствует длинная и короткая форма рецептора, при этом 67 С-концевых аминокислотных остатков короткой формы кодирует LTR (Kapitonov and Jurka, 1999; Fahlbusch et al., 2012). В ходе эволюции гоминид внедрение химерного элемента HERV-E/RTVL-I в интрон перед ORF гена плейотропина определяет его высокую тканеспецифичную экспрессию в трофобластах (Schulte and Wellstein, 1998; Fahlbusch et al., 2012). В районе *HLA-DQ* и возле гена белка системы комплемента C2 описано несколько элементов HERV-K (Hohenadl et al., 1996). В интроне гена *ZNF91* найден LTR HERV-K (Khil et al., 1998). Большая доля среди регулируемых LTR ERV генов тех, которые кодируют белки цинковых пальцев, интересна тем, что сами эти белки являются регуляторами транскрипции и нередко участвуют в контроле активности генов развития и дифференцировки (Di Cristofano et al., 1995a; Gonzalez-Hernandez et al., 2012; Schoorlemmer et al., 2014). Ретротранспозиция LTR может изменить уровень регуляции экспрессии транскрипционных факторов и, таким образом, воздействовать на большое число генов.

*Ремоделирование хроматина* – еще один путь воздействия LTR на транскрипцию (Sverdlov, 1998; Kejnovsky and Lexa, 2014). Подобный эффект оказывает LTR вируса MMTV, гомологичного HERV класса II. LTR MMTV изменяет близлежащий хроматин в геноме, препятствуя связыванию транскрипционных факторов с ДНК (Eisfeld et al., 1997). Помимо этого, LTR включает участок связывания рецептора глюкокортикоидных гормонов. Связывание комплекса глюкокортикоида и его рецептора с LTR вызывает разворачивание хроматина и обеспечивает контакт с транскрипционными факторами, реализуя гормональную индукцию LTR. Ранее отмечалось взаимодействие LTR HERV-K с транскрипционным фактором YY1, который также принимает участие в ремоделировании хроматина (Knossl et al., 1999; Thomas and Seto, 1999). Ремоделирование хроматина – ответственный этап регуляции транскрипции, а вовлеченность LTR в этот процесс может означать их влияние на близко расположенные гены без непосредственного участия в транскрипции (Sverdlov, 1998; Kejnovsky and Lexa, 2014).

#### *2.3.7.3. Защита от повторного заражения*

В ходе эволюции положительную роль для потомков-носителей интегрированных HERV могла играть невосприимчивость к повторному заражению экзогенными ретровирусами, обусловленная встроенными HERV (Leib-Mosch et al., 1992; Gonzalez-Hernandez et al., 2012). Вероятно, этот эффект мог содействовать закреплению провирусов, которые, вследствие эндогенной природы, испытывали меньшее селективное давление (Lower et al., 1996).

#### *2.3.7.4. Придание пластичности геному*

Рекомбинации и транспозиции являются эффективными инструментами для придания пластичности геному. Например, образование двух аллелей и двух форм рецептора гормона роста у человека произошло вследствие делеции третьего экзона с фрагментами второго и третьего интронов в результате

гомологичной рекомбинации двух LTR HERV-P (Pantel et al., 2000). Причиной вариабельности генома может являться и присутствие обратной транскриптазы, основными поставщиками которой являются HERV, ретротранспозоны и LINE-ретропозоны (Lander et al., 2001; Finnegan, 2012).

### *2.3.8. Белковые факторы транскрипции, способные связываться с LTR*

Проявление LTR различных типов активности обеспечивается при связывании ими регуляторных белков. С использованием метода торможения в геле выявили связывание LTR HERV-H с транскрипционными факторами семейства Sp1. Было продемонстрировано, что LTR HERV-H могут инициировать транскрипцию как собственных, так и соседних клеточных генов. Промоторная активность характерна лишь для одной субпопуляция LTR HERV-H, содержащей GC/GT мотив с 3'-стороны от TATA-бокса. В клетках NTera2-D1 с GC/GT боксом связываются факторы транскрипции Sp1, Sp3 и BTEB (Nelson et al., 1996; Sjøttem et al., 1996; Fuchs et al., 2011). В последовательности LTR HERV-H были идентифицированы сайты связывания регуляторного белка Myb и выявлено семикратное увеличение активности промотора при котрансфекции с конструкцией, экспрессирующей Myb (de Parseval et al., 1999; Manghera and Douville, 2013). Последовательности LTR семейства HERV-K были также проверены на наличие участков связывания белковых факторов клетки-хозяина с использованием торможения в геле и ультрафиолетовой сшивки. Было показано, что последовательности LTR двух подсемейств HERV-K содержат специфические сайты связывания по крайней мере для трёх белковых факторов, ERLBF1, ERLBF2 и ERLBF3. По своим характеристикам эти белки не похожи на уже известные факторы транскрипции. Участок их связывания расположен в 5'-области U3-элемента LTR. Сохранение специфического белкового сайта связывания в различных LTR HERV-K, предполагает их возможную роль в регуляции близлежащих генов (Акоров et al., 1998).

### 3. Материалы и методы

#### 3.1. Материалы

##### Ферменты

Термостабильная ДНК-полимераза *Thermus aquaticus* (Taq-полимераза), термостабильная ДНК-полимераза *Thermus aquaticus* (Taq-полимераза) с антителами - ИБХ РАН, Россия.

Смесь термостабильных ДНК-полимераз Encyclo PCR kit - Евроген, Россия.

Эндонуклеазы рестрикции *Eco* RV, *Xba* I, *Not* I, *Nhe* I (*Asu* NHI), *Cla* I, *Sal* I, *Xho* I, *Pst* I, *Sac*II, *Ssp* I *Eco* 47III, T4 ДНК лигаза, фрагмент Кленова ДНК-полимеразы I *E. coli*, SAP (щелочная фосфатаза креветки) - MBI Fermentas, Литва.

Эндонуклеаза рестрикции *Psp* EI - СибЭнзим, Россия.

Эндонуклеаза рестрикции *Bam* HI - New England BioLabs, США.

Лизоцим марки В – Реахим, Россия.

РНК-аза А из поджелудочной железы быка (10 мг/мл) – Serva, США.

Протеиназа К – Promega, США.

Трипсин – Calbiochem, США.

##### Буферные растворы

Буфер для ПЦР реакции (10×) для Taq-полимеразы (ИБХ): 500 мМ KCl, 100 мМ Трис-HCl pH 9,0, 1% Triton X-100.

Буфер для нанесения ДНК на агарозный гель (6x): 0,25% бромфеноловый синий, 0,25% ксиленцианол, 30% глицерин

Буфер TE, pH 7,6: 10 мМ Трис-HCl pH 7,6, 1 мМ EDTA

Буфер TE, pH 8,0: 10 мМ Трис-HCl pH 8,0, 1 мМ EDTA

Буфер TAE: 40 мМ Трис-ацетат pH 8,2, 1 мМ EDTA

Буфер TBEx0,5: 45 мМ трис-борат, 1 мМ EDTA, pH 8,0.

Буфер В (12 мМ HEPES–KOH, pH 7,9, 12% глицерин, 60 мМ KCl, 0,3 мМ EDTA, 0,6 мМ ДТТ).



Буфер для элюции ДНК из полиакриламидного геля: 0,5 М ацетат аммония, 1% SDS, 1 мМ EDTA.

Буфер PBS (Phosphate Buffered Saline): 1,7 мМ  $\text{KH}_2\text{PO}_4$ , 5,2 мМ  $\text{Na}_2\text{HPO}_4$ , 150 мМ NaCl.

Буфер TNM (5x) для отжига олигонуклеотидов: 200 мМ Трис-HCl pH 8,0, 250 мМ NaCl; 100 мМ  $\text{MgCl}_2$ .

20x SSC: 3 М NaCl, 0,3 М цитрат натрия.

50x реагент Денхардта: 5 г фиколл, 5 г поливинилпирролидон, 5 г БСА на 500 мл  $\text{H}_2\text{O}$ .

Денатурирующий раствор для переноса по Саузерну: 0,5 М NaOH, 1,5 М NaCl.

Нейтрализующий раствор для переноса по Саузерну: 1 М Трис-HCl pH 7,4, 1,5 М NaCl.

Предгибризационный буфер: 6xSSC, 0,5% SDS, 5x раствор Денхардта, 100 мкг/мл денатурированной ДНК спермы лосося.

Лизирующий буфер для выделения геномной ДНК: 10 мМ Tris HCl pH 8,5, 5 мМ EDTA, 0,2% SDS, 0,2 М NaCl, 0,1 мг/мл Proteinase K).

#### Наборы реактивов

Wizard Plus Minipreps DNA Purification System - Promega, США.

Wizard Plus Midipreps DNA Purification System - Promega, США.

Wizard SV Gel and PCR Clean-Up System - Promega, США.

Wizard SV Genomic DNA Purification System - Promega, США.

Dual-Luciferase Reporter Assay System - Promega, США.

QIAquick PCR purification kit – Qiagen, ФРГ.

#### Маркеры длин ДНК

100 п.о. GeneRuler 100bp DNA Ladder Plus – Fermentas, Литва.

1 т.п.о.: 1Kb DNA Ladder – GibcoBRL, США.

## Среды

Среда LB: 10 г Vacto-Tryptone, 5 г Vacto-yeast extract, 5 г NaCl на 1 литр стерильной дистиллированной воды.

Среда LB с ампициллином: к среде LB добавляли раствор ампициллина до концентрации 75 мкг/мл.

Агаризованная среда LB с ампициллином: к 1 л среды LB добавляли 15 г агара и автоклавировали, после остывания до 50°C добавляли раствор ампициллина до конечной концентрации 75 мкг/мл.

Сухие смеси для приготовления культуральных сред (DMEM, DMEM:F12 HAM, RPMI-1640) – GibcoBRL, США.

Эмбриональная сыворотка теленка, GibcoBRL, США.

БСА (бычий сывороточный альбумин, 10 мг/мл) - Biolabs, США.

HEPES, Lipofectamine 2000 - Invitrogen, США.

Polybrene - Sigma-Aldrich Co., США.

## Бактериальные клетки

Для трансформации использовали компетентные клетки *E.coli* штамма DH5-альфа, приготовленные Медведевой Н.И в лаборатории структуры и функций генов человека.

## Олигонуклеотиды

В зависимости от целей последовательности праймеров конструировали вручную или с использованием программы Primer3. Праймеры синтезированы Скапцовой Н.В. в лаборатории структуры и функций генов человека ИБХ РАН на синтезаторе ASM 700 (BIOSSET, Новосибирск, Россия), либо компанией Евроген.

## 3.2.Методы

В разделе подробно изложены лишь методы, предложенные либо модифицированные нами. Остальные методы изложены сжато и сопровождаются соответствующими ссылками.

### *3.2.1. Стандартные методики*

Стандартные генно-инженерные манипуляции (обработка ДНК рестриктазами, лигирование, трансформация *E. coli*, выделение плазмидной и геномной ДНК, выделение РНК, синтез кДНК, электрофорез ДНК, РНК и белков в агарозном и полиакриламидном геле, блот-гибридизация и др. проводили согласно руководству по молекулярному клонированию (Sambrook and Russell, 2001).

### *3.2.2. Культивирование клеток*

Клеточные линии НЕК-293 (CRL-1573, клетки почки эмбриона человека, трансформированные ДНК аденовируса Ad5), CHO-K1 (CCL-61, клетки яичника китайского хомячка), HeLa (CCL-2, аденокарцинома шейки матки человека), HL-60 (CRL-1964, промиелоцитарная лейкемия человека), Jurkat (TIB-152, острая Т-лимфобластоподобная лейкемия человека), PANC-1 (CRL-1469, карцинома поджелудочной железы человека), NIH/3T3 (CRL-1658, эмбриональные фибробласты мыши) и Phoenix-AMPHO (клетки пакующей линии, производные от клеток линии 293-T), а также линии тератокарциномы человека – NT2/D1 (CRL-1973) и Tera-1 (HTB-105), HD3 (клон А6 линии LSCC куриных эритроцитов), DT40 (CRL-2111, линия лимфоидных клеток кур) культивировали в средах и при условиях рекомендованных American Type Culture Collection для этих линий. Клеточная линия NGP-127 (нейробластома человека) была любезно предоставлена проф. Paul S. Meltzer. Ее культивировали при условиях, описанных в (Elkahloun et al., 1996). Клеточная линия GS (карцинома почки) была получена из коллекции Института Цитологии (Санкт-Петербург, Россия) (Vydra et al., 1988). Все клеточные линии культивировали при 37°C в атмосфере 5% CO<sub>2</sub>. При пассировании всех адгезивных культур использовали обработку смесью трипсин-Версен (0,5% Trypsin-EDTA).

### *3.2.3. Трансфекция клеток*

Трансфекция клеток с использованием DEAE-декстрана. Для трансфекции суспензионных клеточных культур применяли метод с использованием DEAE-декстрана (Milman and Herzberg, 1981; Sompayrac and Danna, 1981). Перед трансфекцией готовили суспензию клеток с концентрацией  $10^6$  клеток/мл. Для трансфекции отбирали по 10 мл клеточной суспензии на образец. Готовили 25 мкл раствора ДНК, содержащего 5 мкг плазмиды серии pGL3 с клонированным фрагментом и 1 мкг нормирующей плазмиды, и вносили его в 1,4 мл раствора А (DEAE-декстран 10 мг/мл/Трис-НСl 1М рН 7,4/DMEM = 1/4/75). Клетки отмывали от сыворотки, ресуспендировали в содержащем ДНК растворе А и инкубировали 1 час при 37°C. Далее клетки промывали средой RPMI-1640 без сыворотки, ресуспендировали в 6 мл ростовой среды (90% RPMI-1640, 10% FCS) и оставляли на 40-48 часов при 37°C в 5% CO<sub>2</sub>.

Трансфекцию клеток с помощью реагентов Lipofectin и Lipofectamine 2000 осуществляли согласно рекомендациям фирмы-производителя (Invitrogen, США).

### *3.2.4. Трансфекция клеток электропорацией*

Электропорацию клеток осуществляли с помощью прибора Gene Pulser Xcell (BioRad, США) согласно рекомендациям фирмы производителя. За 20 час до трансфекции  $\sim 2 \times 10^6$  клеток высевали в 25 см<sup>2</sup> флакон и культивировали при 37°C в 5% CO<sub>2</sub>. Перед трансфекцией клетки снимали трипсин-версеном, промывали и ресуспендировали в среде OPTI-MEM без сыворотки. В кювету для электропорации помещали 5 мкг линейаризованной плазмиды, затем  $10^6$  клеток в объеме 600-800 мкл среды OPTI-MEM и суспензию подвергали воздействию однократного импульса с экспоненциальным затуханием напряжения (20 мс, 360 В, 800 мкФ). Клетки инкубировали 10-15 мин во льду, переносили в 25 см<sup>2</sup> культуральный флакон и культивировали в 5 мл ростовой среды при 37°C в 5% CO<sub>2</sub> в течение 24 часов.

### 3.2.5. Позитивно-негативная селекция последовательностей потенциальных инсуляторов

Клетки CHO трансфицировали линейризованными плазмидами методом электропорации. Через сутки после трансфекции клеток в среду роста добавляли 500 мкг/мл генетицина G418 (Geneticin®, GibcoBRL, США). По достижении монослоя клетки пересевали и продолжали культивировать в среде с селективным агентом. Через две недели в ростовую среду вместе с генетицином добавляли до 4 мкМ ганцикловира (Цимевен®, Roche, Швейцария). Через две недели устойчивые к генетицину G418 и ганцикловиру клетки собирали, промывали PBS и выделяли геномную ДНК. Участок интегрированной плазмиды между промотором и энхансером CMV амплифицировали при помощи двухстадийной-ПЦР. Амплифицированный пул фрагментов расщепляли *Xho* I, клонировали в вектор pGEM-T Easy и полученные после трансформации *E. coli* колонии ранжировали в 96-луночных планшетах.

### 3.2.6. Получение вирусных частиц и инфицирование клеток

Через 18-20 час после трансфекции клеток Phoenix-AMPHO ретровирусными конструкциями с использованием реагента Lipofectamine 2000 среду роста заменяли на свежую и инкубировали клетки еще 24 часа при 32°C в атмосфере 5% CO<sub>2</sub>. Затем среду с вирусными частицами собирали и центрифугировали 5 мин при 12000 об/мин. Полученный супернатант использовали для инфицирования клеток HeLa, а также для определения титра вирусных частиц.

За 24 часа до инфицирования  $\sim 2 \times 10^6$  клеток HeLa высевали в 25 см<sup>2</sup> культуральные флаконы. Через сутки среду во флаконах удаляли и добавляли 3 мл среды с вирусными частицами и 2 мл среды роста, содержащей 40 мкг полибрена, и инкубировали клетки 20 часов при 32°C в 5% CO<sub>2</sub>. Затем среду заменяли на 5 мл свежей и инкубировали клетки 24 часа при 37°C в 5% CO<sub>2</sub>. После этого добавляли селектирующий агент.

### *3.2.7. Определение титра вирусных частиц*

Определение титра проводили методом, предложенным в работе (Вуун et al., 1996). Для этого клетки NIH/3T3 рассеивали в 24-луночный планшет в количестве 10000 на лунку, оставляли клетки для прикрепления в течение 2-3 часов. Затем делали последовательные 10, 100, 500 и 1000- кратные разведения вируса в среде с полибреном (8 мкг/мл), добавляли к клеткам и инкубировали при 32°C в атмосфере 5% CO<sub>2</sub>. Через 20 часов инкубации заменяли среду на свежую с добавлением G418 (1 мг/мл) и инкубировали 72 часа, после чего проводили подсчет выживших клеток.

### *3.2.8. Селекция клеток на среде с генетицином G 418*

Через 48 час после заражения к клеткам в среду добавляли 0,8 мг/мл генетицина G-418. Селекцию проводили 12-14 дней, смену среды проводили каждые 3 дня. После первого раунда селекции собирали клетки, выделяли ДНК и готовили библиотеку. С полученной библиотекой проводили второй раунд трансфекции-инфекции-селекции. Образцы ДНК, выделенные из клеток после второго раунда селекции, использовали для определения энхансерной активности выявленных фрагментов.

### *3.2.9. Получение библиотеки фрагментов ДНК локуса FXYD5-COX7A хромосомы 19 человека*

По 1 мкг смеси равных количеств ДНК из 30 индивидуальных космид расщепляли 20 ед. рестриктазы Sau3A или Csp6-I в буфере, рекомендованном производителями. Фрагменты космидной ДНК, полученные после расщепления каждой из рестриктаз, смешивали и лигировали с праймером и соответствующим адаптором в следующей смеси: 10 мкл смеси фрагментов ДНК, 2 мкл 10x буфера для лигирования (200 mM Трис-НСl pH 7,6, 100 mM MgCl<sub>2</sub>, 50 mM 2-меркаптоэтанола, 100 мкг/мл бычьего сывороточного альбумина, 10 mM АТР), по 2 мкл (1 нмоль) соответствующего адаптора и праймера, 8 ед. ДНК-лигазы фага

T4, в конечном объеме 20 мкл. Реакцию проводили в течение ночи при 13°C для гидролизатов Sau3A и 4°C для Csp6-I.

Структуры праймера и адапторов, используемых для получения библиотеки локуса:

5'-GATCTGTTCATGG-3' (адаптор для *Sau* 3A).

5'-GТАCTGTTCATGG-3' (адаптор для *Csp* 6-I).

5'-ACTGAGGTCGACTATCCATGAACA-3' (библиотечный праймер LP, используемый для получения библиотек для поиска энхансеров)

5'-ACTGAGCTCGAGTATCCATGAACA-3' (библиотечный праймер 27Pr, используемый для получения библиотек, предназначенных для поиска инсуляторов).

### 3.2.10. Радиоактивное мечение праймеров

Реакционную смесь в объеме 35 мкл, содержащую 50 пмоль олигонуклеотида, 12 мкл ( $\gamma$ -<sup>32</sup>P)АТР (120 мкКи), 20 ед. полинуклеотидкиназы фага T4 в соответствующем буфере, инкубировали при 37°C в течение 30 мин. Фермент инактивировали нагреванием при 70°C в течение 20 мин.

### 3.2.11. Очистка меченой ДНК

Меченые фрагменты ДНК разделяли в 10% полиакриламидном геле в 0,5x TBE буфере (45 mM трис-борат, 45 mM борная кислота, 1 mM EDTA, pH 8,0). После электрофоретического разделения и радиоавтографии меченой ДНК часть геля, соответствующую искомому фрагменту ДНК, вырезали и измельчали. Для элюции ДНК из геля добавляли 150 мкл 0,5 M ацетата аммония, 1% SDS, 1 mM EDTA и инкубировали при 37°C в течение ночи. Затем элюат отфильтровывали и переосаждали этанолом.

*Связывание меченых фрагментов ДНК с белками in vitro (Electrophoretic mobility shift assay, EMSA)*

Реакционная смесь содержала 2 мкл ~20 нг (50000 имп/мин) меченого с использованием полимеразной цепной реакции фрагмента ДНК, 2 мкл (2 мкг) альтернирующего двунитевого гетерополимера poly(dI-dC), 4 мкл буфера TE, 4 мкл (2 мкг суммарного белка) ядерного экстракта клеток HeLa, 12 мкл буфера (12 mM HEPES–KOH, pH 7,9, 12% глицерин, 60 mM KCl, 0,3 mM EDTA, 0,6 mM ДТТ). Смесь инкубировали в течение 20 минут при комнатной температуре.

По окончании инкубации смесь, содержащую ДНК-белковые комплексы, наносили на 7,5%-ный полиакриламидный гель (сшивка 30:1) приготовленный на 0,5x TBE. По прохождении ксиленцианолом 2/3 длины геля электрофорез останавливали, гель закрывали лавсановой пленкой и проводили радиоавтографию в течение 16-20 часов при +4°C.

При анализе способности последовательностей ДНК конкурировать с меченым фрагментом за связывание с белками ядерного экстракта HeLa добавляли в реакционную смесь соответствующий избыток немеченых конкурирующих фрагментов ДНК.

### *3.2.12. Определение активности люциферазы*

Для определения активности люциферазы проводили котрансфекцию клеток HeLa в 24-луночных культуральных планшетах плазмидами pGL3, содержащими потенциальные энхансеры, с нормирующей плазмидой pRL-TK с применением реагента Lipofectamine 2000 согласно рекомендациям фирмы-производителя.

Через 48 часов клетки лизировали в однократном буфере PLB и замораживали при -70°C. Измерение активности люциферазы проводили с использованием набора реактивов Dual-Luciferase Reporter Assay System (Promega, США) в соответствии с рекомендациями фирмы-изготовителя на люминометре GENios Pro (Tecan, Австрия). По результатам трех независимых трансфекций проводили компьютерный анализ полученных данных с использованием программы Microsoft Excel 2007 (Microsoft, США).



### *3.2.13. Фиксация и окрашивание клеток с помощью Coomassie Blue*

Для подсчета колоний, образующихся в результате селекции инфицированных клеток в 25 см<sup>2</sup> флаконах, их промывали 1хPBS, затем дважды раствором для фиксации (7% уксусная кислота, 20% метанол) и окрашивали раствором Coomassie Blue (0,25 % Coomassie Blue, 45 % метанол, 10 % ледяная уксусная кислота) в течение 15 мин при комнатной температуре. Окрашенные клетки промывали 2-3 раза раствором для фиксации и подсчитывали колонии.

### *3.2.14. Картирование последовательностей на геноме*

Сравнение последовательностей с международными базами данных проводили при помощи программы BLAST (Altschul et al., 1997) и сервера Национального института здравоохранения США ([www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/)). Для картирования использовали Human Genome Browser (<http://genome.ucsc.edu/cgi-bin/hgGateway>, (Kent, 2002)).

### *3.2.15. Метод двумерного EMSA*

Первое направление двумерного EMSA проводили так, как описано выше (см. раздел «Связывание меченых фрагментов ДНК с белками *in vitro*»). После радиоавтографии полоску геля (~ 0,3x12 см, толщина 0,7 мм), соответствующую полосе ДНК-белковых комплексов, вырезали из геля и инкубировали 20-30 минут при 50°C в 40 мл буфера, содержащего 0,1xTBE, 0,5% SDS. Полосу геля помещали на поверхность 5% полиакриламидного геля (сшивка 30:1), размером 20x20 см и толщиной 1 мм, содержащего 0,1% SDS, для разделения фрагментов ДНК в направлении, перпендикулярном первому. Электрофорез во втором направлении проводили при силе тока 15мА. После электрофореза проводили радиоавтографию в течение 16-40 часов при 4°C. Вырезали из геля полосу диагонали, соответствующую ДНК-белковому комплексу. Фрагменты, связывающиеся с белком, элюировали из геля и подбирали оптимальное количество циклов ПЦР для второго раунда селекции, как описано нами ранее (Vetchinova et al., 2006)

### *3.2.16. Иммунопреципитация хроматина*

Иммунопреципитацию хроматина проводили по описанному методу (Orlando, 2000). Клетки в стадии логарифмического роста фиксировали 1% по объему формальдегидом в ростовой среде в течение 10 мин при 37°C. Клетки промывали PBS, содержащим 1 мМ AEBSF и 1 мкг/мл коктейля ингибиторов протеиназ (Sigma, США), снимали раствором Версена в случае адгезивных культур, центрифугировали в течение 5 мин при 2000 об/мин, ресуспендировали в 200 мкл лизирующего буфера (50 мМ Трис-НСl pH 8,0, 1% SDS, 10 мМ EDTA) и инкубировали 10 мин на льду. Клетки обрабатывали ультразвуком при помощи дезинтегратора CP750 (Cole-Parmer) в течение 20 циклов (3 с импульс/10 с интервал) при амплитуде 30%. Клеточный дебрис удаляли центрифугированием (10 мин, 13000 об/мин, 4°C), супернатант разводили десятикратно буфером, содержащим 16,7 мМ Трис-НСl, pH 8,0, 16,7 мМ NaCl, 1,2 мМ EDTA, 1% Triton X-100, 0.01% SDS, 1 мМ PMSF, и 1 мкг/мл коктейля ингибиторов протеиназ. На этой стадии отбирали аликвоту для контроля (input chromatin). Клеточный лизат предварительно очищали инкубацией с частицами иммобилизованного на сефарозе А-белка и инкубировали в течение ночи при 4°C и перемешивании с 1,5 мкг кроличьих поликлональных антител против белка CTCF (ab10571, Abcam), и контрольных антител к растительному белку тауматину (любезно предоставлены Е.А. Стукачевой, ИБХ РАН) или бета-галактозидазе (D19, Roche Diagnostics).

Частицы сефарозы, несущие комплексы антитело-белок-ДНК, собирали центрифугированием, промывали и ДНК-белковые комплексы элюировали 2x15 мин при комнатной температуре 1% SDS, 0,1 М NaHCO<sub>3</sub>. К элюату добавляли NaCl до 0,2 М, РНК-азу и протеиназу К и инкубировали при 65°C в течение 4 час для обращения сшивок. Реакционную смесь очищали экстракцией смесью фенола и хлороформа, осаждали этанолом в присутствии 20 мкг гликогена, растворяли в воде и использовали для дальнейшего ПЦР-анализа.

### *3.2.17 Двумерный электрофорез с ультрафиолетовой сшивкой*

Смешивали меченый с использованием полимеразной цепной реакции двухцепочечный олигонуклеотид (около 200000 имп/мин) с 5 мкг poly-d(IC)-poly-d(IC) и 50 мкг тотального белка ядерного экстракта в объёме 60 мкл 12 мМ NERES–KOH, pH 7,9, 12 % глицерина, 60 мМ KCl, 0,3 мМ EDTA, 0,6 мМ ДТТ с добавлением 3 мМ леупептина, и инкубировали в течение 20 мин при комнатной температуре. Смесь наносили на 10% полиакриламидный гель, приготовленный на 0,5-X TBE, и разделяли в том же буфере (первое направление). Гель облучали УФ-лампой (254 нм, 6 мкВт/см<sup>2</sup>) в течение 3 мин, автордиографически определяли локализацию полосы первого направления, вырезали ее из геля и помещали в 50 мМ Трис-HCl, pH 6,8, 1% SDS, 2% 2-меркаптоэтанол, 0,025% бромфеноловый синий на 15 мин при 55°C. Затем полоску геля помещали на поверхность 12% полиакриламидного геля в присутствии SDS (Laemmli and Johnson, 1973) и белки, пришитые к меченому олигонуклеотиду, разделяли во втором направлении в соответствии с молекулярными массами. Олигонуклеотиды, пришитые к белкам незначительно изменяют их электрофоретическую подвижность (Nikolaev et al., 1988).

### *3.2.18. Приготовление ДНК-аффинного носителя*

Синтезированные олигонуклеотиды (примерно по 250 мкг каждого) очищали электрофорезом в ПААГ и растворяли в 100 мкл воды. По 100 мкл раствора каждого из олигонуклеотидов помещали в пробирку и добавляли 25 мкл 10х буфера для полинуклеотидкиназы (0,5 М Трис-HCl pH 7,9, 100 мМ MgCl<sub>2</sub>, 50 мМ ДТТ, 1 мМ спермидина, 1 мМ EDTA), 5 мкл 100 мМ АТР, и полинуклеотидкиназу (150 ед). Объем реакции доводили до 250 мкл, инкубировали при 37°C в течение 2 час, добавляли EDTA до 20 мМ, и инкубировали при 65°C в течение 15 мин. После этого производили отжиг с понижением температуры от 80°C до 4°C в течение ночи. К смеси добавляли NH<sub>4</sub>OAc до 2 М, 2,5 объёма этанола и инкубировали не менее 6 час при –20°C. Двухцепочечный олигонуклеотид осаждали, промывали метанолом, высушивали и растворяли в 200 мкл TE. К

раствору добавляли  $MgCl_2$  до 15 мМ, 25 мкл 3 М NaOAc pH 5,5 и 2,5 объёма этанола. Инкубировали ночь при  $-20^{\circ}C$ .

Олигонуклеотиды осаждали и растворяли в 100 мкл буфера для лигирования (40 мМ Трис-HCl pH 7,9, 10 мМ  $MgCl_2$ , 10 мМ ДТТ, 1 мМ спермидина). Добавляли 5 мкл 100 мМ АТФ и 2 мкл (40 Weiss units) ДНК-лигазы фага Т4. Инкубировали при  $12^{\circ}C$  в течение ночи и еще 2 часа при комнатной температуре. Степень лигирования проверяли, анализируя аликвоту в 2% агарозном геле.

Лигазную смесь экстрагировали один раз смесью фенола, хлороформа и изоамилового спирта (25:24:1), добавляли 10 мкл 3 М NaOAc pH 5,5, 2,5 объёма этанола, осаждали ДНК в течение ночи при  $-20^{\circ}C$ , дважды промывали метанолом, высушивали и растворяли в 10 мМ HEPES-Na pH 8,0.

10 мл сефарозы CL-2B (Pharmacia) промывали, ресуспендировали в 20 мл воды и помещали на водяную баню при  $15^{\circ}C$ . BrCN (1,1 г) растворяли в 2 мл ацетонитрила и по каплям добавляли к суспензии сефарозы при перемешивании на магнитной мешалке. Добавляли по каплям NaOH (5 М, около 1,8 мл) в течение 10 мин, поддерживая pH 10. Реакцию останавливали добавлением ледяной воды (100 мл) и немедленным промыванием на стеклянном фильтре 300 мл ледяной воды и 100 мл 10 мМ К-фосфатного буфера pH 8. Использовали немедленно, не давая просохнуть.

К промытой сефарозе добавляли 10 мМ К-фосфатный буфер pH 8,0 так, чтобы сформировалась плотная суспензия (около 0,4 мл). Затем добавляли олигонуклеотиды после лигирования, растворённые в буфере HEPES (около 0,5 мг в 100 мкл), инкубировали при перемешивании в течение 16 час при комнатной температуре. Сефарозу собирали на фильтре, промывали 200 мл 10 мМ К-фосфатного буфера pH 8,0, 100 мл 1 М этаноламина и ресуспендировали в 1М этанолаmine в конечном объёме 14 мл. Суспензию инкубировали 6 час при комнатной температуре при перемешивании. Сефарозу собирали, промывали на стеклянном фильтре 100 мл 10 мМ К-фосфатного буфера pH 8,0, 100 мл 1 М К-фосфатного буфера pH 8,0, 100 мл 1 М KCl, 100 мл TE и 100 мл буфера с азидом

натрия (10 мМ Трис-НСl pH 7,5, 1 мМ EDTA, 0,3 М NaCl, 0,02% (w/v) азид натрия). Хранили в том же буфере при 4°C.

Для проверки эффективности связывания 100 мкл осадка ДНК-сефарозы и 100 мкл осадка исходной сефарозы CL-2В промывали дважды 0,5 мл буфера для микрококковой нуклеазы (50 мМ Трис-НСl pH 7,9, 3 мМ CaCl<sub>2</sub>), ресуспендировали в 1 мл этого буфера, добавляли 0,5 ед. микрококковой нуклеазы и оставляли при перемешивании на 1 час при 37°C. Сефарозу осаждали центрифугированием, оптическую плотность супернатанта относительно контроля измеряли при 260 нм и рассчитывали количество ДНК, связавшейся с носителем. Количество пришитой ДНК составляло 5 мкг/мл носителя.

### *3.2.19. Фракционирование ядерного экстракта на ДНК-аффинной колонке*

Все процедуры проводили при 4°C. К 0,8 мл ядерного экстракта (концентрация тотального белка примерно 2,5 мг/мл) добавляли 0,4 мл воды, и разведённый экстракт наносили на проточно-охлаждаемую колонку с 0,5 мл гепарин-агарозы (Sigma), уравновешенной буфером С (12 мМ HEPES–KOH pH 7,9, 60 мМ KCl, 0,12 мМ EDTA, 0,3 мМ DTT, 0,3 мМ PMSF, 12% glycerol).

Нанесение проводили при скорости 0,1 мл/мин; проскок собирали и вновь наносили на колонку. Процесс повторяли 5 раз. Затем колонку промывали 5 объёмами буфера С, содержащего 0,2 М KCl и еще 5 мл буфера С. Белки, связавшиеся с гепарином, элюировали буфером С, содержащим 0,5 М KCl. Белки собирали фракциями по 3-4 мл и хранили при -70°C. Содержащие белок фракции объединяли, и доводили концентрацию KCl до 60 мМ. Затем фракции наносили на колонку со 100 мкл ДНК-сефарозы, уравновешенной буфером С. Процесс повторяли 5 раз. Колонку с ДНК-сефарозой промывали буфером С, содержащим 0,1 М KCl и еще 1 мл буфера С. Специфически связавшиеся с ДНК-сефарозой белки элюировали 500 мкл буфера С, содержащего 0,4 М KCl. Белок высаживали 12% ТХУ, промывали и растворяли в 20 мМ HEPES pH 7,9.

### *3.2.20. Фракционирование ядерного экстракта методом аффинной элюции*

Все процедуры проводили при 4°C. К 0,8 мл ядерного экстракта с концентрацией тотального белка примерно 2,5 мг/мл добавляли 0,4 мл воды, и разведённый экстракт наносили на проточно-охлаждаемую колонку и промывали так, как указано выше для ДНК-аффинной хроматографии. Для аффинной элюции, LTR-связывающие белки элюировали 5 мл буфера С с добавлением двухцепочечного олигонуклеотида С1/С2 (AGGATCGGGGCGGGGCGATC и СТGATCGCCCCGCCCCGATC), содержащего сайт связывания белкового комплекса, при конечной концентрации 30 нМ. Фракции собирали, объединяли, белок высаживали 12% ТХУ, промывали и перерастворяли в 20 мМ HEPES pH 7,9. Колонку промывали 5 мл буфера С и для повторного использования промывали 1М KCl и уравнивали 5 мл буфера С.

### *3.2.21. Обработка гелей, гидролиз трипсином и экстракция пептидов*

Обработку гелей, гидролиз трипсином и экстракцию пептидов проводили согласно протоколам (Shevchenko et al., 1996) с некоторыми модификациями. Из окрашенных Brilliant Blue R белковых полос вырезали фрагменты геля объемом 3 мм<sup>3</sup>, каждый фрагмент промывали 150 мкл 50 мМ NH<sub>4</sub>HCO<sub>3</sub> в 30% ацетонитриле в течение 45 мин при комнатной температуре и перемешивании. Супернатант отбрасывали, фрагменты геля высушивали на роторном испарителе. К каждому фрагменту геля добавляли 7 мкл раствора модифицированного свиного трипсина (6,25 нг/мкл; Promega, США) в 20 мМ NH<sub>4</sub>HCO<sub>3</sub>. Инкубировали при 4°C в течение 45 мин, а затем в течение ночи при 37°C. По окончании гидролиза фрагменты геля высушивали на роторном испарителе. Для удаления солей к фрагментам добавляли по 20 мкл деионизованной воды и вновь высушивали. Для экстракции пептидов к фрагментам геля добавляли по 20 мкл 0,1% раствора трифторуксусной кислоты в деионизованной воде и инкубировали при перемешивании в течение 20 мин при комнатной температуре. Супернатант отбирали в отдельные пробирки. Проводили повторную экстракцию, добавляя к фрагментам геля 20 мкл 0,1% трифторуксусной кислоты в 50% ацетонитриле, инкубировали в течение 40 мин при комнатной температуре. Соответствующие

супернатанты объединяли и концентрировали на роторном испарителе до конечного объема 1-2 мкл и затем добавляли до 5 мкл 0,1% раствор трифторуксусной кислоты в 50% ацетонитриле. Приготовленные образцы применяли для нанесения на мишень масс-спектрометра.

### *3.2.22. Анализ триптических пептидов при помощи масс-спектрометрии.*

0,2-1 мкл образца смешивали на мишени с таким же объемом раствора 2,5-дигидроксibenзойной кислоты (20 мг/мл в 20% ацетонитриле, 0,1% трифторуксусной кислоте), и высушивали на воздухе. В другом варианте, 0,2-1 мкл образца наносили на мишень и к капле немедленно добавляли такой же объем насыщенного раствора  $\alpha$ -циано-4-гидроксикоричной кислоты (Sigma, США; дважды перекристаллизованная из метанола) в 80% ацетонитриле, 0,1% трифторуксусной кислоте. Высушивали на воздухе. Масс-спектры получали на MALDI-TOF-масс-спектрометре Ultraflex, оборудованном PAN<sup>TM</sup> delay extraction system (Bruker, Германия) с УФ лазером (336 нм) в режиме положительных ионов в диапазоне масс от 500 до 8000 Да. Полученные масс-спектры калибровали, используя известные массы внутренних стандартов. Идентификацию белков по наборам значений масс пептидов после трипсинолиза проводили с использованием опции Peptide Fingerprint программы Mascot ([http://www.matrixscience.com/search\\_form\\_select.html](http://www.matrixscience.com/search_form_select.html)) с точностью определения массы  $MH^+$ , равной 0,01%, допуская возможность модификации цистеинов акриламидом и окисления метионинов. При поиске использовали базы данных NCBI (<http://www.ncbi.nlm.nih.gov>).

## 4. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

### 4.1. Стратегия идентификации энхансер-подобных элементов в протяженных областях сложных геномов.

Энхансеры – регуляторные элементы генома, способные усиливать транскрипцию генов, находясь на значительных расстояниях от них. Основываясь на способности большинства энхансеров активировать гетерологичные промоторы, нами предложена стратегия для выявления энхансер-подобных элементов в протяженных участках генома по способности этих элементов активировать минимальный промотор, контролирующий селективный ген. Общая схема стратегии показана на рис. 1.

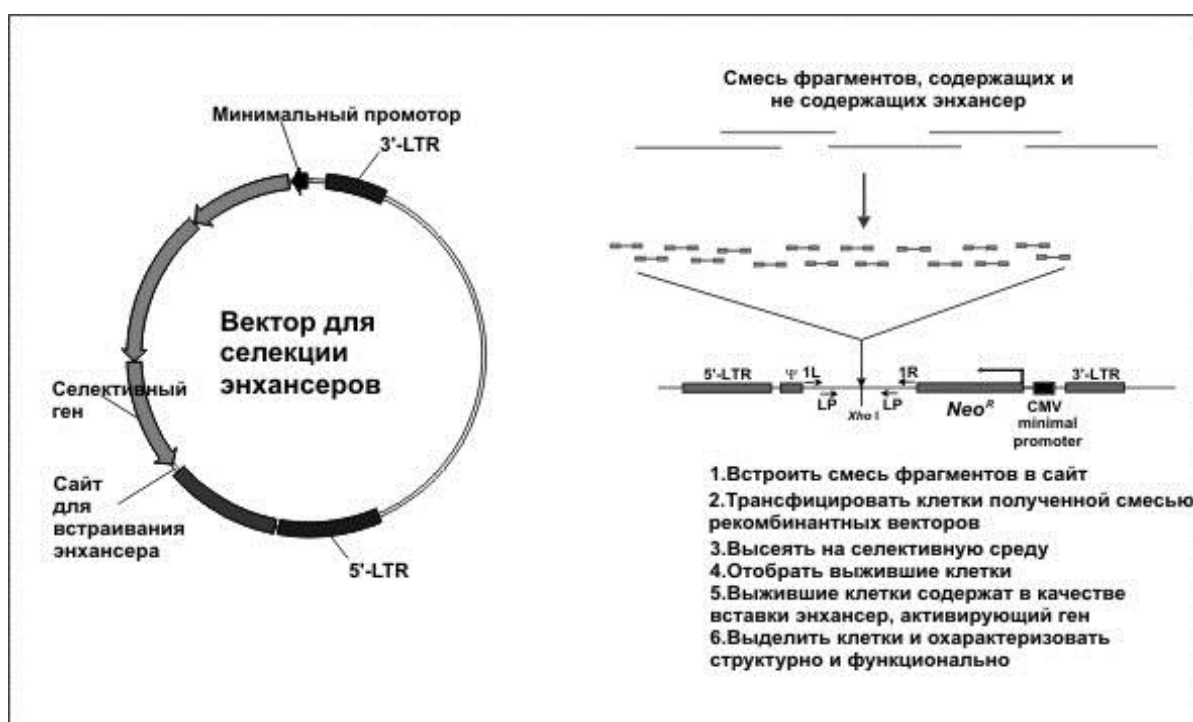


Рис. 1. Стратегия выявления энхансер-подобных элементов в протяженных участках генома.

Эту общую стратегию, разработанную совместно с Л.Г. Николаевым и И.П. Черновым, для практического использования следовало детализировать. Для селекции использовали ретровирусные векторы, преимущественно встраивающиеся в активно транскрибирующиеся участки генома, в результате



чего формируются стабильные клеточные линии, обладающие высоким уровнем экспрессии исследуемых генов и не содержащие внутренних перестроек.

Важным элементом стратегии является использование самоинактивирующегося ретровирусного вектора. Принцип его действия заключается в том, что после первого раунда репликации происходит удаление регуляторных элементов вектора, необходимых для осуществления следующего раунда репликации. Таким образом предотвращается влияние этих регуляторных элементов (промотор, энхансер) на репортерный ген.

#### **4.1.2. Экспериментальная идентификация и картирование энхансер-подобных элементов в локусе *FXVD5-COX7A1* хромосомы 19 человека**

Отбор потенциальных энхансеров проводили из фрагментов библиотеки локуса *FXVD5-COX7A1* хромосомы 19 человека.

#### *4.1.3. Создание векторных конструкций для селекции энхансер-подобных элементов*

Для создания конструкций, предназначенных для поиска энхансеров, мы использовали коммерческий экспрессионный самоинактивирующийся ретровирусный вектор pQCXIX (Clontech, США) и сконструированную Е.В. Снежковым плазмиду pCMV-SGTN, содержащую ген составного белка *GTN*, под контролем минимального промотора и энхансера цитомегаловируса (CMV). Ген *GTN* собран из фрагментов, кодирующих тимидинкиназу вируса простого герпеса (*HSV-tk*), неомицинфосфотрансферазу II (*NPTII*), придающую клеткам устойчивость к генетицину (G-418), и ген зеленого флуоресцентного белка EGFP. Из вектора pQCXIX обработкой эндонуклеазами рестрикции *EcoRV* и *XbaI* были удалены промотор и энхансер цитомегаловируса, внутренний сайт посадки рибосомы (IRES) и полилинкер. Одновременно плазмиду pCMV-SGTN обработали эндонуклеазой *NotI*, выступающие рестриктные концы достроили фрагментом Кленова ДНК-полимеразы I *E. coli*, и провели гидролиз эндонуклеазой *AsuNHI*. Часть плазмиды, содержащая минимальный промотор цитомегаловируса и ген *GTN*, была лигирована с фрагментом pQCXIX

*EcoRV-XbaI*. В полученную кассету по сайту *ClaI* был лигирован синтетический полилинкер (CGATCTAGACTCGAGAATTTC), содержащий сайты рестрикции эндонуклеаз *ClaI*, *XbaI*, *XhoI* и *EcoRI*. Полученную ретровирусную конструкцию назвали pQCXIX-ENH. Схема вектора pQCXIX-ENH представлена на рис. 2. Для предотвращения интерференции ретровирусных промоторов ген *NPTII* транскрибируется в направлении, противоположном тому, в котором транскрибируются ретровирусные гены. Структуру полученного вектора проверили секвенированием.

#### 4.1.4. Клонирование и селекция энхансер-подобных элементов

В полученную конструкцию клонировали фрагменты библиотеки локуса *FXYD5-COX7A1*. Кроме того, были приготовлены две контрольные плазмиды. Положительный контроль pQCXIX-Enh(+) содержал кассету *GTN* под контролем энхансера и промотора цитомегаловируса (CMV). Для этого плазмиду pCMV-SGTN гидролизовали *ClaI* и *AsuNI* и выступающие концы достраивали фрагментом Кленова. Фрагмент, содержащий ген *GTN*, минимальный промотор и энхансер цитомегаловируса, лигировали с фрагментом pQCXIX *EcoRV-XbaI*. Конструкция содержала в регуляторной области перед геном *GTN* минимальный промотор и энхансер цитомегаловируса. Отрицательный контроль pQCXIX-Enh(-) содержал кассету *GTN* без промотора и энхансера. Для этого плазмиду pCMV-SGTN обрабатывали *BamHI* и *AsuNI* и выступающие концы достраивали фрагментом Кленова. Фрагмент, содержащий только ген *GTN*, лигировали с фрагментом pQCXIX *EcoRV-XbaI*.

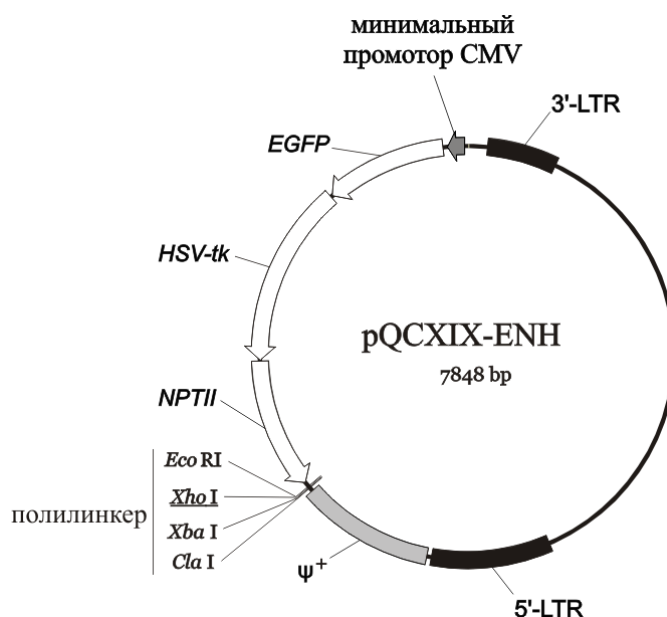


Рис. 2. Самоинактивирующийся ретровирусный вектор pQCXIX-ENH, предложенный для выявления энхансеров в протяженных геномных последовательностях. Вектор включает 5'- и 3'- длинные концевые повторы (*LTR*) и сигнал упаковки ( $\psi^+$ ). Фрагменты библиотеки лигировали в вектор по сайту рестрикции *XhoI*; *HSV-tk* – ген тимидинкиназы вируса простого герпеса; *NPTII* – ген неомизинфосфотрансферазы II, придающий клеткам устойчивость к генетицину (G-418); *EGFP* - ген зеленого флуоресцентного белка.

Для выявления фрагментов, обладающих энхансерными свойствами, использовали геномную библиотеку, содержащую фрагменты локуса *FXVD5-COX7A1* хромосомы 19 человека, приготовленную, как описано выше. Далее продукты ПЦР обрабатывали эндонуклеазой рестрикции *Sall* и лигировали с линейризованным по сайту *XhoI* вектором pQCXIX-Enh. В результате лигирования фрагментов библиотеки в вектор был получен пул плазмид pQCXIX-Enh(L).

Лигирование комплементарных выступающих концов, образованных в результате обработки ферментами *Sall* и *XhoI*, приводит к потере этих сайтов. В то же время, при лигировании может происходить замыкание вектора самого на себя без включения вставки, что приведет к восстановлению сайта *XhoI*. Поэтому перед трансформацией плазмидами клеток *E. coli* (штамм DH-5 $\alpha$ ) лигазную смесь обрабатывали эндонуклеазой *XhoI*. В результате трансформации клеток *E. coli* было получено около 16 тыс. колоний.

Для оценки эффективности лигирования проводили ПЦР-скрининг колоний с использованием праймеров 1L и 1R, последовательности которых фланкируют сайт *XhoI* в векторе pQCXIX-Enh (рис. 3А). 83% (10 из 12) анализированных клонов содержали вставки различной длины (рис. 3Б). Следовательно, из 16 тысяч выросших клонов свыше 13 тысяч содержат фрагменты библиотеки. Такого количества клонов вполне достаточно для представления всех уникальных последовательностей библиотеки в пуле плазмид pQCXIX-Enh(L), так как расчетное количество уникальных последовательностей в библиотеке локуса длиной 1 млн. п.о. при средней длине фрагментов 500 п.о. равно  $4 \times 10^3$  (длина локуса /средняя длина фрагментов библиотеки)  $\times$  2 (поскольку использовали две рестриктазы для приготовления библиотеки). Таким образом, расчетное количество выросших клонов, содержащих вставку ( $\sim 1,3 \times 10^4$ ), более чем в четыре раза превышает число уникальных последовательностей, что указывает на высокую репрезентативность библиотеки.

Из выросших клонов выделяли плазмидную ДНК, представляющую собой ретровирусные конструкции pQCXIX-Enh(L), содержащие фрагменты библиотеки локуса *FXVD5-COX7A1* хромосомы 19 человека.

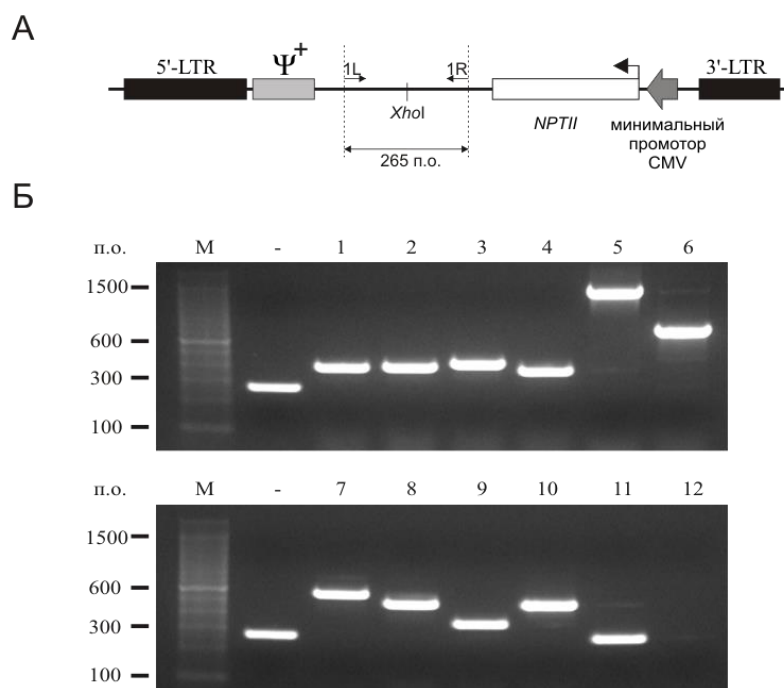


Рис. 3. Тестировка трансформированных pQCXIX-ENH(L) клонов на наличие вставки с использованием ПЦР. А) Линейризованная конструкция pQCXIX-ENH

(*XhoI* - уникальный сайт узнавания эндонуклеазы рестрикции, по которому лигировали фрагменты библиотеки участка FXVD5-COX7A1 хромосомы 19 человека). Б) Электрофореграмма продуктов ПЦР-амплификации клонов, трансформированных pQCXIX-ENH(L). Для ПЦР использовали праймеры 1L и 1R, фланкирующие сайт *XhoI*. M – маркер, (-) – ПЦР-амплификация конструкции pQCXIX-ENH (контрольный вектор без вставки). Цифрами обозначены номера клонов (11 и 12 не содержали вставки).

Далее проводили отбор последовательностей ДНК, обладающих энхансерной активностью. Схема предложенного нами метода селекции энхансер-подобных элементов представлена на рис. 4. Клетки пакующей линии Phoenix-AMPHO (Kinsella and Nolan, 1996) трансфицировали пулом плазмид pQCXIX-Enh(L), а также контрольными конструкциями. Линия клеток Phoenix-AMPHO представляет собой популяцию клеток 293-T, способных экспрессировать вирусные белки, необходимые для образования амфотропных вирусных частиц. После попадания ретровирусных конструкторов в клетки они способны продуцировать два типа транскриптов, один из которых представляет собой мРНК гена *GTN*, второй, более протяженный, синтезируется в противоположном направлении и находится под контролем промотора, расположенного в 5'-длинном концевом повторе (5'LTR). Этот транскрипт содержит такие ретровирусные элементы, как фрагменты длинных концевых повторов и сигнал упаковки  $\psi^+$ , благодаря которому он упаковывается в вирусные частицы, а после проникновения в клетку и процесса обратной транскрипции его кДНК-копия интегрирует в геном хозяйской клетки.

Всего было получено четыре популяции вирусных частиц pQCXIX-Enh, pQCXIX-Enh(L), pQCXIX-Enh(+) и pQCXIX-Enh(-). Для определения концентрации частиц, способных инфицировать клетки использовали, экспресс-метод *Byun et al.* (Byun et al., 1996). Метод позволяет определить нижний предел значений титра вируса путем инфицирования клеток NIH3T3 с их последующей селекцией на среде, содержащей генетицин G418. Высокий титр вирусных частиц pQCXIX-Enh(L), содержащих фрагменты библиотеки, важен для сохранения репрезентативности библиотеки. В наших экспериментах для конструкций pQCXIX-Enh(L) титр составил не менее  $5 \times 10^3$  частиц/мл. Фактически титр был

значительно выше, так как не все вирусные частицы популяции pQCXIX-Enh(L) способны обеспечить зараженные клетки необходимым уровнем неомифосфотрансферазы II (G<sup>TN</sup>) из-за отсутствия энхансерной активности у части фрагментов библиотеки.

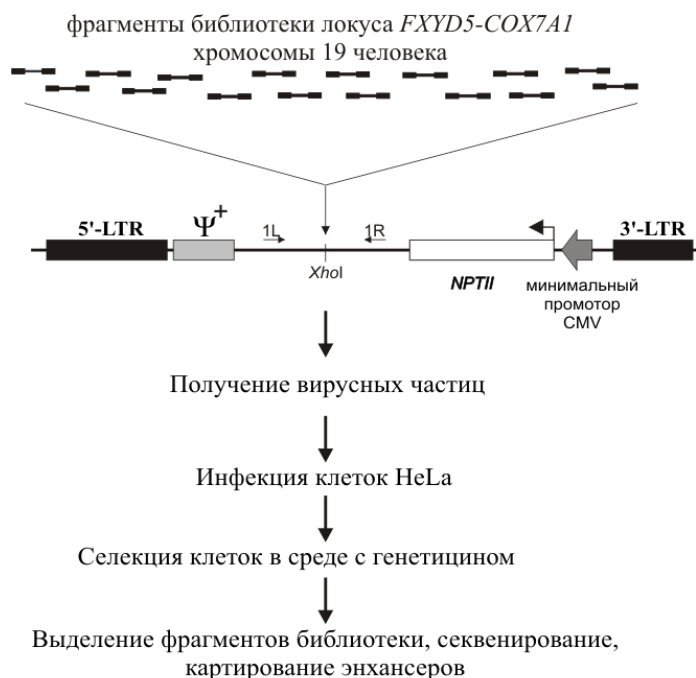


Рис. 4. Общая схема селекции последовательностей, обладающих энхансерной активностью. Фрагменты библиотеки локуса *FXVD5-COX7A1*, полученной путем расщепления ДНК 30-ти космид, содержащих перекрывающиеся последовательности фрагмента хромосомы 19 человека длиной 1 млн. п.о., клонировали в конструкцию pQCXIX-Enh(+), по сайту XhoI. Полученным набором конструкций (pQCXIX-Enh(L)) трансфицировали клетки пакующей линии Phoenix-AMPHO для получения вирусных частиц. Вирусными частицами инфицировали клетки HeLa и культивировали в среде с генетицином G418. Из клеток, прошедших селекцию, выделяли геномную ДНК с фрагментами библиотеки, обладающими энхансерной активностью. Выделенные фрагменты секвенировали и определяли их положение относительно генов локуса хромосомы 19 человека.

Полученными вирусными частицами трансдуцировали клетки линии HeLa. Инфицированные клетки культивировали в среде с добавлением 800 мкг/мл G418 в течение двух недель. Концентрацию селективного агента определяли методом титрования. Использование G418 в концентрации 800 мкг/мл вызывает полную гибель интактных клеток HeLa примерно за 10-12 дней. Эту концентрацию антибиотика использовали в дальнейшем.

При инфицировании вирусными частицами происходит интеграция кДНК-копии вирусной мРНК в геном хозяйской клетки. В случае, если интегрированный провирус содержал фрагмент библиотеки, проявляющий энхансерную активность, он увеличивал активность гена селективного маркера (*NPTII* в составе белка GTN), и клетки, содержащие такую конструкцию, выживали в среде с генетицином. После раунда селекции из клеток выделяли геномную ДНК и анализировали при помощи двустадийной ПЦР (nested-PCR) с использованием а) плазмидных праймеров (1L и 1R) и б) с помощью библиотечного праймера 27Pr (рис.5А, раунд 1).

Интеграция ретровирусной ДНК происходит преимущественно в активно транскрибирующиеся области генома (Scherdin et al., 1990), поэтому гены, находящиеся в составе ретровирусной кассеты, после интеграции в геном могут оказаться под влиянием эндогенных *цис*-регуляторных элементов. При воздействии эндогенного энхансера будут отобраны фрагменты библиотеки, не обладающие энхансерной активностью (ложно-позитивные). Подобный результат может быть получен и в случае множественной интеграции, когда в геном одной клетки встраивается несколько конструкций, среди которых хотя бы одна попадает под воздействие энхансера.

Для уменьшения числа ложнопозитивных последовательностей проводили второй раунд селекции. После первого раунда фрагменты библиотеки, амплифицированные с геномной ДНК популяции рQСХІХ-Enh(L) (рис. 5А, раунд 1), лигировали в вектор рQСХІХ-Enh, которым трансформировали клетки *E.coli*. Количество выросших на этом этапе колоний составило примерно 10 тысяч. Из них выделяли плазмидную ДНК, представляющую набор ретровирусных конструкций с фрагментами библиотеки, прошедшими отбор в первом раунде. Полученным набором ретровирусных конструкций, а также контрольными конструкциями, трансфицировали клетки Phoenix-AMPHO для получения вирусных частиц, которыми затем инфицировали клетки HeLa.

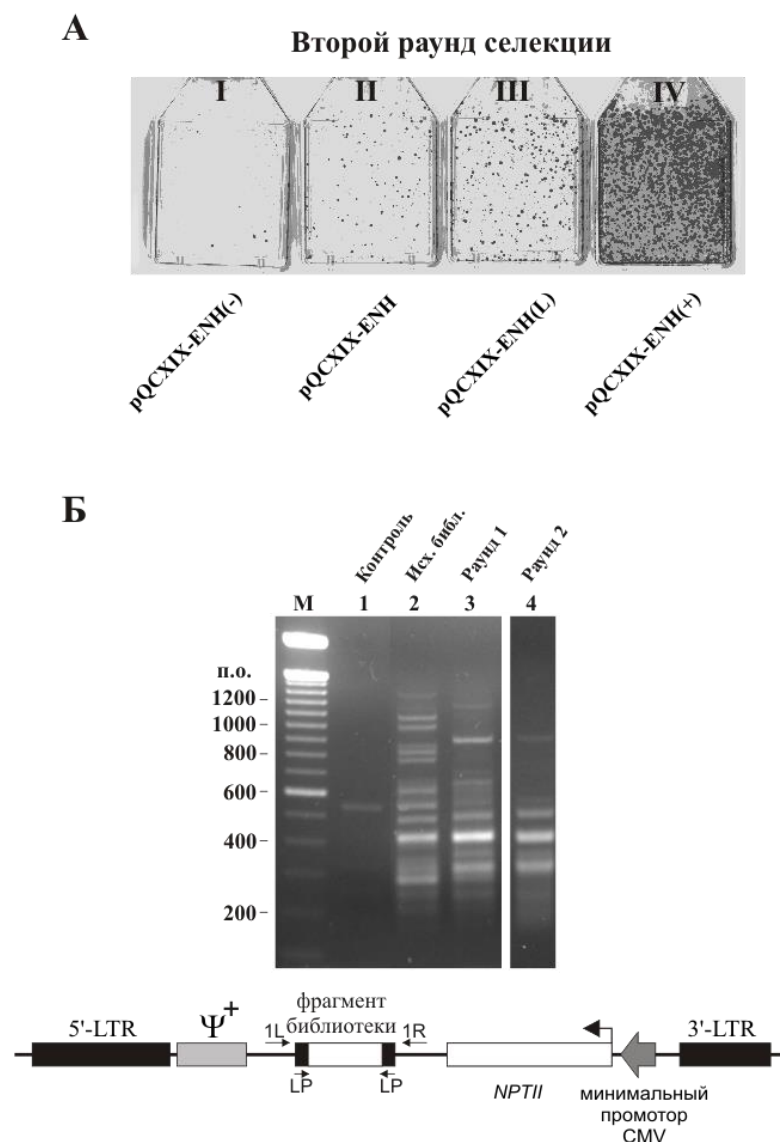


Рис. 5. Отбор потенциальных энхансеров из фрагментов библиотеки локуса *FXVD5-COX7A1*, расположенного на хромосоме 19 человека. А) Четыре популяции инфицированных клеток HeLa после второго раунда селекции (окраска Coomassie Blue). Б) Электрофорез в агарозе продуктов двухстадийной ПЦР-амплификации фрагментов библиотеки: дор. 2 - исходной библиотеки ; дор. 3 - после первого раунда; дор. 4 - после второго раунда селекции. На первой стадии амплификации использовали внешние праймеры 1L и 1R, на второй стадии использовали праймер LP, фланкирующий фрагменты библиотеки.

После инфицирования клеток HeLa каждую популяцию при пассировании высевали на два флакона и проводили селекцию в среде с генетицином на протяжении примерно двух недель. После селекции по одному из двух флаконов с популяциями клеток, инфицированных контрольными и ретровирусными конструкциями, содержащими фрагменты библиотеки, использовали для подсчета селектированных колоний. Для этого клетки во флаконах фиксировали



при помощи метанола и уксусной кислоты и окрашивали Coomassie Blue (рис. 5). Клетки из оставшихся флаконов собирали и выделяли геномную ДНК для дальнейшего анализа. Из рис. 5А видно, что наибольшее количество выживших после селекции и образующих колонии клеток содержит популяция, инфицированная позитивным контролем pQСХІХ-Enh(+), содержащим энхансер и промотор цитомегаловируса в регуляторной области гена *GTN*. Популяция клеток, инфицированная конструктором pQСХІХ-Enh(-), содержащим только минимальный промотор (негативный контроль), содержит гораздо меньше выживших клеток. Флакон с клетками, в которых проходил отбор фрагментов библиотеки pQСХІХ-Enh(L), содержит также меньше клеток по сравнению с позитивным контролем, но больше чем в популяции, инфицированной конструкцией с негативным контролем pQСХІХ-Enh(-), что свидетельствует о присутствии среди фрагментов библиотеки последовательностей, обладающих энхансерной активностью. Среди клеток, инфицированных конструкцией pQСХІХ-Enh(-), выживают единичные клетки, несмотря на то, что сама конструкция pQСХІХ-Enh(-) содержит ген *GTN* без энхансера и промотора. Вероятно, интеграция в этих клетках происходила вблизи активного промотора. Подобные события весьма редки, что и демонстрирует малое количество выросших колоний.

Геномную ДНК, выделенную из клеток после второго раунда селекции, анализировали при помощи двустадийной ПЦР (рис. 5Б, раунд 2). Как видно из рисунка, на дорожке 4 (раунд 2) выявляется меньше фрагментов ДНК, чем в исходной библиотеке (дор. 2) и в библиотеке после первого раунда (дор. 3), что свидетельствует об отборе последовательностей, обладающих энхансерной активностью.

#### 4.1.5. Анализ отобранных потенциальных энхансерных фрагментов

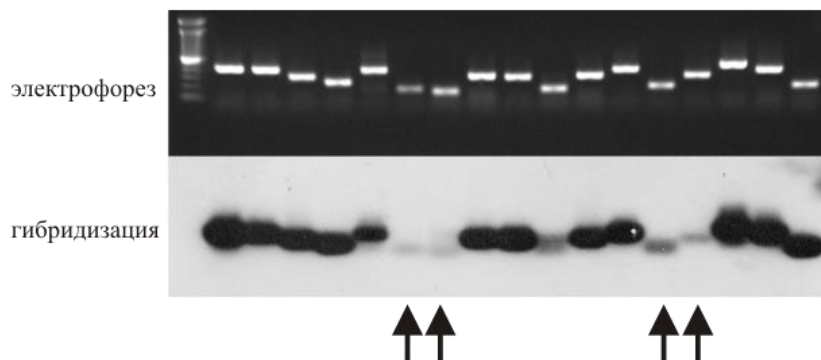
Фрагменты библиотеки, амплифицированные после второго раунда селекции, лигировали в плазмидный вектор pGEM-T Easy (Promega) и трансформировали ими клетки *E.coli*. Полученные клоны ранжировали в 96-

луночные планшеты. Полученную клонотеку потенциальных энхансеров использовали для дальнейшего анализа.

Среди выделенных фрагментов, невзирая на два раунда селекции, часть могла быть представлена последовательностями, не являющимися энхансерами. Для отсеечения этих последовательностей, доля которых среди отобранных фрагментов незначительна, проводили анализ ранжированных клонов с помощью блот-гибридизации по Саузерну. С этой целью проводили ПЦР-амплификацию ДНК индивидуальных клонов при помощи праймера LP. Для уменьшения эффекта ПЦР селекции фрагментов использовали минимальное количество циклов. Продукты ПЦР разделяли в 1% агарозном геле (рис. 6А, сверху), переносили на нейлоновую мембрану и проводили гибридизацию с меченой ( $\alpha$ - $^{32}\text{P}$ )-dATP порцией тотальной ДНК, содержащей фрагменты библиотеки, амплифицированные после двух раундов селекции (рис. 6А, снизу). После гибридизации отбирали клоны с наиболее сильным гибридизационным сигналом. Такие клоны содержали последовательности библиотеки, высоко представленные во фракции ДНК, полученной после двух раундов селекции. Клоны со слабым гибридизационным сигналом исключали из дальнейшего анализа.

Из отобранных клонов выделяли плазмиды со вставками и секвенировали. Среди 50 секвенированных клонов было обнаружено 15 последовательностей, 6 из которых встречались при анализе полученных данных по одному разу, 2 последовательности - по два раза, 4 – три раза, 2 – по четыре раза и одна последовательность двенадцать раз. На рис. 6Б приведен график зависимости числа обнаруженных уникальных последовательностей от общего числа секвенированных клонов. Кривая на графике выходит на плато, и показывает, что идентификация новых энхансеров при дальнейшем секвенировании библиотеки маловероятна. Свойства идентифицированных последовательностей представлены в таблице 4.

А



Б

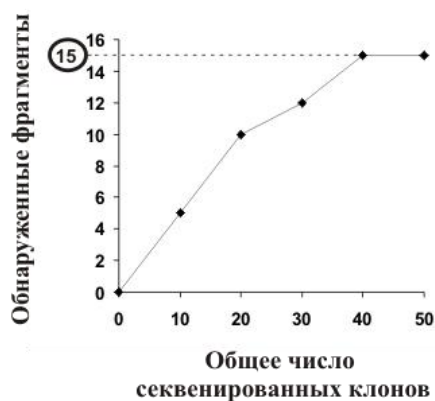


Рис. 6. Анализ библиотеки потенциальных энхансеров. А) Гибридизация индивидуальных клонов с тотальной библиотекой, полученной после двух раундов селекции (приведена часть проанализированных клонов). Верхняя панель – электрофорез в 1% агарозном геле продуктов ПЦР-амплификации фрагментов библиотеки, выделенных из индивидуальных клонов. Нижняя панель – результат гибридизации амплифицированных фрагментов с тотальной библиотекой, полученной после двух раундов селекции. Клоны, давшие слабый гибридизационный сигнал (отмечены стрелками), исключали из дальнейшего анализа. Б) Оценка числа независимых клонов в библиотеке. Приведена зависимость числа выявленных уникальных последовательностей от общего числа секвенированных клонов.

#### 4.1.6. Способность отобранных фрагментов связываться с клеточными белками

Далее была определена функциональная активность полученных фрагментов.

Отличительной особенностью функциональных элементов, в том числе энхансеров, является их способность формировать ДНК-белковые комплексы.

Большинство моделей функционирования энхансеров предполагает образование комплекса с белками, взаимодействие с промоторным комплексом, и

выпетливание ДНК между ними. Мы проверили способность нескольких выявленных энхансер-подобных элементов взаимодействовать с ядерными белками при помощи метода сдвига электрофоретической подвижности (EMSA), позволяющего выявлять ДНК-белковые взаимодействия. EMSA-анализ проводили с использованием ядерного экстракта из клеток HeLa, поскольку отбор энхансеров проводили с использованием этой клеточной линии.

Для анализа выбрали фрагменты 7, 8, 10, 11 и 12. Энхансер цитомегаловируса использовали в качестве положительного контроля. ДНК радиоактивно метили при помощи ( $\alpha$ - $^{32}$ P)-dATP и инкубировали с ядерным экстрактом. Для доказательства специфичности ДНК-белковых взаимодействий в реакционную смесь с мечеными фрагментами ДНК добавляли в качестве конкурирующих те же немеченые фрагменты в трех- и шестикратном избытке (рис. 7). В случае специфичного взаимодействия немеченные фрагменты конкурируют с мечеными за связывание с белками, что приводит к ослаблению интенсивности зон торможения ДНК-белковых комплексов. Все проанализированные последовательности с высокой специфичностью образовывали комплексы с белками ядерного экстракта HeLa. Использование неспецифического конкурента (фрагмент промоторной области гена *c-myc* человека), взаимодействующего с ядерным белком CTCF (Filippova et al., 1996), не оказывало влияния на способность анализируемых фрагментов образовывать комплексы ДНК-белок (данные не представлены).

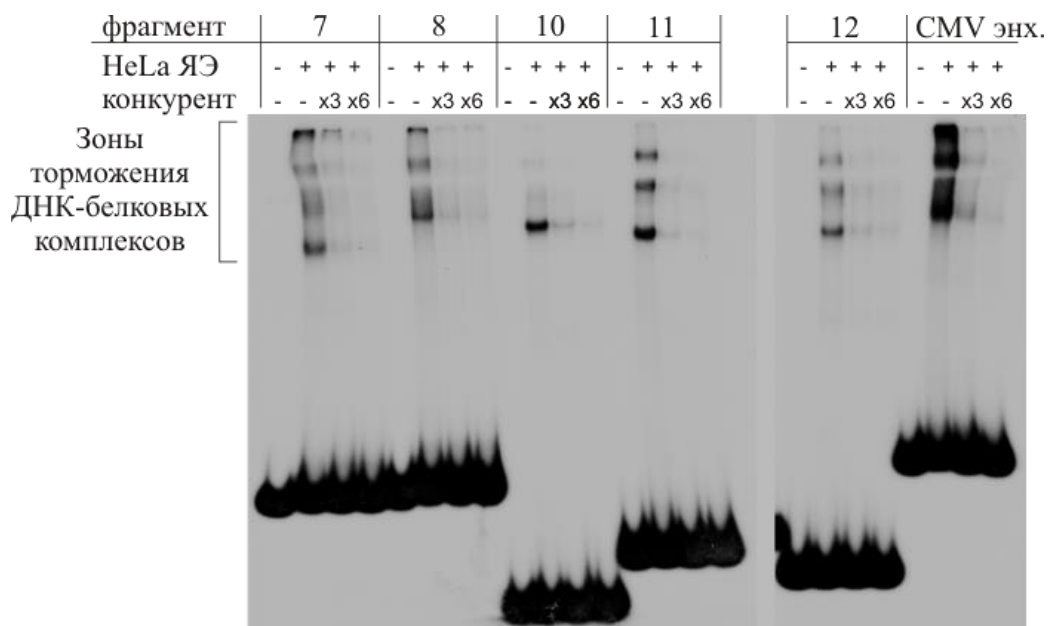


Рис. 7. Проверка связывания обнаруженных потенциальных энхансеров с белками ядерного экстракта (ЯЭ) из клеток линии HeLa методом сдвига электрофоретической подвижности (EMSA).

Из рис. 7 видно, что все исследованные фрагменты образуют одну или несколько выраженных зон торможения, что указывает на способность этих фрагментов взаимодействовать с белковыми компонентами ядерных экстрактов. ДНК-белковые комплексы некоторых фрагментов отличаются по электрофоретической подвижности. Вероятно, эти фрагменты образуют комплексы разного состава или же с белками разного размера.

#### 4.1.7. Способность отобранных фрагментов активировать минимальный промотор с репортерным геном

Энхансерную активность 15 выявленных последовательностей измеряли с использованием двойной люциферазной системы (Promega). Плазмиды, содержащие 15 потенциальных энхансеров, обрабатывали *SalI*, поскольку каждый фрагмент фланкирован сайтом узнавания *SalI*, расположенным в праймере LP. Далее 15 фрагментов с выступающими концами лигировали в линейаризованный по сайту *SalI* вектор pGL3 promoter vector (pGL3pv, Promega).

Полученными конструкциями проводили транзientную трансфекцию клеток HeLa с использованием реагента Lipofectamine 2000. В качестве контролей

использовали pGL3pv без вставки и pGL3 control vector (pGL3cv), содержащий промотор и энхансер SV40.

Из пятнадцати проанализированных потенциальных энхансеров тринадцать проявляли энхансерную активность в данной системе, усиливая активность гена люциферазы в 2-5 раз, что соответствует 15-30% от активности сильного энхансера SV40 (рис. 8). Таким образом, около 85 % (13 из 15) прошедших отбор фрагментов библиотеки проявляют энхансерную активность, что указывает на высокую эффективность предложенного подхода.

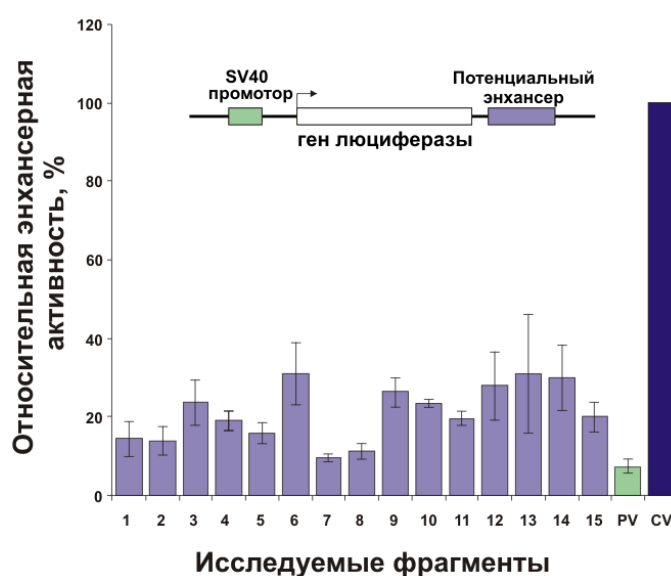


Рис. 8. Способность выявленных потенциальных энхансеров активировать промотор SV40 в клетках HeLa. Приведены результаты, полученные на основе трех независимых трансфекций. Активность потенциальных энхансеров представлена в процентах от активности энхансера SV40, принятой за 100%. PV – в трансфекции использовали контрольный вектор pGL3pv, содержащий ген люциферазы под контролем промотора SV40. CV - в трансфекции использовали контрольный вектор pGL3cv, содержащий ген люциферазы под контролем промотора и энхансера SV40.

#### 4.1.8. Анализ расположения потенциальных энхансеров в геноме. Построение карты расположения энхансеров

Характеристика обнаруженных потенциальных энхансеров представлена в таблице 4. Положение энхансер-подобных элементов в исследуемом локусе изображено на рис. 9. Шесть фрагментов были локализованы внутри генов (в интронах), пять фрагментов располагаются с 5'-стороны от генов, один

предполагаемый энхансер расположен с 3'-стороны от гена и три фрагмента в межгенных областях (ближайшие гены находятся на расстоянии более 10 т.п.о.). Представляет интерес то, что два из трех потенциальных энхансеров, выявленных в межгенных областях, перекрываются с повторяющимися элементами, Alu-элементом и длинным концевым повтором семейства HERV-K эндогенных ретровирусов человека. Поскольку гены (включая интроны) составляют не более 50% всей длины локуса, можно заключить, что обнаруженные потенциальные энхансеры преимущественно располагаются внутри, либо вблизи генов локуса.

Таблица 4 Характеристика потенциальных энхансеров, выявленных в области *FXYD4-COX7A1* хромосомы 19 человека.

№ Энхансера	Координаты на хромосоме 19**	Длина, п.о.	Положение относительно генов локуса	Активность*	Перекрытие с повторяющимися элементами
1	40341821-40342118	298	Интрон 3 гена <i>FXYD5</i>	2.2	
2	40344506-40344897	392	Интрон 4 гена <i>FXYD5</i>	2.1	
3	40427994-40428292	299	~2 т.п.о. от 5'-конца гена <i>LSR</i>	3.5	1-232 Alu-Y
4	40634421-40635308	888	1 т.п.о. от 3'-конца гена <i>FFAR2</i>	2.8	
5	40718720-40719018	299	Интрон 1 гена <i>GAPDHS</i>	2.3	
6	40748237-40748490	254	2 т.п.о. от 5'-конца гена <i>ATP4A</i>	4.7	
7	40757795-40758182	388	Межгенное***	1.4	1-388 HERV-K LTR
8	40763469-40763799	331	Межгенное***	1.6	
9	40764184-40764486	303	Межгенное***	3.9	1-240 Alu-Y
10	40792389-40792543	156	3 т.п.о. от 5'-конца гена <i>KIAA0841</i>	2.8	
11	40812850-40813109	260	Интрон 2 гена <i>RBM42</i>	2.8	1-220 Alu-Sc
12	40927410-40927666	257	Интрон 2 гена <i>U2AFIL4</i> ; 0.7 т.п.о. от 5'-конца гена <i>PSENN</i>	4.3	
13	40965824-40966019	196	Интрон 15 гена <i>SNX26</i>	4.9	
14	41091517-41091714	198	0.5 т.п.о. от 5'-конца гена <i>TYROBP</i>	4.3	
15	41110960-41111236	277	8 т.п.о. от 5'-конца гена <i>LRFN3</i>	2.9	70-277 Alu-Sx

\*Данные по транзитным трансфекциям в двойной люциферазной системе детекции (Рис. 8).

\*\*Данные Human Genome Browser (сборка от марта 2006 г.).

\*\*\*Расстояние > 10<sup>3</sup> п.о. от ближайших генов.





Некоторые из обнаруженных последовательностей (фрагменты 3, 9, 11 и 15) перекрываются с ретроэлементами, в частности Alu-элементами и длинным концевым повтором (LTR) семейства HERV-K (фрагмент 7). Энхансерная активность LTR HERV-K будет описана ниже (Domanskii et al., 2002; Ruda et al., 2004; Ilarionova et al., 2007), и выявление среди потенциальных энхансеров фрагмента, расположенного в этом классе повторяющихся элементов генома, лишь подтверждает достоверность предложенного метода. Гораздо меньше накоплено данных об энхансерной активности Alu-элементов. Так, было показано, что энхансер гена *BRCA1* представляет собой Alu-элемент со свойствами эстроген-зависимого энхансера (Norris et al., 1995).

Два энхансера было выявлено вблизи ранее обнаруженных активных промоторов (Kim et al., 2005). Впрочем, не обязательно указанные энхансеры участвуют в регуляции именно этих промоторов, так как действие энхансера может распространяться на значительные расстояния.

#### *4.1.9. Функциональный анализ активности потенциального энхансера “U2AF1L4”*

После выявления участков ДНК генома человека, проявляющих энхансерную активность в составе ретровирусной конструкции, интегрированной в геном клеток HeLa, мы провели детальную функциональную диссекцию одного из обнаруженных участков, а именно, энхансера 12 (Табл. 4), и попытались подробно проанализировать энхансерную активность этого участка и его фрагментов в системе экспрессии репортерного гена.

Как видно из рис. 10, энхансер находится во втором интроне гена *U2AF1L4* на хромосоме 19 человека (далее энхансер “*U2AF1L4*”). В этом локусе отмечается кластеризация сайтов связывания белков, таких как инсуляторный белок CTCF и компоненты комплекса SWI/SNF (BAF155, BAF170 и Ini1) (Celniker et al., 2009).

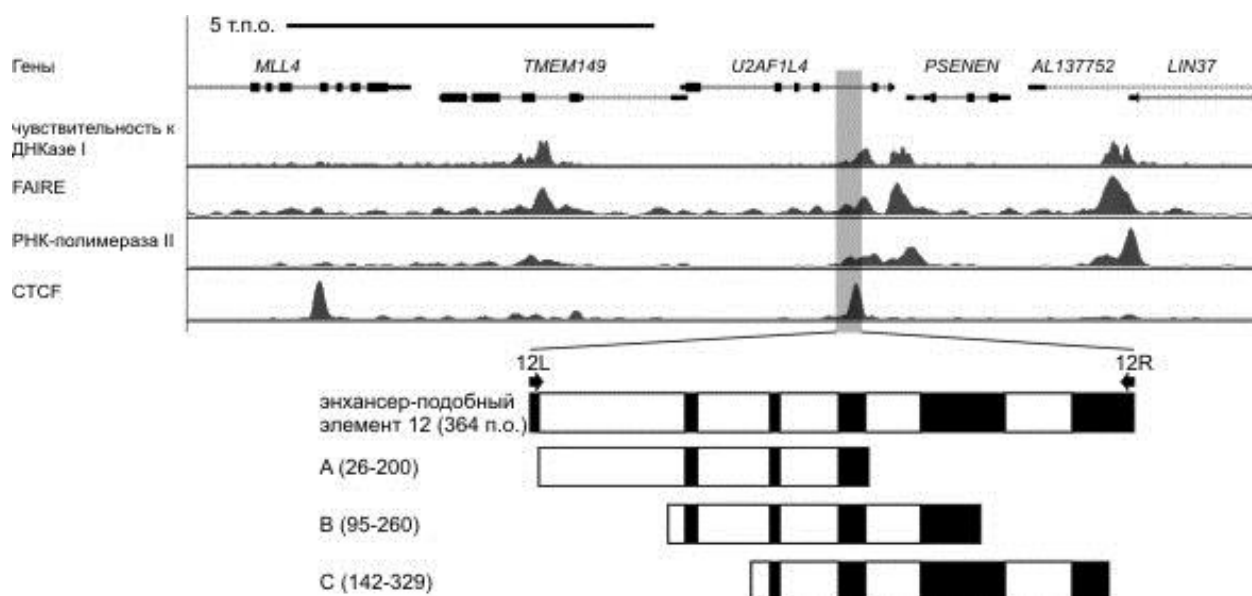


Рис. 10. Позиция энхансера “*U2AF1L4*” относительно ближайших генов на хромосоме 19 человека. Представлены данные геномного браузера UCSC (<http://genome.ucsc.edu/goldenPath/hgTracks.html>) по плотности распределения в геноме клеток HeLa участков чувствительности к ДНК-азе I, регуляторных участков, обнаруженных с помощью метода выявления свободных от нуклеосом фрагментов ДНК (FAIRE – Formaldehyde Assisted Isolation of Regulatory Elements), участков, взаимодействующих с РНК-полимеразой II и инсуляторным белком CTCF. Серым цветом отмечена область, содержащая энхансер “*U2AF1L4*”. Внизу показаны полноразмерный энхансер “*U2AF1L4*” и его перекрывающиеся делеционные фрагменты (А, В и С), для которых проведен функциональный анализ энхансерной активности (рис. 11).

Полноразмерную последовательность и три перекрывающихся фрагмента энхансера клонировали в двух ориентациях относительно промотора в вектор pGL3-PV, содержащий ген люциферазы светлячка под контролем промотора SV40 (рис. 11) для определения участка энхансера “*U2AF1L4*”, ответственного за его энхансерную активность.

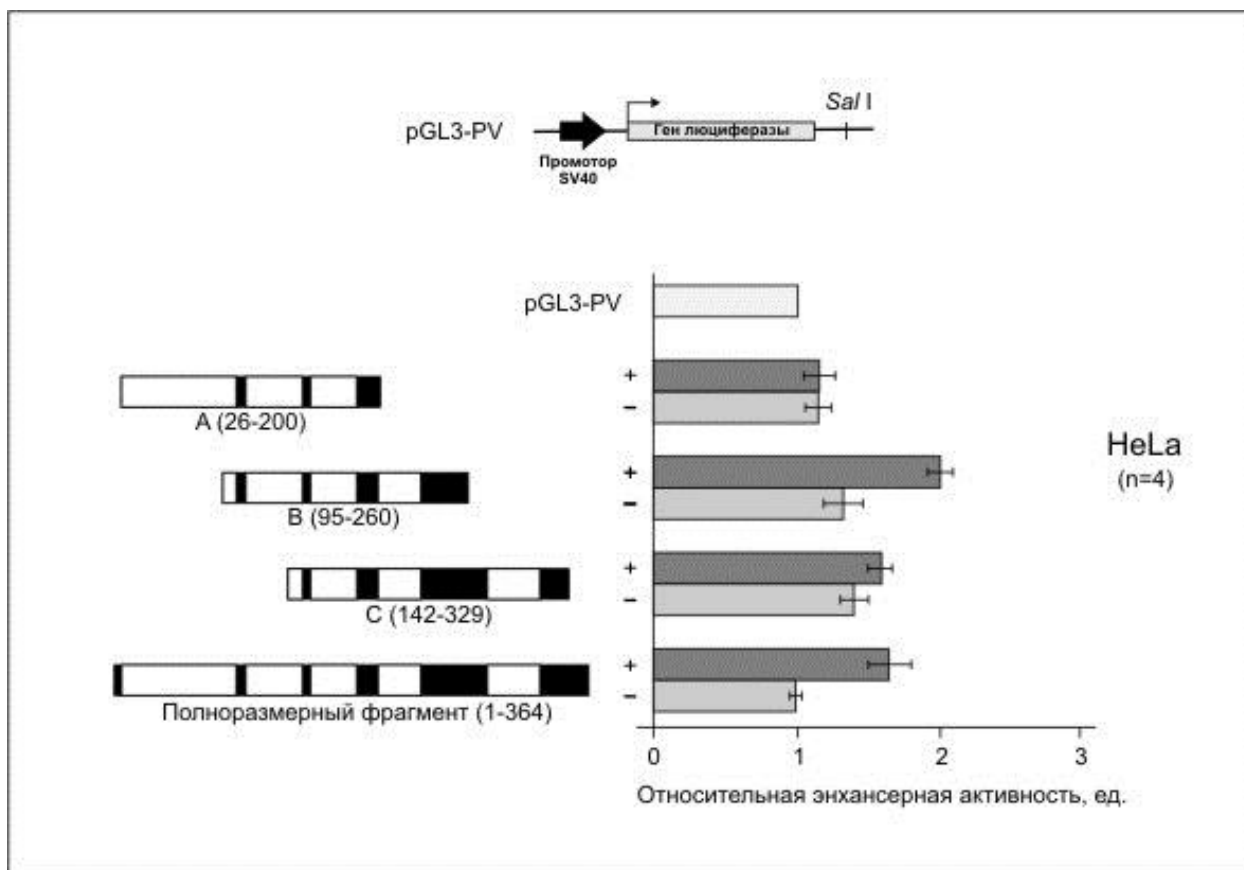


Рис. 11. Исследование энхансерной активности полного энхансера “*U2AF1A*” и его фрагментов в прямой (+) и обратной (–) ориентациях в клеточной линии HeLa с использованием двойной системы люциферазной детекции. Исследуемые фрагменты клонировали в вектор pGL3pv (сверху представлена схема вектора в линейаризованном виде) по сайту *Sal I* в обеих ориентациях. Значения активности фрагментов указаны относительно активности промотора SV40 (pGL3pv). Показана стандартная ошибка среднего для четырех независимых экспериментов (n=4).

Клетки HeLa трансфицировали полученным набором плазмид, а также исходным вектором pGL3-PV, используемым в качестве контрольного (рис. 11). Фрагмент В проявлял максимальную активность, усиливая активность промотора SV40 в два раза, что указывает на наличие в этом фрагменте участка, ответственного за энхансерную активность. Фрагмент С содержит участок связывания инсуляторного белка CTCF который способен взаимодействовать с сайленсерами и. таким образом снижать активность полноразмерной последовательности энхансера, содержащей фрагмент С (Klochkov et al., 2006). Основная активность энхансер-подобного элемента сосредоточена в 3'-области фрагмента В. Относительно невысокий уровень энхансерной активности можно

объяснить тем, что в клетках HeLa сильный промотор SV40 слабо подвержен модуляции со стороны клеточных энхансеров, и только в присутствии сильных энхансеров его активность может значительно повышаться.

Активность энхансерных фрагментов зависит от их ориентации относительно промотора. Полноразмерный энхансер “*U2AF1L4*”, а также его фрагменты В и С проявляли наибольшую активность в прямой ориентации (рис. 11).

#### **4.2. Стратегия экспериментального поиска инсуляторов в протяженных областях сложных геномов**

Инсуляторы – регуляторные элементы генома, способные блокировать взаимодействие между энхансером и промотором, если находятся между ними. Инсуляторы также защищают фланкированный ими транскрипт от эффекта положения (West and Fraser, 2005; Gaszner and Felsenfeld, 2006; Valenzuela and Kamakaka, 2006). Основываясь на способности большинства инсуляторов блокировать гетерологичные энхансеры, нами предложена стратегия для выявления инсулятор-подобных элементов в протяженных участках генома. Мы разработали функциональный подход, который позволяет отбирать инсуляторы из смеси геномных фрагментов. В нём применяется процедура позитивно-негативной селекции с использованием гена тимидинкиназы вируса простого герпеса (рис. 12). Анализируемый фрагмент, проявляющий инсуляторную активность, в конструкции стабильно интегрированной в геном, блокирует взаимодействие энхансера с промотором, что приводит к подавлению экспрессии тимидинкиназы вируса простого герпеса (HSV-tk). Тимидинкиназа сначала фосфорилирует ганцикловир, который является синтетическим аналогом 2-дезоксигуанозина, затем фосфорилированный ганцикловир фосфорилируется клеточными киназами до трифосфата и встраивается в растущую при репликации цепь ДНК, блокируя ее, что приводит к гибели клеток. При селекции клеток ганцикловиром выживут клетки, в которых подавлен синтез *HSV-tk* вследствие блокирования энхансера инсулятором.

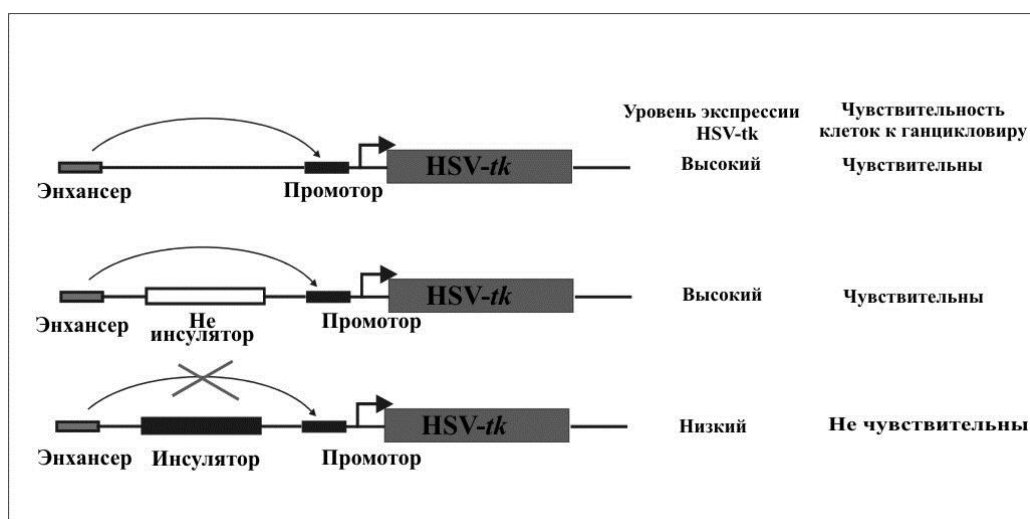


Рис. 12. Механизм селекции инсулятор-подобных последовательностей.

Стратегия идентификации инсулятор-подобных последовательностей в протяженных фрагментах геномов (рис. 13) предполагает:

1. Конструирование вектора с двумя селективными маркерами
2. Клонирование в вектор смеси анализируемых фрагментов ДНК
3. Трансфекцию клеток смесью векторов
4. Отбор рекомбинантных клеток, содержащих векторы при помощи первого селективного агента - неомифосфотрансферазы *NeoR* (позитивная селекция)
5. Отбор клеток, содержащих инсулятор-подобные элементы при помощи второго селективного агента (*HSV-tk*)
6. Структурную и функциональную характеристику идентифицированных инсулятор-подобных элементов.

#### 4.2.1. Система для экспериментального поиска инсуляторов

Общая схема метода представлена на рис. 13А. Для селекции потенциальных инсуляторов была приготовлена конструкция, названная рPNT/EmP, схема которой представлена на рис. 13Б. За основу была взята плаزمида рPNT [Tubulewicz, 1991], содержащая ген устойчивости к неомифину ( $Neo^R$ ) и ген тимидинкиназы вируса простого герпеса, придающий

экспрессирующим его клеткам чувствительность к ганцикловиру (*Herpes Simplex Virus, HSV-tk, GANC<sup>S</sup>*), позволяющие проводить позитивно-негативную селекцию. В полученной конструкции происходит эффективная экспрессия гена *HSV-tk*, однако, если между промотором и энхансером этого гена клонировать последовательность ДНК, способную блокировать действие энхансера (инсулятор), экспрессия гена *HSV-tk* в стабильно трансфицированных такой плазмидой клетках прекращается или значительно снижается, и содержащие эту последовательность клетки становятся устойчивыми к ганцикловиру.

В качестве положительного контроля была сконструирована плазида рPNT/mP, содержащая в регуляторной области гена *HSV-tk* только минимальный промотор цитомегаловируса без энхансера, и придающая клеткам фенотип *GANC<sup>R</sup>*, т.е. клетки, в геном которых интегрирует такая конструкция, устойчивы к ганцикловиру. В качестве отрицательного контроля использовали плазмиду рPNT/CMV-EmP придающую содержащим ее клеткам чувствительность к *GANC* (фенотип *GANC<sup>S</sup>*).

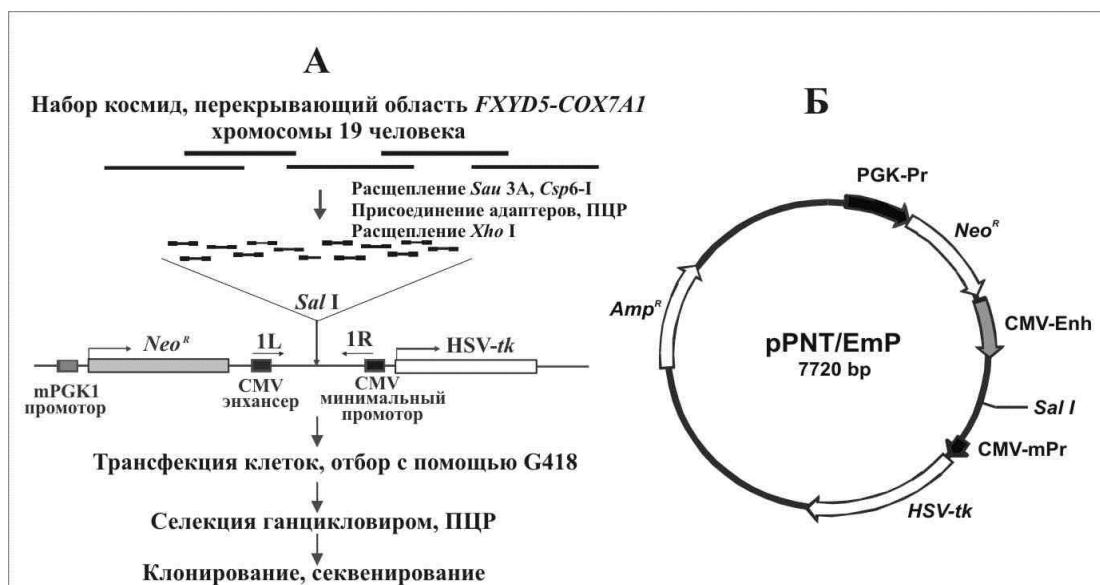


Рис. 13. (А) Общая схема отбора инсуляторов. mPGK1 - промотор фосфоглицераткиназы человека; *Neo<sup>R</sup>* - ген неомифосфотрансферазы; *HSV-tk* - ген тимидинкиназы вируса простого герпеса. (Б) Схема плазмиды рPNT/EmP для позитивно-негативной селекции инсуляторов.

Поиск потенциальных инсуляторов проводили в локусе хромосомы 19 человека длиной  $10^6$  пар оснований, расположенного между генами *FXVD5* и

*COX7A1*. Для получения библиотеки фрагментов, предназначенных для отбора потенциальных инсуляторов, ДНК 30 космид, перекрывающих область *FXYD5–COX7A1*, расщепляли по-отдельности рестриктазами *Sau3A* и *Csp6I*. К полученным фрагментам (длиной 200–700 п.о.) при помощи синтезированных адапторов был лигирован библиотечный праймер LP, обеспечивающий ПЦР-амплификацию фрагментов библиотеки и содержащий сайт расщепления *XhoI*. Затем обе библиотеки смешивали. Полученную библиотеку амплифицировали при помощи ПЦР с использованием праймера LP и обрабатывали рестриктазой *XhoI*. Продукты, полученные после рестриктной обработки, клонировали в плазмиду pPNT/EmP по сайту *Sall*. Для репрезентативной библиотеки было получено 12000 колоний, что соответствует 4-кратному перекрыванию области *FXYD5–COX7A1*. Выросшие клоны смывали с чашек Петри и использовали для выделения плазмидной ДНК.

Клетки линии СНО трансфицировали электропорацией полученными на предыдущем этапе плазмидами, которые предварительно линеаризовали при помощи эндонуклеазы рестрикции *SspI*. Условия электропорации (20 мс, 360 В, 800 мкФ) для клеток культуры СНО были отработаны автором в лаборатории проф. Боде в Национальном Центре Биотехнологии (Германия), и обеспечивают интеграцию одной копии линеаризованной плазмиды в геном. В тех же условиях проводили трансфекции клеток контрольными плазмидами pPNT/mP и pPNT/EmP. На этапе позитивной селекции клетки, в геном которых интегрировал вектор, отбирали по устойчивости к генетицину (G418). После этого проводили негативную селекцию генетицин-устойчивых клеток ганцикловиром. 100%-ная гибель клеток, содержавших контрольную плазмиду pPNT/EmP, определяющую чувствительность клеток к ганцикловиру, служила подтверждением успешно проведенной селекции. При этом популяция клеток, трансфицированных контрольной плазмидой pPNT/mP, была устойчива к ганцикловиру.

Чувствительность клеток к ганцикловиру может наблюдаться в ситуации, если клонированный фрагмент ДНК обладает свойствами сайленсера. Впрочем, последовательность со свойствами сайленсера не обладает направленностью, и

будет подавлять транскрипцию гена устойчивости к неомицину, что приведет к гибели таких клеток на этапе позитивной селекции. Для экспериментального подтверждения этого предположения мы клонировали три из идентифицированных последовательностей (№ 2, 5 и 8,) в плазмиду pGL3-promoter vector (Promega) с 5'-конца от промотора SV40 и регистрировали люциферазную активность этих конструкций в транзиентной трансфекции клеток HeLa. Сайленсерной активности не наблюдалось во всех трех конструкциях (не показано).

После стадии негативной селекции выделяли геномную ДНК из клеток, устойчивых к ганцикловиру. Эту ДНК использовали в качестве матрицы для ПЦР-амплификации участка между энхансером и промотором CMV при помощи праймеров 2L и 2R. Продукты амплификации повторно клонировали в вектор pPNT/EmP для проведения второго раунда селекции.

На рис. 14А представлены результаты ПЦР с праймерами 2L (GGATTTCCAAGTCTCCAGGGGAT) и 2R (ACCTCCCACCGTACACGCCT) пула плазмид перед трансфекцией (дор. 1) и результаты ПЦР с теми же праймерами на матрице геномной ДНК из клеток, трансфицированных библиотекой в плазмиде pPNT/EmP и прошедших негативную и позитивную селекции (дор. 2). Видно, что в результате селекции исходная библиотека, содержащая набор большого числа фрагментов разной длины, превращается в «лестницу» из ограниченного числа фрагментов.

Ранее мы идентифицировали и проанализировали в локусе хромосомы 19 небольшую часть инсулятор-подобных фрагментов [Акоров, 2006]. В данной работе мы попытались клонировать все или подавляющую часть потенциальных инсуляторов исследуемой области. Для этого отобранные фрагменты ДНК (дорожка 2, рис. 14 А) клонировали в вектор pGEM-T («Promega», США), и полученные клоны проверяли на наличие и длину вставок ПЦР с теми же праймерами (рис. 14 Б). Вставку наблюдали во всех 12 проанализированных клонах, а длины вставок различались для разных клонов. 20 клонов были секвенированы, четыре идентифицированные последовательности содержали



фрагменты ДНК *E. coli* и человеческую ДНК других хромосом, то есть не принадлежали к исследуемой области, анализ остальных клонов выявил 10 новых последовательностей. Эти последовательности нанесли на карту исследуемой хромосомы 19 между генами *FXVD5* и *COX7A1* (рис. 9, инсуляторы 9-18). Их свойства представлены в таблице 5.

На основании этих и полученных данных примерное количество инсуляторов в пересчете на геном человека составило ~20000, что соответствует экспериментально определенной средней длине хроматинового домена (80-300 т.п.о.), что соответствует примерно 10000-40000 пограничных элементов в геноме млекопитающих [Heng, 2001].

#### 4.2.2..Анализ расположения инсуляторов в геноме

Считается, что инсуляторы подразделяют геном на домены, содержащие экспрессирующиеся при сходных условиях гены, и вместе с генами, промоторами и энхансерами образуют многоуровневую сеть регуляции транскрипционной активности.

**Таблица 5.** Позиция потенциальных инсуляторов в локусе *FXVD5-COX7A1* хромосомы 19 человека.

№	Координаты на хромосоме 19 человека <sup>1</sup>	Длина (п.о.)	Позиция относительно генов
9	40347492-40347685	194	6-й интрон гена <i>FXVD5</i>
10	40406076-40406590	515	~1 т.п.о. с 3'-стороны от гена <i>TMEM162</i>
11	40437627-40437699	73	2-й интрон гена <i>LSR</i>
12	40909389-40909561	173	14-й интрон гена <i>MLL4</i>
13	40932935-40933049	114	1-й интрон гена <i>LIN37</i>
14	41067647-41067745	99	~4 т.п.о. с 3'-стороны от гена <i>NFKBID</i>
15	41116085-41116309	225	~4 т.п.о. с 5'-стороны от гена <i>LRFN3</i>
16	41147417-41147594	178	Межгенное <sup>2</sup>
17	41350295-41350437	143	Межгенное <sup>2</sup>
18	40801090-40801286	197	10-й и 11-й интроны, 11-й экзон гена <i>KIAA0841</i>

<sup>1</sup>По данным Human Genome Browser (Kuhn et al., 2007), сборка от марта 2006 г.

<sup>2</sup>Расстояние более 10 т.п.о. от ближайших генов.

Как отмечалось выше, термином «инсулятор» обозначают как энхансер-блокирующие элементы, так и пограничные элементы генома, отделяющие участки хроматина с различной структурой или активностью. Поскольку мы

идентифицировали инсуляторы только по энхансерблокирующей активности, то, в принципе, они могут не проявлять «барьерной» активности, хотя она и располагается, скорее всего, вблизи этих элементов. Тем не менее, в отдельных случаях можно выделить гипотетические домены, заключенные между инсуляторами.

В локусе *FXVD5–COX7A1* содержится кластер генов, кодирующих рецепторы, ассоциированные с G-белком, – *FFAR1*, *FFAR2*, *FFAR3* и *GPR42P*. Ранее было показано, что гены *FFAR3* и *FFAR2* имеют различные профили экспрессии (Brown et al., 2003). Позиции инсулятора 6, находящегося между генами *FFAR3* и *FFAR2*, и инсулятора 2, отделяющего *FFAR2* от гена *KRTDAP*, экспрессирующегося преимущественно в эпидермальных кератиноцитах (Bazzi et al., 2007), соотносятся с независимой экспрессией этих генов. Гены *ATP4A* и *KIAA0841* располагаются каждый в своем гипотетическом домене (между инсуляторами 5 и 1, 1 и 18 соответственно). Ген *ATP4A* транскрибируется преимущественно в тканях желудка (Chalaya et al., 2006; Herrmann et al., 2007) и явно отличается в этом отношении от окружающих генов.

Гены *HCST* и *TYROBP* расположены между инсуляторами 14 и 15 и обладают высокой тканеспецифичностью экспрессии, отличной от специфичности окружающих их генов. Так, эти гены обладают высоким уровнем экспрессии в миелоидных клетках (Wu et al., 1999; Takaki et al., 2006), тогда как отделенный от них инсулятором 14 ген *APLP1* экспрессируется преимущественно в коре головного мозга (Kim et al., 1995), а отделенный инсулятором 15 ген *LRFN3* – в нейрональных клетках (Morimura et al., 2006). Эти данные соответствуют представлениям о том, что последовательности, обладающие инсуляторной активностью, разделяют цепь ДНК на независимые домены, в каждом из которых находятся гены, характеризующиеся специфичным для них профилем экспрессии.

Из 10 идентифицированных последовательностей восемь содержат фрагменты повторяющихся элементов генома (табл.6). Инсулятор 9 представляет собой фрагмент LINE (Long Interspersed Element), а инсуляторы 10 и 17 – LTR

эндогенных ретровирусов человека MaLR и ERVL, соответственно. Другие потенциальные инсуляторы перекрываются с Alu- повторами (инсуляторы 12, 15–17). Последовательность 14 представляет собой повторяющийся элемент из семейства транспозонов Tc2.

**Таблица 6.** Повторяющиеся элементы генома, входящие в состав потенциальных инсуляторов.

Повторяющиеся элементы генома		Порядковый № инсулятора	Перекрытие инсулятора с повторяющимися элементами, %
SINE	AluSx	12	31
	AluSx	15	100
	AluSg	16	100
	AluSg	17	29
LINE	L2	9	98
LTR	ERVL	17	59
	MaLR	10	35
DNA	Tc2	14	100

Имеются указания на то, что некоторые Alu-повторы могут обладать инсуляторными свойствами (Willoughby et al., 2000; Ruiz-Narvaez and Campos, 2008). Область *FXVD5–COX7A1*, однако, содержит более 700 Alu-элементов, и обнаружение только двух из них среди предполагаемых инсуляторов указывает на то, что лишь отдельные Alu-повторы обладают такой активностью.

Ретротранспозиция элементов, подобных Alu и LINE, таким образом, может быть одним из механизмов перестройки доменной структуры хроматина (Arnold et al., 2000). Можно сделать предположение, что одной из регуляторных функций ретроэлементов является их участие в образовании функциональных доменов. Подобное предположение высказывалось для ретроэлементов, содержащих S/MARs, картированных нами ранее (Chernov et al., 2004).

Следует отметить, что исследованная нами область может содержать инсуляторы, которые не были идентифицированы из-за ограниченных возможностей использованного подхода. Идентифицируемые при данном подходе инсуляторы должны быть активны в клетках СНО и по отношению к паре «промотор-энхансер» цитомегаловируса. При этом некоторые ткане- или

энхансер-специфичные инсуляторы могли быть пропущены. Подобные недостатки, однако, могут быть преодолены путем использования других клеточных линий и промоторно-энхансерных конструкций.

Таким образом, учитывая перечисленные слабые места, предложенный метод может использоваться как важный подход функциональной геномики.

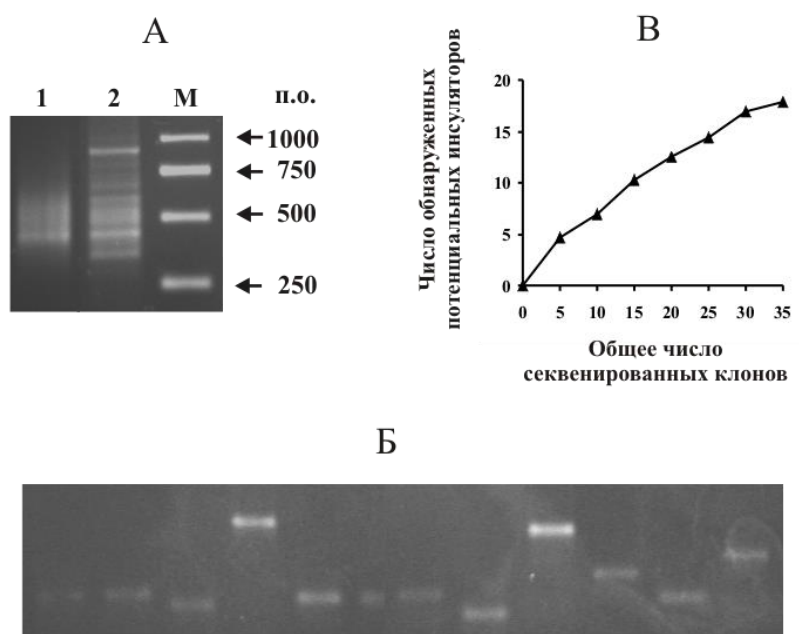


Рис. 14. Анализ библиотеки потенциальных инсуляторов.

А) Результаты селекции инсуляторов. ПЦР-амплификация с использованием праймеров 2L и 2R на матрицах: 1 – пул плазмид до трансфекции, 2 – геномная ДНК клеток, которые трансфицировали библиотекой коротких фрагментов, клонированной в векторе рNNT/EmP после позитивной (G-418) и негативной (GCV) селекций. М – маркер длин ДНК.

Б) ПЦР-амплификация с использованием праймеров 2L и 2R 12-ти случайно отобранных клонов из библиотеки потенциальных инсуляторов.

В) Оценка числа независимых клонов в библиотеке. Приведена зависимость числа клонов с уникальными последовательностями от общего числа секвенированных клонов.

#### 4.2.3. Функциональный анализ инсуляторов

После идентификации и картирования инсуляторов актуальным является наличие системы, позволяющей сравнивать их активности при помощи функционального теста. С этой целью мы разработали подход, основанный на трансфекции клеток линейризованной плазмидой (Recillas-Targa et al., 1999).

Мы отобрали четыре фрагмента ДНК, обладающих энхансер-блокирующей активностью, выявленные нами методом позитивно-негативной селекции (Акоров et al., 2006; Didych et al., 2009). Свойства и характеристика этих фрагментов приведены в таблице 7.

Таблица 7. Свойства потенциальных инсуляторов, выявленных методом негативно-позитивной селекции.

Фрагмент	Положение на хромосоме 19 человека (сборка hg19)	Length, bp	Состояние хроматина*, **	Связывание CTCF*	Связывание p300*
Ins1	36547955-36548158	204	Слабый энхансер (GP) Транскрибируется (K)	+ K	-
Ins2	36033501-36033722	222	Слабый энхансер (G) Транскрибируется (KP)	++ G	-
Ins3	36034844-36035321	478	Слабый энхансер /weak promoter (GKP)	+ K	+/- K
Ins4	35668221-35668887	667	Слабый энхансер/транскрибируется (K)	-	-

\*Буквы соответствуют линиям клеток, в которых регистрировали указанный тип активности: H - HeLa, G - GM12878, P - HepG2, K - K562, I - IMR90, A - A549. Степень связывания факторов транскрипции указана в соответствии с данными ChIp-seq ENCODE (Rosenbloom et al., 2009)– от (-) – отсутствие связывания до (+++) – сильное связывание.

\*\*Структура хроматина, установленная по девяти эпигенетическим маркерам, включающим связывание фактора транскрипции CTCF и модификации гистонов различного типа (Ernst and Kellis, 2010; Ernst et al., 2011).

Классический CTCF-связывающий инсулятор cHS4 из бета-глобинового локуса кур (Chung et al., 1993; Bell et al., 1999) проявляет энхансер-блокирующую активность в различных системах (Recillas-Targa et al., 1999). Для проверки его активности мы клонировали инсулятор cHS4 в обеих ориентациях между энхансером и промотором SV40, а также между промотором SV40 и энхансером E2. В контрольных конструкциях cHS4 клонировали вне пары промотор-энхансер. Конструкциями трансфицировали клетки HeLa и определяли в них активность люциферазы (рис. 15). Если инсулятор cHS4 был расположен между энхансером и промотором, он понижал активность пары промотор-энхансер в 2-3 раза (рис. 15а). Подавления активности не наблюдали при клонировании между энхансером и промотором контрольного фрагмента ДНК фага лямбда.

Аналогичный результат получили при использовании пары промотор SV40–

энхансер E2. В обоих случаях энхансер-блокирующая активность сHS4 слабо зависела от ориентации инсулятора.

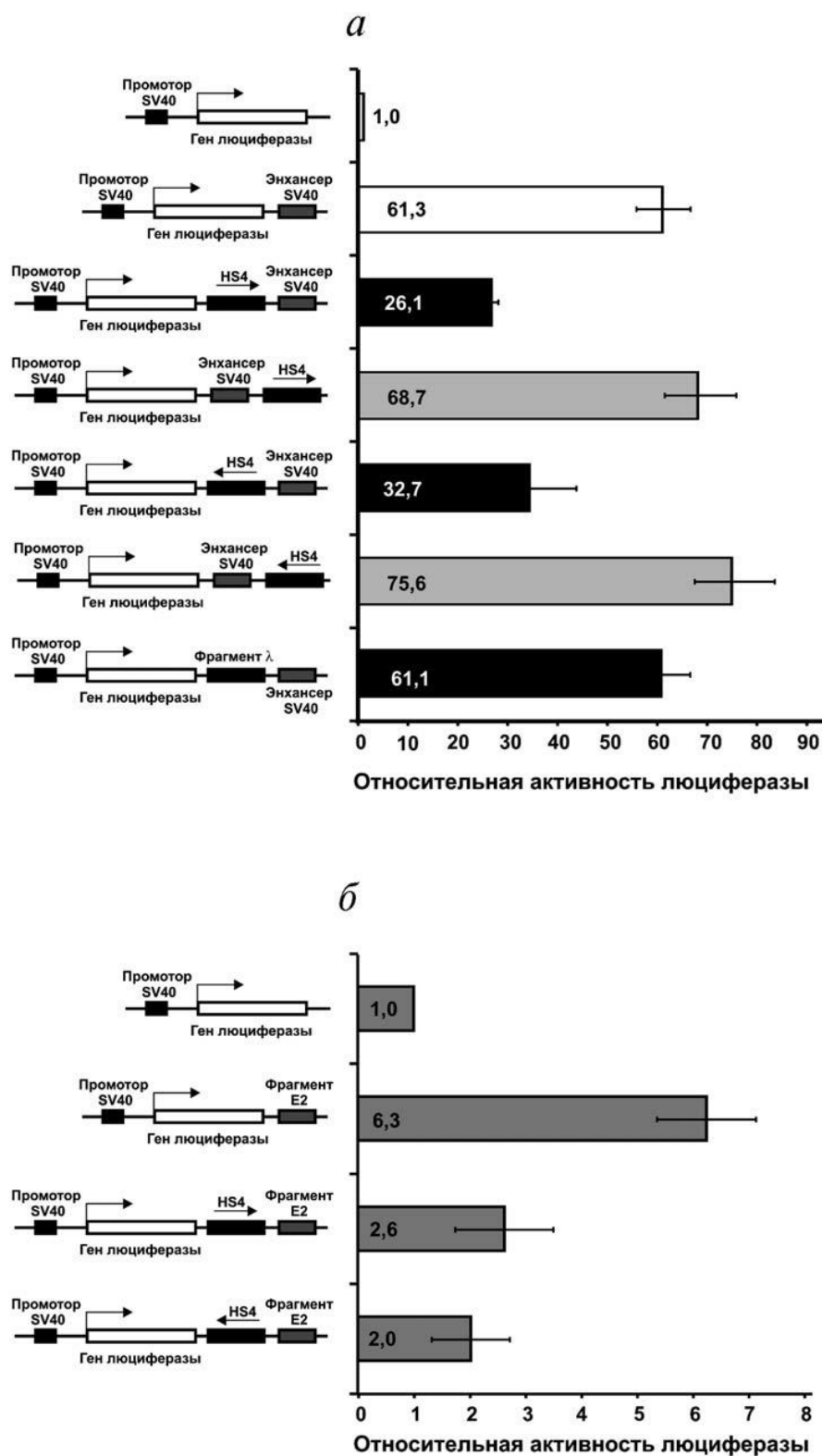


Рис. 15. Энхансер-блокирующая активность инсулятора сHS4 из бета-глобинового локуса кур в клетках HeLa по отношению к паре промотор SV40–энхансер SV40 (а) и паре промотор SV40–энхансер E2. За единицу принята

активность люциферазы после трансфекции клеток плазмидами, содержащими промотор SV40 и соответствующий энхансер.

Четыре фрагмента, отобранных при помощи позитивно негативной селекции (табл. 7) клонировали в плазмиду pGL4EPV2 между энхансером и промотором SV40. Полученные конструкции линейаризовали и трансфицировали ими клетки CHO, поскольку исходно отбор энхансер-блокирующих фрагментов проводили в клетках этой линии, и клетки HeLa для сравнения. Из результатов трансфекций следует (рис. 16), что три из четырех (75%) отобранных фрагментов заметно (в 2-3 раза) ослабляли активность пары промотор-энхансер SV40, при этом их активность не зависела от ориентации относительно промотора (не показано). Более того, энхансер-блокирующая активность двух элементов (Ins2 и Ins3) выявлялась только в клетках CHO и практически отсутствовала в клетках HeLa, а активность Ins4 была одинаковой в этих клетках. При этом, Ins4 связывал фактор транскрипции CTCF, а Ins2 и Ins3 связывали CTCF (табл. 7).

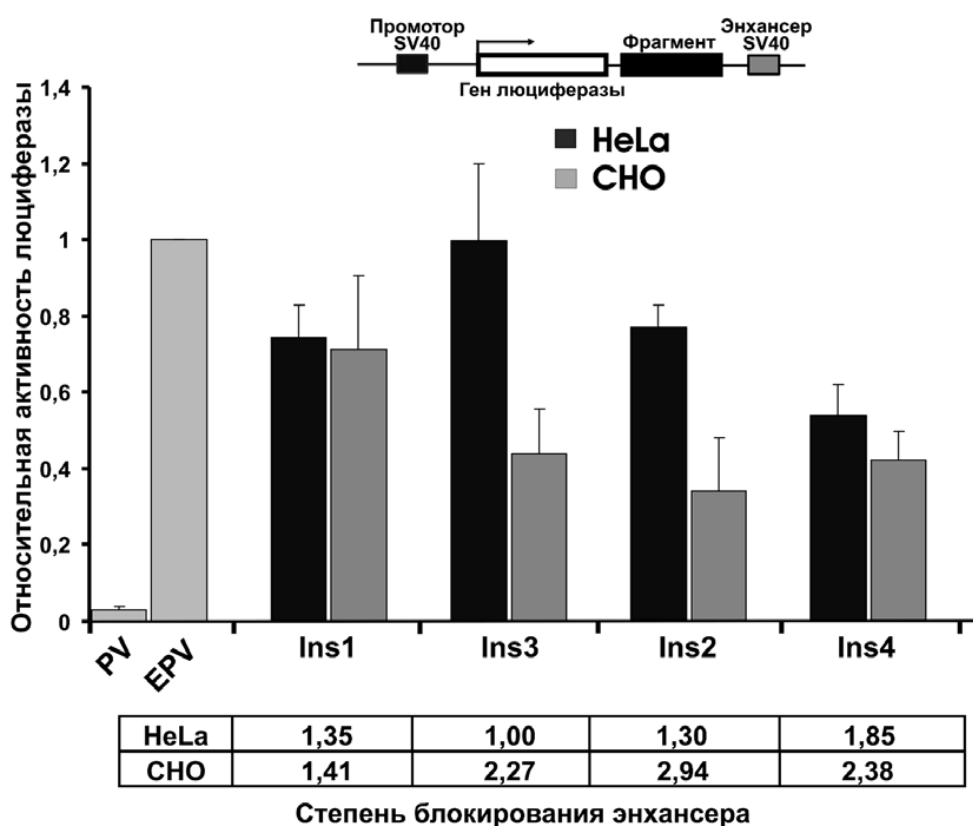


Рис. 16. Энхансер-блокирующая активность по отношению к паре промотор SV40–энхансер SV40 потенциальных инсуляторов, выявленных в геноме человека методом позитивно-негативной селекции. Приведены данные для

клеточных линий HeLa и CHO. За единицу принята активность люциферазы после трансфекции клеток плазмидами, содержащими промотор SV40 и соответствующий энхансер.

Таким образом, в системе с использованием транзientной трансфекции линейризованными конструкциями мы регистрировали выраженную энхансер-блокирующую активность как классического инсулятора cHS4 из бета-глобинового локуса кур, так и фрагментов из генома человека, выбранных на основании разработанного нами экспериментального подхода или идентифицированных другими авторами. Энхансер-блокирующая активность различных последовательностей характеризуется определенной тканеспецифичностью и зависимостью от их ориентации относительно промотора. Предложенная нами система во многих случаях пригодна для количественного анализа энхансер-блокирующей активности.

### **4.3. Энхансер-блокирующая активность CTCF-связывающих последовательностей**

Как уже отмечалось в обзоре литературы, CTCF представляет собой полифункциональный ядерный белок, присутствующий в большинстве тканей, который принимает участие в формировании пограничных элементов генома (инсуляторов). Число сайтов связывания CTCF в геномах млекопитающих составляет ~40000 и значительно (в 2-50 раз) превышает число сайтов связывания других факторов транскрипции. Поэтому вполне вероятно, что существенная часть сайтов связывания CTCF может приходиться не на инсуляторы, а определять другие функции CTCF.

Ранее при помощи разработанного нами метода двумерного сдвига электрофоретической подвижности мы выявили и картировали в локусе хромосомы 19 человека длиной 1 млн п.о., расположенном между генами *FXYD5* и *COX7A1*, десять фрагментов ДНК, содержащих участки связывания транскрипционного фактора CTCF (Vetchinova et al., 2006). Представлялось интересным определить, обладают ли эти фрагменты инсуляторной (энхансер-



блокирующей) активностью в описанной выше системе позитивно-негативной селекции.

Десять фрагментов, выявленных по связыванию с транскрипционным фактором CTCF (Vetchinova et al., 2006), и клонированных в плазмидный вектор pGEM-T, были вырезаны с использованием рестриктазы *Xho* I и клонированы в обеих ориентациях по сайту *Sal* I в плазмиду pPNT/EmP (Акоров et al., 2006) между энхансером и минимальным промотором цитомегаловируса.

В качестве одного из контролей использовали плазмиду pPNT/mP, содержащую в регуляторной области перед геном тимидинкиназы только минимальный промотор CMV и придающую клеткам фенотип устойчивости к ганцикловиру (Акоров et al., 2006).

В качестве контроля функциональной активности мы приготовили плазмиду pPNT/E-sns-mP, содержащую между энхансером и минимальным промотором CMV инсулятор *sns* морского ежа *Paracentrotus lividus* (Melfi et al., 2000; Di Simone et al., 2001). Для этого плазмиду pBS KS+, любезно предоставленную Р. Мелфи и Дж. Спинелли, содержащую инсулятор *sns*, последовательно гидролизовали рестриктазами *Hind*III и *Sma*I и с помощью фрагмента Кленова ДНК-полимеразы I достраивали 3'-концы фрагментов ДНК. После рестрикции продукты разделяли электрофорезом в 1,5 % агарозном геле, фрагмент ДНК длиной ~300 п.о., содержащий инсулятор *sns*, элюировали из геля и лигировали в линейаризованный по сайту *Sal* I вектор pPNT/EmP с достроенными 3'-концами.

В качестве негативного контроля мы приготовили плазмиду pPNT/E- $\lambda$ -mP, содержащую между энхансером и минимальным промотором CMV фрагмент ДНК фага  $\lambda$ . Для этого ПЦР-амплифицированный участок ДНК фага  $\lambda$ , содержащий на 5'-концах сайты расщепления *Xho* I, лигировали в плазмидный вектор pGEM-T Easy, и полученную плазмиду гидролизовали рестриктазой *Xho* I. После рестрикции фрагмент ДНК фага  $\lambda$  длиной ~200 п.о. элюировали из геля и лигировали в линейаризованный по сайту *Sal* I вектор pPNT/EmP.

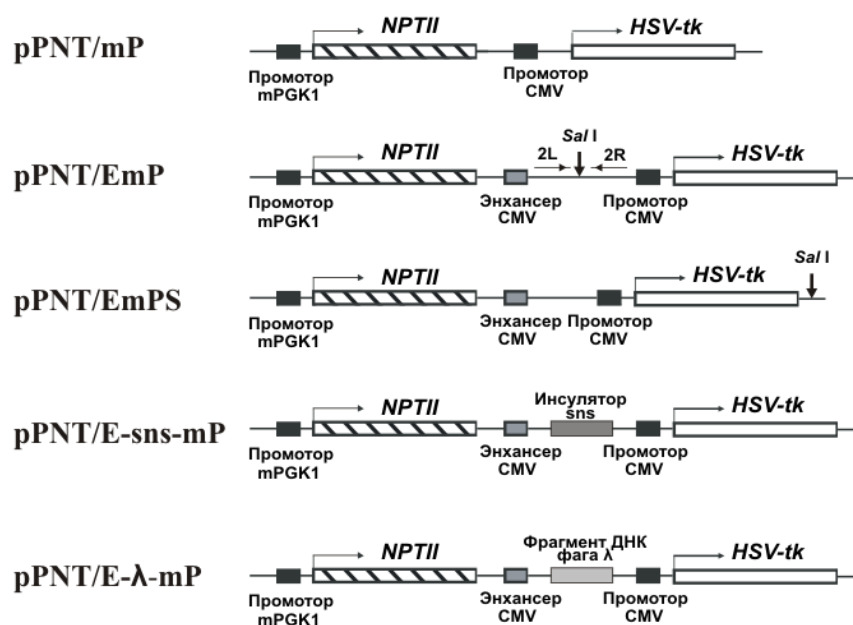


Рис. 17. Схемы плазмидных конструкций.

*mPGK-1* - промотор фосфоглицераткиназы-1 мыши;

*NPTII* - ген неомифосфотрансферазы *E. coli*;

*HSV-tk* - ген тимидинкиназы вируса простого герпеса.

Для проверки возможного наличия сайленсерной активности CTCF-связывающих фрагментов из плазмиды pPNT/EmP был удален сайт рестрикции *Sal I*, расположенный между энхансером и минимальным промотором CMV. С этой целью плазмиду pPNT/EmP линейаризовали по сайту *Sal I*, образовавшиеся 3'-концы достраивали с помощью фрагмента Кленова ДНК-полимеразы I, и лигировали вектор сам на себя. После этого плазмиду расщепляли рестриктазой *PspE I*, сайт узнавания которой расположен с 3'-конца от гена *HSV-tk*, и лигировали полилинкер, содержащий сайт рестрикции *Sal I*. Приготовленная плазида была обозначена как pPNT/EmPS (рис. 17). В нее по сайту *Sal I* были клонированы три CTCF-связывающих фрагмента (3, 7 и 8). Схемы приготовленных конструкций приведены на рис. 17.

Для анализа десяти CTCF-связывающих фрагментов на способность блокировать энхансер в системе позитивно-негативной селекции, их клонировали в плазмиду pPNT/EmP между энхансером и минимальным промотором цитомегаловируса. Таким образом приготовили 20 плазмид, с одним CTCF-связывающим фрагментом в каждой, клонированным в прямой и обратной

ориентациях. Анализ энхансер-блокирующей активности фрагментов проводили в обеих ориентациях, поскольку для отдельных CTCF-зависимых инсуляторов была показана зависимость их активности от ориентации (Hark et al., 2000).

Перед трансфекцией клеток плазмиды pPNT/EmP со вставками линейаризовали по сайту *Eco47III*, и все 20 плазмид объединили в общую смесь в равных количествах. В качестве внутренних контролей в смесь добавили в эквивалентных количествах линейаризованные плазмиды pPNT/E-sns-mP, содержащие инсулятор *sns* в обеих ориентациях, а также pPNT/E- $\lambda$ -mP, содержащие фрагмент ДНК фага  $\lambda$ . Полученным пулом плазмид трансфицировали клетки линии СНО методом электропорации в условиях (20 мс, 360 В, 800 мкФ), приводящих к интеграции единственной копии плазмиды в геном (Baer et al., 2000).

Ранее отмечалось, что 5'HS4 элемент из  $\beta$ -глобинового локуса кур проявлял инсуляторные свойства в клетках линии эритролейкоза человека K-562 (Chung et al., 1993), а инсулятор *sns* морского ежа *Paracentrotus lividus* блокировал действие энхансера на промотор в клетках остеосаркомы человека U2-OS и в клетках человеческой аденокарциномы легкого H1299 (Di Simone et al., 2001). Кроме того, показано, что человеческий CTCF связывается с соответствующими участками в геноме мыши (Filipova et al., 1996). Поскольку инсуляторы организмов одного вида могут сохранять свою активность в клетках организмов других видов, вполне корректно использование клеток линии СНО для поиска инсуляторов среди фрагментов ДНК из локуса хромосомы 19 человека. После позитивно-негативной селекции, которую проводили как описано ранее, из прошедших отбор клеток выделили геномную ДНК.

Образцы геномной ДНК анализировали посредством двухстадийной ПЦР. На первой стадии на геномной ДНК в качестве матрицы с праймерами 2L и 2R амплифицировали фрагменты, находящиеся между энхансером и промотором CMV (рис. 18). После ПЦР получили смесь клонированных ранее CTCF-связывающих фрагментов, фланкированных небольшими фрагментами плазмиды

pPNT/EmP. Эту смесь использовали в качестве матрицы для следующей стадии ПЦР.

К последовательностям каждого CTCF-связывающего фрагмента из локуса *FXYD5-COX7A1*, а также инсулятора морского ежа *sns* конструировали внутренние праймеры (табл. 8), которые в комбинации с плазмидными праймерами (2L или 2R) позволяли определить наличие каждой из вставок в смеси и их ориентацию (рис. 18).

**Таблица 8.** Сочетание праймеров для идентификации CTCF-связывающих последовательностей в прямой и обратной ориентации.

Фрагмент	Комбинации праймеров			
	Прямая ориентация		Обратная ориентация	
	Конструкция pPNT/EmP	Конструкция pPNT/EmPm	Конструкция pPNT/EmP	Конструкция pPNT/EmPm
<b>1</b>	2R, 1for		2R, 1rev	
<b>2</b>	2R, 2for		2R, 2for	
<b>3</b>	2R, 3for	SalF R, 1for	2L, 3for	
<b>4</b>	2R, 4for		2R, 4rev	
<b>5</b>	2R, 5for		2L, 5for	
<b>6</b>	2R, 6for		2R, 6rev	
<b>7</b>	2R, 7for	SalF R, 7for	2R, 7rev	
<b>8</b>	2R, 8for		2R, 8rev	SalF R, 8for
<b>9</b>	2L, 9for		2R, 9rev	
<b>10</b>	2R, 10for		2L, 10for	
<b>sns</b>	2R, snsL		2R, snsR	

Из рис. 18 (верх) видно, что все 10 CTCF-связывающих фрагментов, а также инсулятор *sns*, обнаруживаются в геномной ДНК клеток после селекции с генетицином G418, то есть содержащие их конструкции интегрировали в геном клеток. Те же фрагменты обнаруживаются и после селекции клеток в присутствии 10 мкМ ганцикловира в культуральной среде. Аналогичные результаты получили при селекции с концентрацией селективного агента 4 мкМ (не показано). Таким образом, можно заключить, что все 10 клонированные нами CTCF-связывающих фрагментов, как и инсулятор *sns*, будучи клонированными между промотором и энхансером, придают клеткам устойчивость к ганцикловиру.

Устойчивость к ганцикловиру трансфицированным клеткам, помимо инсуляторов, могут придавать сайленсеры, т.е. последовательности, которые, в отличие от инсуляторов, способны подавлять активность промотора независимо от их положения относительно последнего (Ogbourne and Antalis, 1998). В таком случае, сайленсер способен инактивировать и промотор гена неомицинфосфотрансферазы, и клетки, содержащие такую конструкцию должны погибать на стадии позитивной селекции (Акоров et al., 2006). Однако для тестирования наличия потенциальной сайленсерной активности у CTCF-связывающих фрагментов три из них (3, 7 и 8) мы клонировали в плазмиду pPNT/EmPS с 3'-стороны от гена *HSV-tk* (Рис. 17). Клетки, трансфицированные этими плазмидами, погибали на стадии негативной селекции с ганцикловиrom, т.е. CTCF-связывающие последовательности 3, 7 и 8 не оказывали ослабляющего действия на активность энхансера CMV и не являлись сайленсерами.

Таким образом, анализ в системе позитивно-негативной селекции фрагментов, картированных в локусе *FXYD5-COX7A1* и связывающихся с транскрипционным фактором CTCF показал, что все эти фрагменты обладали энхансер-блокирующими свойствами, причем эта их активность проявлялась в обеих ориентациях относительно промотора. Энхансер-блокирующая активность инсулятора морского ежа *sns* в системе позитивно-негативной селекции также не зависела от ориентации.

Несмотря на то, что все проанализированные нами CTCF-связывающие фрагменты обладали энхансер-блокирующей активностью, следует учесть, что CTCF (и, соответственно, образующие с ним комплекс фрагменты ДНК) реализуют в геноме, кроме инсуляторной, большое количество различных функций. В искусственной системе энхансер-блокирующая активность фрагмента ДНК может быть отражением его более сложной функции в геноме, реализуемой не только посредством взаимодействия с энхансерными белками. В использованной нами системе энхансер-блокирующая функция выявляется главным образом потому, что энхансер и связанные с ним белки помещаются вблизи CTCF-связывающего участка.

В целом, в результате применения системы позитивно-негативной селекции удалось идентифицировать 18 фрагментов ДНК с энхансер-блокирующей активностью (Акоров et al., 2006), принадлежащих к тому же участку хромосомы 19, что и 10 CTCF-связывающих фрагментов, для которых также была продемонстрирована энхансер-блокирующая активность (Vetchinova et al., 2006).

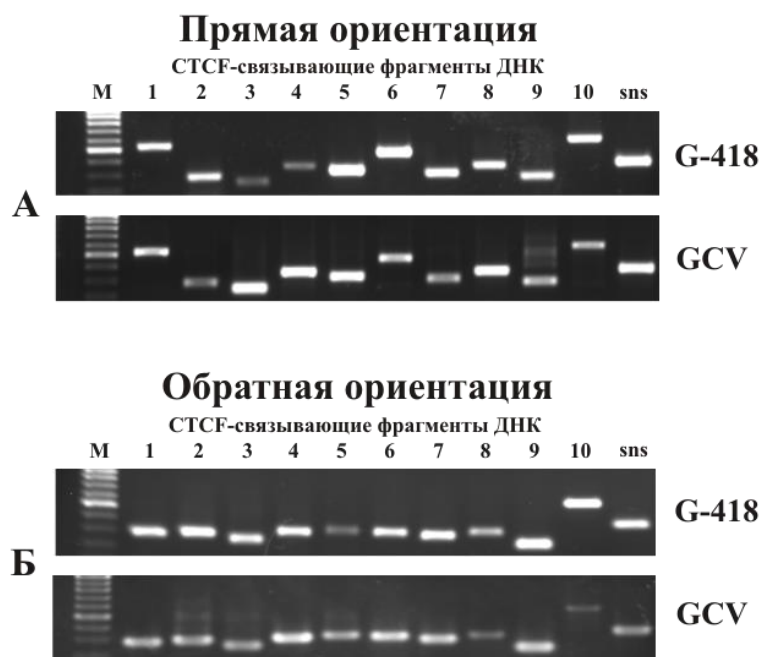


Рис. 18. ПЦР анализ с использованием в качестве матрицы геномной ДНК из трансфицированных клеток СНО после позитивной (G418) и негативной (GCV) селекции с праймерами, специфичными к 10 CTCF-связывающим последовательностям ДНК. (А) "Прямая" и (Б) "обратная" ориентации относительно промотора. М - маркер длин фрагментов ДНК.

Следует заметить, что в большинстве своем последовательности, выявленные по их энхансер-блокирующей активности и способности к связыванию с CTCF не пересекаются, хотя все эти фрагменты способны блокировать действие энхансера и связывать CTCF. Можно сделать предположение, что фрагменты, с высокой аффинностью взаимодействующие с CTCF (именно такие фрагменты мы отбирали в экспериментах (Vetchinova et al., 2006), обладают относительно слабой энхансер-блокирующей активностью, и наоборот, сильные инсуляторы относительно слабо связывают CTCF. Для проверки подобного предположения необходимы разработка и применение

методики, позволяющей количественно оценивать величину энхансер-блокирующей (инсуляторной) активности.

#### **4.4. Идентификация и картирование CTCF-связывающих последовательностей в глобиновом локусе птиц**

Домен, включающий тканеспецифично экспрессирующиеся гены альфа-глобина кур, исследуется на протяжении ряда лет и является удобной моделью для изучения регуляции активности генов. Внутри этого домена идентифицирован ряд регуляторных элементов, в частности энхансеры, участки связывания с ядерным матриксом, инсуляторы и локус-контролирующий район (Guerrero et al., 2007; Razin and Ioudinkova, 2007). Однако, задача выяснения точных механизмов регуляции активности генов внутри домена далека от полного разрешения. По-видимому, внутри домена имеется большое число еще неизвестных регуляторных элементов, которые оказывают существенное влияние на его функционирование.

Картирование сайтов связывания транскрипционного фактора CTCF в альфа-глобиновом домене кур и исследование их функциональных свойств представляется весьма актуальным и позволяет значительно расширить наши представления о регуляции транскрипции генов и функционировании протяженных участков генома в целом.

Куриный белок CTCF получали с использованием системы экспрессии в клетках Cos-1. Для этого в экспрессионную плазмиду рННс вводили последовательность, кодирующую 6 остатков гистидина. Полученную плазмиду назвали рSG5/HisCTCF. Плазмида рННс (Klenova et al., 1997) была приготовлена на основе вектора рSG5 (Stratagene) путем вставки кДНК куриного CTCF и любезно предоставлена С.В. Разиным.

Плазмидой рSG5/HisCTCF трансфицировали клетки линии Cos-1, культивировали в течение 48 час, снимали раствором Версена, обрабатывали ультразвуком, и белок CTCF очищали методом никель-аффинной хроматографии. Качество очищенного белка проверяли при помощи вестерн-блоттинга с

антителами к CTCF, а также при помощи EMSA, конкурентного EMSA и супершифта с известными фрагментами ДНК, способными связывать CTCF.

ДНК искусственной бактериальной хромосомы (BAC), содержащей исследуемую область генома кур (любезно предоставлена О.В. Яровой), расщепляли в двух параллельных образцах либо Csp6-I, либо Sau3A, к концам полученных фрагментов лигировали синтетические праймеры-линкеры для последующей ПЦР-амплификации. Образцы смешивали и амплифицировали. Полученную библиотеку коротких фрагментов из области генома кур, содержащей кластер альфа-глобиновых генов, радиоактивно метили при помощи ПЦР, смешивали с очищенным белком CTCF и подвергали предложенной нами ранее (Vetchinova et al., 2006) процедуре двумерного EMSA. При этом связывающие CTCF фрагменты образовывали на двумерном геле дополнительную диагональ. Область геля, предположительно содержащую CTCF-связывающие фрагменты, вырезали, фрагменты ДНК элюировали, проводили второй цикл отбора и клонировали. Клонированную библиотеку потенциальных CTCF-связывающих фрагментов ранжировали на 96-луночном планшете.

11 случайно отобранных клонов библиотеки методом EMSA проверили на их способность связывать куриный CTCF (рис. 19). Оказалось, что все они связывают белок CTCF сравнимо с положительным контролем (CTCF-связывающим фрагментом ДНК из лизоцимного сайленсера кур).



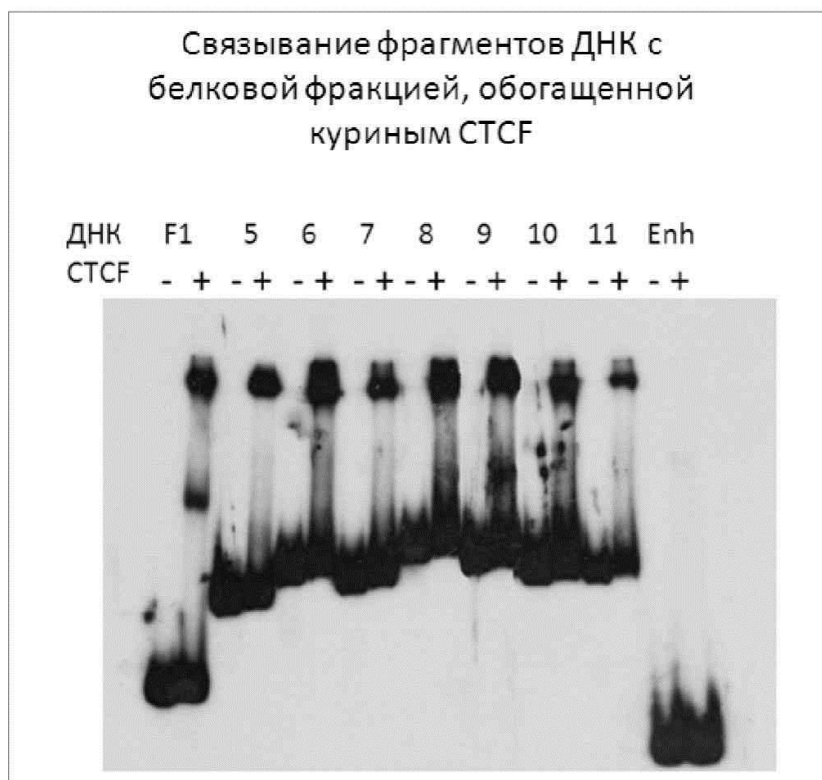


Рис. 19. Проверка эффективности связывания с белком CTCF фрагментов ДНК из библиотеки, полученной после селекции методом EMSA. F1 - CTCF-связывающий фрагмент ДНК из лизоцимного сайленсера кур (положительный контроль); Enh - фрагмент ДНК, содержащий энхансер бета-глобиновых генов кур (отрицательный контроль).

Далее было секвенировано 230 клонов библиотеки, среди которых найдено 77 уникальных последовательностей (рис. 20). Поскольку кривая рис. 20 практически достигает насыщения, полученная библиотека вполне представительна.

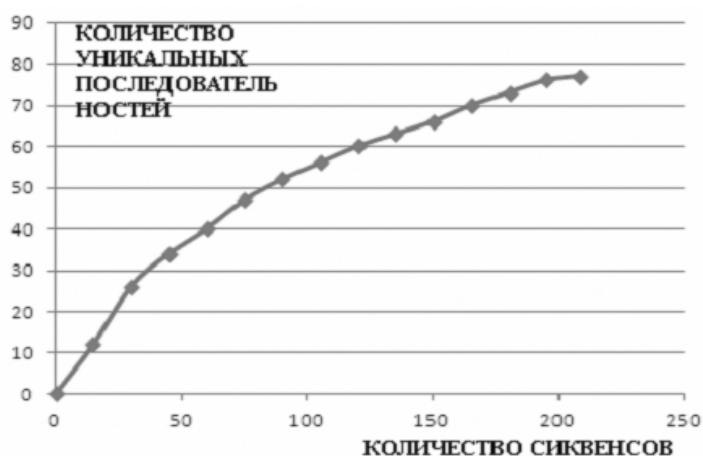


Рис. 20. Оценка числа независимых клонов в библиотеке. Показана зависимость числа уникальных последовательностей от полного количества секвенированных клонов.

Выявленные последовательности были нанесены на физическую карту локуса генома кур, содержащего альфа-глобиновые гены (рис. 21). На карту также были нанесены сайты связывания CTCF, идентифицированные различными методами в других работах (Valadez-Graham et al., 2004; Klochkov et al., 2006; Martin et al., 2011). Анализ распределения последовательностей показывает, что число сайтов генома, способных связывать CTCF *in vitro* по данным двумерного EMSA с применением очищенной ДНК значительно превышает число сайтов, связывающихся с CTCF в составе хроматина и идентифицируемых методом иммунопреципитации хроматина. Этот вывод следует из сравнения полученных нами данных с результатами работы (Martin et al., 2011), в которой для нашего участка генома было идентифицировано лишь 6 участков связывания CTCF в одном типе клеток, т.е. на порядок меньше, чем нами. Четыре из 6 участков перекрываются с идентифицированными в нашей работе. Причиной такого расхождения может быть недостаточно высокая чувствительность метода иммунопреципитации хроматина. При этом участки, относительно слабо связывающиеся с CTCF *in vivo*, не регистрируются, тогда, как связывание со свободной ДНК выявляет все потенциально возможные участки связывания CTCF. Кроме того, степень сродства CTCF к одному и тому же участку связывания в составе свободной ДНК и в хроматине могут не совпадать, в частности, из-за положительной или отрицательной кооперативности связывания в присутствии других белков хроматина.

Далее мы провели эксперименты по выявлению фрагментов, связывающих CTCF *in vivo*, методом иммунопреципитации хроматина в клетках эритробластов кур HD3, в которых наблюдается экспрессия генов альфа-глобина, а также в клетках HD3, стимулированных к эритроидной дифференцировке при помощи

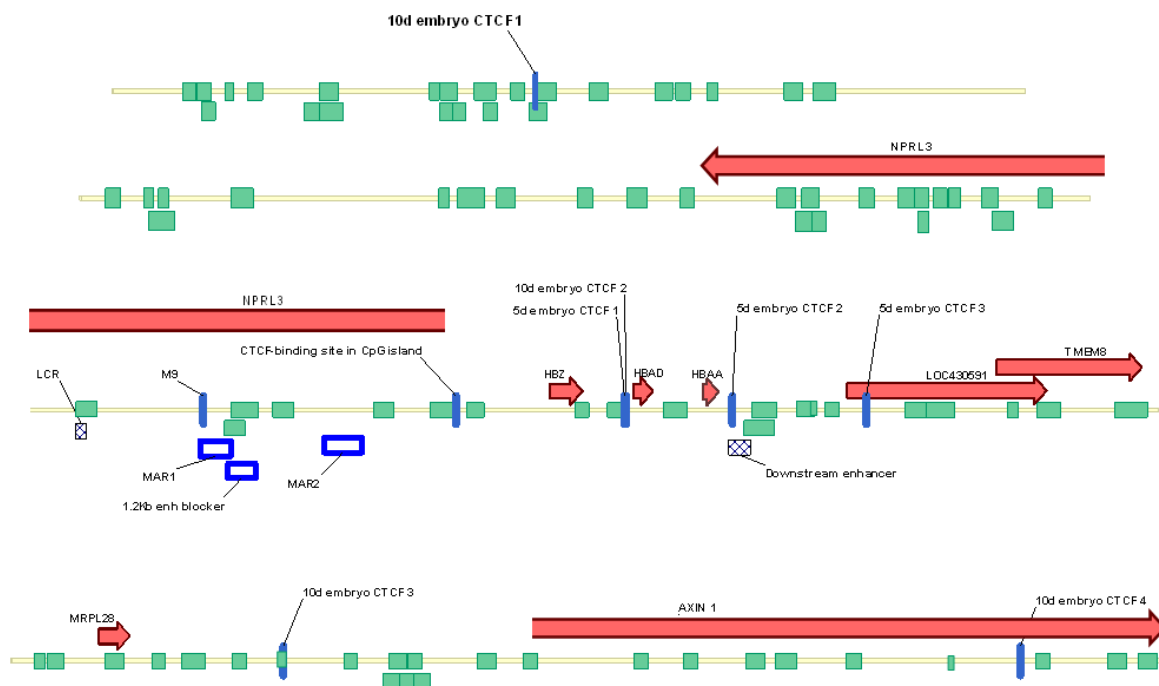


Рис. 21. Карта расположения потенциальных CTCF-связывающих последовательностей в области альфа-глобиновых генов кур. Красными стрелками обозначены расположение и направление транскрипции генов, зелеными прямоугольниками - расположение CTCF-связывающих последовательностей, идентифицированных методом двумерного EMSA. Синими вертикальными полосами обозначены участки, связывающие CTCF в клетках эмбрионов кур по данным иммунопреципитации хроматина (Valadez-Graham et al., 2004; Klochkov et al., 2006; Martin et al., 2011).

обработки геминном и бутиратом при повышенной температуре (Grdisa and White, 2000), а также в лимфоидной линии клеток кур DT40. В качестве отрицательных контролей были использованы не связывающие CTCF последовательности из области энхансера бета-глобинового локуса кур и гена альфа-глобина D, в качестве положительных контролей - связывающие CTCF элементы из промотора гена *MYC* и инсулятора F1 лизоцимного гена кур. Для проверки связывания CTCF были выбраны участки, расположенные в наиболее интересной для нас области альфа-глобинового локуса, непосредственно окружающей гены альфа-глобинов. Предположительно, именно эти участки ответственны за регуляцию транскрипции альфа-глобиновых генов.

На рис. 22 представлены сравнительные результаты экспериментов по иммунопреципитации хроматина клеток кур HD3, индуцированных к эритроидной дифференцировке клеток HD3 и клеток DT40. Как видно из рисунка,

идентифицированный нами фрагмент 12 связывает CTCF *in vivo* в клетках HD3 и индуцированных к дифференцировке клетках HD3 на уровне положительных контролей, зафиксировано также связывание CTCF фрагментами 4, 5 и 9 в индуцированных клетках HD3. Из идентифицированных ранее (Martin et al., 2011) CTCF-связывающих фрагментов лишь один (5d3) показал невысокий уровень связывания CTCF.

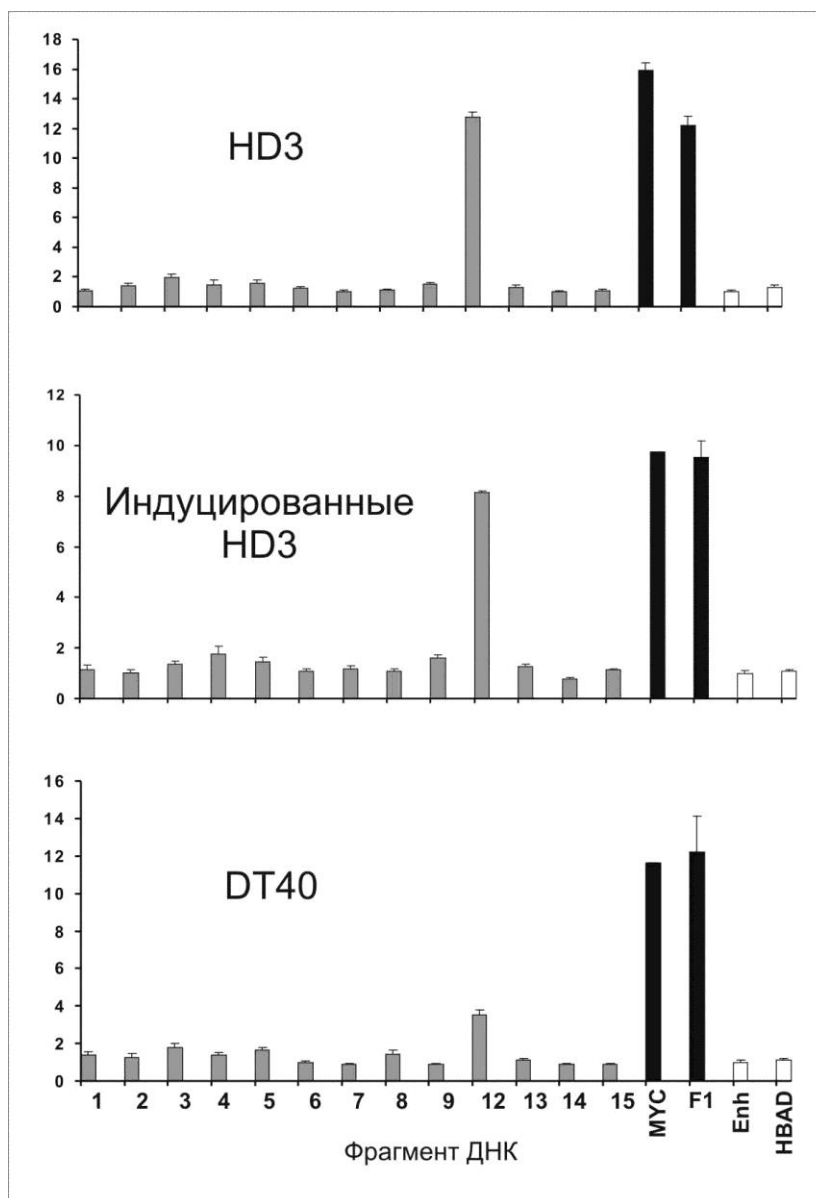


Рис. 22. Результаты экспериментов по иммунопреципитации хроматина с антителами к куриному CTCF для ряда фрагментов ДНК, идентифицированных нами в области генов альфа-глобинов кур (серые столбцы). Номера фрагментов соответствуют их положению на карте рис. 21. Черными столбцами обозначены положительные контроли (фрагменты промотора гена *MYC* и инсулятора *F1*

лизоцимного гена кур), белыми – отрицательные контроли (фрагменты области энхансера бета-глобинового локуса кур и гена альфа-глобина D).

При использовании для иммунопреципитации хроматина клеток DT40, в которых альфа-глобиновые гены не экспрессируются, было показано существенное (в 2-3 раза), снижение связывания CTCF фрагментом 12, которое, тем не менее, осталось достоверно выше отрицательных контролей, и фрагментами 4 и 9. Таким образом, в этих экспериментах выявлена тканеспецифичность связывания CTCF в хроматине, т.е. участки, связывающие CTCF в одном типе клеток, могут быть недоступны для связывания в другом типе, скорее всего, из-за модификации гистонов, метилирования ДНК и/или связывания с этими фрагментами других белков.

Так как во многих случаях фрагменты ДНК, способные связывать CTCF, обладают инсуляторной (энхансер-блокирующей) активностью, мы предприняли попытку определить эту активность у некоторых из отобранных нами фрагментов ДНК. Для этого фрагменты 4, 9 и 12 (см. рис. 21, 22) и контрольные фрагменты 27 и 350 (расположены на хромосоме 14 курицы вне альфа-глобинового локуса) были клонированы между промотором и энхансером SV40 и вне пары промотор-энхансер (рис. 23). Полученные конструкции линейаризовали, трансфицировали ими клетки HeLa и измеряли относительную активность репортерного гена люциферазы с использованием двойной люциферазной системы. Как видно из рис. 23, фрагменты 4 и 9 проявляют невысокую энхансерную активность, усиливая активность промотора вне зависимости от их расположения относительно последнего. Фрагменты 27 и 350 не проявляют заметной активности. Фрагмент 12, для которого характерна высокая степень связывания CTCF, способен усиливать активность промотора SV40 в присутствии энхансера SV40 примерно в два раза, причем только тогда, когда он располагается между энхансером и промотором. Такая “обратная инсуляторная” активность ранее не регистрировалась, и ее исследование требует дополнительных экспериментов.

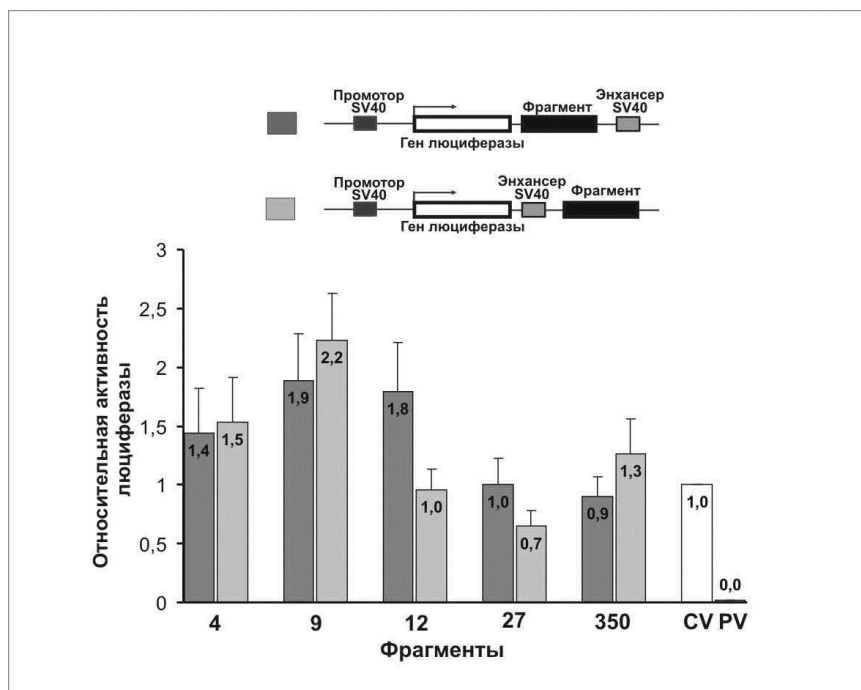


Рис. 23. Определение энхансерной/сайленсерной и энхансер-блокирующей активностей пяти CTCF-связывающих фрагментов ДНК в клетках HeLa. Темно-серые столбцы показывают уровень энхансер-блокирующей активности (фрагмент расположен между энхансером и промотором), светло-серые – уровень энхансерной/сайленсерной активностей (фрагмент расположен вне пары энхансер-промотор). Белый и черный столбцы – активность конструкций, содержащих промотор и энхансер SV40 (CV) и только промотор SV40 (PV). Активность конструкции с промотором и энхансером SV40 (CV) принята за единицу.

#### 4.5. Анализ функциональной архитектуры LTR семейства HERV-K(HML-2) и его взаимодействий с регуляторными компонентами клетки

##### 4.5.1. Характеристика LTR HERV-K

Одиночный LTR HERV-K(HML-2), находящийся на 19 хромосоме человека в области 19q13.31 длиной 38 т.п.о. (номер по GenBank L47334; далее LTR L47334, см. рис. 24), амплифицировали с космидного клона F23280, любезно предоставленного Е. Branscomb (Lawrence Livermore National Laboratory, USA) и использовали для определения тканеспецифичности его промоторной и энхансерной активностей. Эта область не содержит известных генов или их фрагментов на расстояниях до 20 т.п.о. от LTR. Анализ последовательности области L47334 выявил присутствие SINE и LINE повторов, LTR разных

семейств ERV, и несколько гипотетических экзонов (Lebedev et al., 2000). LTR L47334 отсутствует в соответствующем локусе генома шимпанзе и является представителем подгруппы HS-b “молодого” подсемейства семейства HERV-K(HML-2). Эта подгруппа предположительно образовалась около 10 млн. лет назад, а LTR L47334 – менее 5,6 млн. лет назад (Khil et al., 1997b; Lebedev et al., 2000; Buzdin et al., 2003). По другой классификации LTR L47334 может быть отнесен к слабо дивергировавшему девятому кластеру (Medstrand and Mager, 1998).

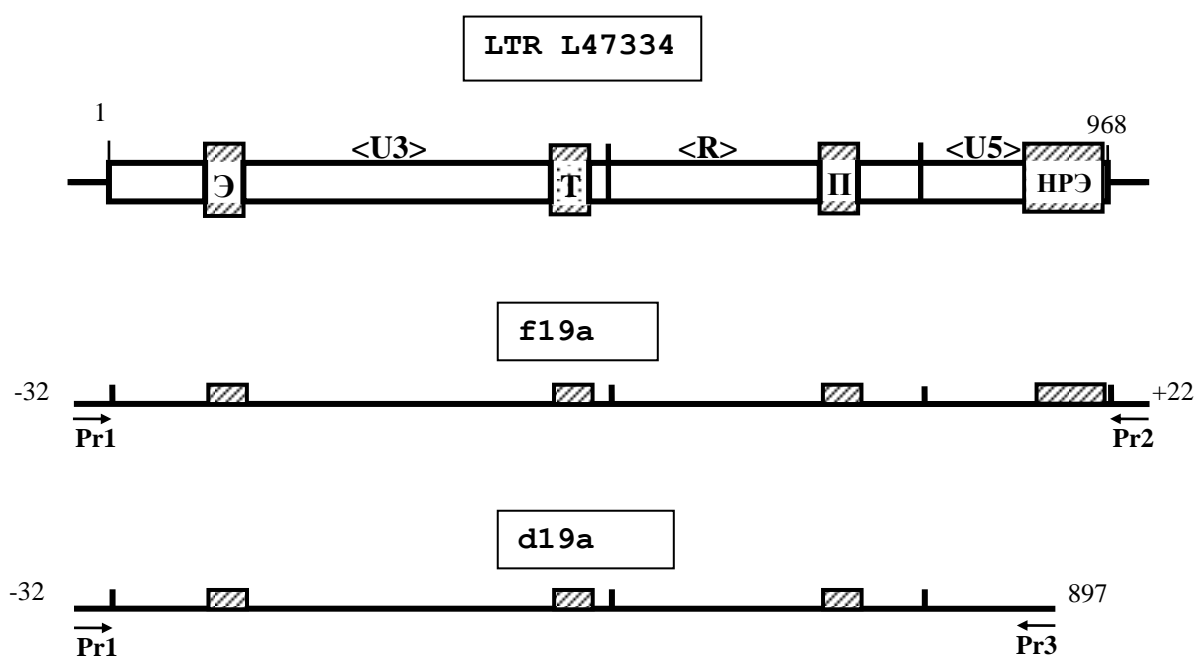


Рис. 24. Схематическое изображение LTR HERV-K и использованных в работе продуктов ПЦР, содержащих его фрагменты. Последовательности праймеров Pr1-Pr3 приведены в таблице 9. Регуляторные элементы изображены заштрихованными прямоугольниками. Э – энхансер, Т – ТАТА-бокс, П – сигнал полиаденилирования, НРЭ – негативный регуляторный элемент. Координаты продуктов ПЦР указаны относительно LTR ((Domansky et al., 2000), с изменениями).

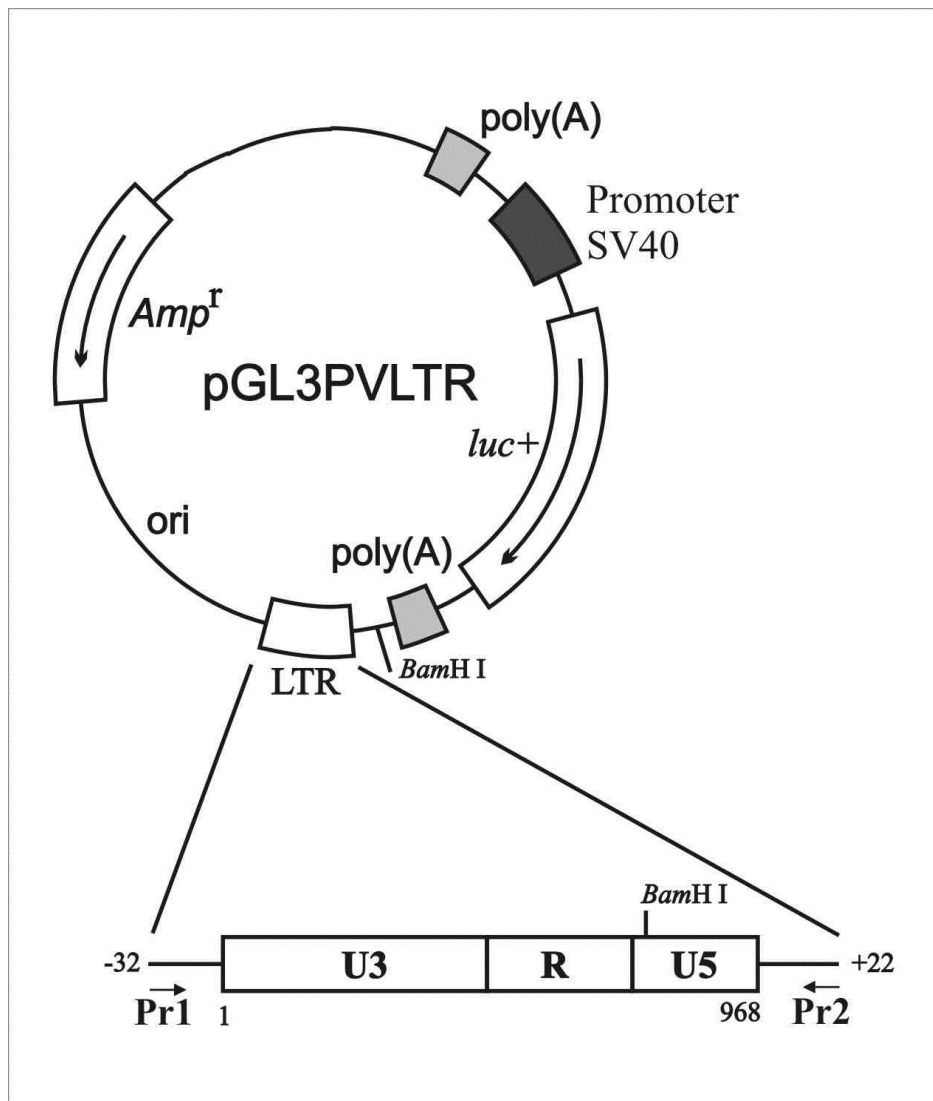


Рис. 25. Схема конструкции с репортерным геном люциферазы (*luc+*) и LTR в качестве энхансера. При проверке промоторной активности LTR находился на месте промотора SV40. Обозначения: *luc+* - ген люциферазы светлячка, *Amp<sup>r</sup>* - ген устойчивости к ампициллину, *ori* - участок начала репликации *E. coli*. Стрелки указывают направление транскрипции.

Полноразмерный LTR и фрагмент с частично делетированной U5-областью амплифицировали с использованием ПЦР. Координаты праймеров и ПЦР-фрагментов в последовательности LTR L47334 приведены на рис.24, а последовательности праймеров в таблице 9.

**Таблица 9.** Структура праймеров для ПЦР

Название	Описание и положение	Последовательность 5'-3', температура отжига (°C)
----------	----------------------	---



	праймера	
Pr1	5'-фланк LTR	CAGTCTTATCTCCTTTACTGACC (55°C)
Pr2	3'-фланк LTR	CCTCGTGTTTGTGCTTG (55°C)
Pr3	в LTR	AAAGACACAGAGACAAAGTATAGA (55°C)
GL2	Праймеры для плазмид	CTTTATGTTTTTGGCGTCTTCCA (60°C)
RV3	серии	CTAGCAAAATAGGCTGTCCC (60°C)
RV4	pGL3	GACGATAGTCATGCCCCGCG (60°C)
LB	Библиотечный прай- мер с сайтом <i>XhoI</i>	ACTGAGCTCGAGTATCCATGAACA (60°C)
FRToutL	pGTN8-плазмидные	GGATTTCCAAGTCTCCAGGGGAT (60°C)
FRToutR	праймеры	GGATTTCCAAGTCTCCAGGGGAT (60°C)

В экспериментах по транзientной трансфекции использовали конструкции, полученные на основе плазмид pGL3-Basic Vector (pGL3bv) и pGL3-Promoter Vector (pGL3pv) (Promega). Эти конструкции содержали полноразмерный LTR (pGL3/f19a) или LTR с частично делетированной U5-областью (pGL3/d19a) в прямой ориентации относительно репортерного гена. Для выявления промоторной активности LTR клонировали в pGL3bv перед геном люциферазы, а для выявления энхансерной активности - в pGL3pv с 3'-стороны от люциферазного гена (рис.25) (Domansky et al., 2000; Domanskii et al., 2002). Для проверки наличия и ориентации вставки проводили рестриктный анализ и ПЦР-амплификацию конструкций с применением праймеров, специфичных для последовательности LTR, и праймеров GL2, RV3 и RV4 (Promega, см. табл. 9), примыкающих к сайтам клонирования LTR. Полученными плазмидами трансформировали клетки *E. coli* DH5 $\alpha$ , выращивали их и выделяли плазмидные ДНК, содержащие репортерные конструкции.

Полученные значения активности люциферазы выравнивали с учетом эффективности трансфекции, которую определяли по активности  $\beta$ -галактозидазы, синтезируемой котрансфицируемой конструкцией pCMV SPORT- $\beta$ gal, содержащей ген  $\beta$ -галактозидазы под контролем промотора цитомегаловируса. Люциферазная активность экстракта клеток,

трансфицированных контрольным вектором pGL3pv, принималась за 100%. Подобная обработка результатов позволяла нам непосредственно сравнивать результаты независимых экспериментов, включая полученные на различных клеточных линиях (Wilson et al., 2003). Каждое значение активности определяли как среднее арифметическое для трех независимых трансфекций.

#### *4.5.2. Промоторная активность LTR*

Анализ промоторной активности LTR L47334 в различных клеточных линиях, т.е. ее тканеспецифичность, проводили в серии транзientных трансфекций конструкцией pGL3bv/f19a десяти клеточных линий различного происхождения. При выборе клеточных линий мы руководствовались литературными данными об особенностях экспрессии последовательностей ретровирусов в различных тканях и клеточных линиях (табл.10,11).

Показано, что клеточные линии, происходящие из эмбриональных тератокарцином и других герминогенных опухолей, содержат транскрипты эндогенных ретровирусов различных семейств, в том числе и HERV-K(HML-2) (Lower et al., 1993; Knossel et al., 1999). В этих клетках также происходит образование вирусных частиц (HTDV), содержащих РНК, гомологичную HERV-K(HML-2) (Lower et al., 1996). Так, в клетках линии Tera-1 тератокарциномы человека показана транскрипция LTR L47334 (Vinogradova et al., 2001), а с использованием клеток линии NT2/D1 тератокарциномы человека было проведено исследование его промоторной активности (Domansky et al., 2000). Имеются данные о зависимости уровня экспрессии генов HERV-K(HML-2) от стадии развития герминогенных опухолей (или опухолей, развившихся из клеток, находившихся на разных стадиях развития). Например, после начала дифференцировки экспрессия снижается (Herbst et al., 1999). Подобно этому, стимуляция дифференцировки клеток NT2/D1, стабильно трансфицированных конструкцией pGL3BV/f19a, приводила к снижению экспрессии люциферазы (Domansky et al., 2000). По этой причине результаты исследования промоторной активности, полученные на клетках линии NT2/D1 тератокарциномы человека,

было интересно сравнить с промоторной активностью LTR в клетках другой линии тератокарциномы человека, Tera-1. Как видно из рис. 26, промоторная активность LTR L47334 в клетках линии Tera-1 сравнима с активностью промотора SV40, что на порядок выше его активности в клетках NT2/D1.

Принято считать, что высокая промоторная активность LTR HERV-K наблюдается преимущественно в герминогенных опухолях и производных линиях клеток. Однако транскрипт, гомологичный последовательности гена *pol* HERV-K10, был выявлен также и в лимфоцитах здоровых доноров клеточной линии Lymph9 (Medstrand and Blomberg, 1993; Schon et al., 2001). Была показана транскрипция LTR L47334 в линии клеток Jurkat (острая Т-лимфобластная лейкемия) (Vinogradova et al., 2001) и связывание ядерного экстракта из этих клеток с U3-областью одного из LTR HERV-K(HML-2) (Акоров et al., 1998) (табл.10, 11). Однако в использованной нами системе LTR L47334 не проявил промоторной активности в клетках линии Jurkat (рис. 26) и в клетках промиелоцитарной лейкемии человека HL-60 (не показано).

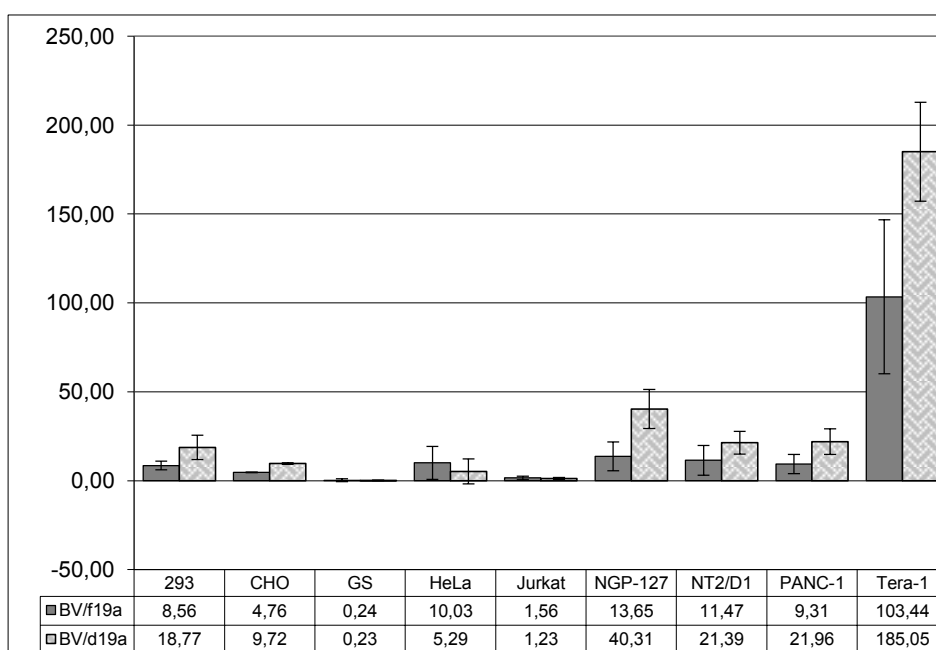


Рис. 26. Относительная промоторная активность HERV-K LTR L47334 (f19a) и его фрагмента (d19a) в различных клеточных линиях. Активность промотора SV40 принята за 100%.

**Таблица 10.** Активность LTR HERV-K(HML-2) в тканях и клеточных линиях.

Типы активности	Клетки/ткани	Ссылка
Специфичное взаимодействие с белками ядерных экстрактов различных клеточных линий	HeLa, Jurkat, CHO; GH, NT2/D1	(Akopov et al., 1998; Knossl et al., 1999)
Промоторная активность	NT2/D1; HeLa, MCF7	(Domansky et al., 2000; Schon et al., 2001)
Промоторная и энхансерная активность	GH, HeLa, NT2/D1	(Knossl et al., 1999)
Транскрипция	Плацента, паренхима яичка, герминогенные опухоли; линии клеток Jurkat, RMS-13, Tera-1	(Vinogradov et al., 2001)

**Таблица 11.** Активность LTR L47334 в различных тканях и клеточных линиях

Типы активности	Клетки	Ссылка
Промоторная активность	NT2/D1	(Domansky et al., 2000).
Транскрипция	Эмбриональная карцинома, смешанные герминогенные опухоли; линии клеток Jurkat и Tera-1	(Vinogradova et al., 2001)
Энхансерная активность	Линия клеток Tera-1	(Domanskii et al., 2002)

Слабую промоторную активность LTR мы зарегистрировали в линии клеток HeLa (рис. 26), что соотносится с данными, полученными ранее другими исследователями (Knossl et al., 1999; Schon et al., 2001).

В нормальных тканях мозга и почки, а также в клетках карциномы почки были обнаружены сплайсированные транскрипты гена *Env* HERV-K(HML-2) (de Parseval et al., 2003; Bannert and Kurth, 2004). Нам представлялось интересным выявить наличие промоторной активности LTR в линиях клеток соответствующего происхождения: NGP-127 (нейробластома человека), HEK-293

(клетки почки эмбриона человека, трансформированные ДНК аденовируса типа 5) и GS (карцинома почки человека). Было установлено, что LTR L47334 проявляет промоторную активность в клетках линий HEK-293 и NGP-127, в то время как в клетках GS активность не детектировалась.

Относительно высокая промоторная активность LTR HERV-K была выявлена нами в линии клеток PANC-1. Обнаружение активности LTR в линии клеток карциномы поджелудочной железы актуально также и потому, что существует предположение о связи экспрессии кодируемого геном *env* HERV-K суперантигена с инсулин-зависимым диабетом (Insulin-Dependent Diabetes Mellitus, IDDM). Аутоантиген этот был выявлен у пациентов с IDDM, что указывает на вероятную связь генетической предрасположенности к диабету с присутствием элементов HERV-K или с вариациями их экспрессии (Conrad et al., 1997; Medstrand and Mager, 1998). Впоследствии эта гипотеза не получила экспериментального подтверждения, однако, для окончательного решения данного вопроса необходимы дополнительные исследования (Nelson et al., 2003).

Мы исследовали промоторную активность LTR HERV-K в линии клеток различных видов млекопитающих. Прежде тканеспецифичная промоторная активность LTR HERV-K была описана для семенников трансгенных мышей и линии клеток эмбрионального рака мыши (Casau et al., 1999). Представлялось интересным выяснить, является ли промоторная активность LTR HERV-K типичной для грызунов, в геномах которых отсутствуют эндогенные ретровирусы этого типа. Нами была выбрана линия клеток яичника китайского хомячка CHO. Исследуемый LTR проявлял в ней промоторную активность, сравнимую с активностью, регистрируемой в некоторых человеческих линиях (рис. 26).

Таким образом, полученные нами результаты свидетельствуют о наличии промоторной активности LTR L47334 в нескольких исследованных линиях клеток: HEK-293, CHO, NGP-127, PANC-1 и Tera-1. Не выявлена промоторная активность в линиях GS, HL-60 и Jurkat. Активность LTR HERV-K в качестве промотора в значительной степени зависит от типа клеток. Например, промоторная активность LTR в клетках линии Tera-1 в десять раз превышает его

активность в других клеточных линиях человеческого происхождения. Следует оговориться, что эти значения справедливы с тем допущением, что активность промотора SV40 мало зависит от типа клеток, то есть не является тканеспецифичной.

Наличие промоторной активности LTR в клетках линии СНО позволяет высказать предположение, что для его функционирования достаточно тех факторов транскрипции, которые имеются в клетках грызуна, и, что эти факторы не являются специфичными для человека. Следует отметить, что линия клеток СНО происходит из ткани, в которой у трансгенных мышей HERV-K(HML-2) неактивны (Casau et al., 1999), однако одиночный LTR проявляет промоторную активность. Это может быть связано с изменениями, происходящими в клетках при трансформации, поскольку, хотя данных об активности HERV-K в нормальных яичниках человека нет, гены HERV-K экспрессируются в опухолях яичника эмбрионов (Herbst et al., 1998).

По сравнению с промотором SV40 LTR L47334 является относительно слабым промотором во всех протестированных линиях клеток, за исключением Tera-1, где активности этих промоторов сравнимы (рис. 26). Эти результаты несколько отличаются от полученных нами ранее, что может быть связано с влиянием на LTR его геномного окружения. Так, ранее не было обнаружено транскриптов U5-области LTR L47334 в Tera-1, свидетельствующих о его промоторной активности, а в линии клеток Jurkat, где, напротив, ранее наблюдали транскрипцию (правда, только после теплового шока) (Vinogradova et al., 2001), нам не удалось выявить такой активности.

Для нескольких LTR семейства HERV-H была показана тканеспецифичность промоторной активности, различная для разных представителей семейства (Schon et al., 2001). Подобные результаты были описаны для пары высокогомологичных LTR семейства HERV-E, один из которых проявлял промоторную активность в плаценте, а другой – в печени (Medstrand et al., 2001; Landry and Mager, 2003). Нами описаны различия в уровнях транскрипции отдельных представителей LTR семейства HERV-K(HML-

2) в нормальных и опухолевых тканях, а также в различных клеточных линиях человека (Vinogradova et al., 2001). Описана точечная замена в LTR MMTV, которая приводила к значительному снижению уровня транскрипции в молочной железе и не влияла на уровень транскрипции в лимфоидной ткани (Qin et al., 1999).

#### *4.5.3. Негативный регуляторный элемент (НРЭ)*

Ранее были описаны различия промоторной активности в разных тканях для частично делетированных и полноразмерных LTR MMTV (Bramblett et al., 1995; Zhu et al., 2000). Так, ММТ-вирус с модифицированными LTR заражал лимфоциты, а не клетки молочной железы. При анализе делеционных мутантов эндогенных ретровирусов свиньи были обнаружены участки LTR, тканеспецифично активирующие или репрессирующие транскрипцию (Wilson et al., 2003). Основываясь на данных о предполагаемом негативном регуляторе транскрипции, полученных ранее при делеционном мутагенезе U5-области LTR HERV-K (Domansky et al., 2000), мы решили сравнить промоторную активность полноразмерного LTR (f19a) с активностью его фрагмента (d19a) с удаленным НРЭ в нескольких линиях клеток.

Фрагмент LTR с удаленным НРЭ (d19a) проявлял промоторную активность в тех же линиях, в которых был активен полноразмерный LTR L47334 (рис. 26). В отличие от регуляторных элементов LTR других ретровирусов, активность которых зависела от типа ткани и вида организма (Zhu et al., 2000; Wilson et al., 2003), исследуемый нами НРЭ не обладал выраженной тканеспецифичностью. Нами было показано, что (1) фрагмент LTR (d19a), не содержащий НРЭ, проявлял промоторную активность во всех линиях, где был активен полноразмерный L47334, (2) во всех исследованных линиях клеток относительное увеличение промоторной активности по сравнению с полноразмерным LTR практически одинаково (рис. 26). Поскольку большая часть одиночных LTR HERV-K (HML-2) в геноме человека представлена 5' и 3' делетированными фрагментами, эти

результаты заслуживают особенного внимания (Lebedev et al., 1995; Vinogradova et al., 1997).

Показанная нами тканеспецифичность исследуемого НРЭ коррелирует с нашими результатами о широком распространении регуляторных белков, взаимодействующих с НРЭ. Возможно, регуляторная активность этого участка LTR обусловлена именно связыванием с ним белкового комплекса, как было показано для HTLV-I (Okumura et al., 1996). Таким образом, различия в промоторной активности LTR L47334, по-видимому, опосредованы не активностью НРЭ, а связаны с активностью энхансера в составе LTR.

#### *4.5.4. Энхансерная активность LTR*

Известно, что, энхансеры могут регулировать промоторы, находясь от них на значительном расстоянии. В частности, человеческий LTR HERV-E служит специфичным для плаценты энхансером гена рецептора эндотелина В (*EDNRB*), находясь на расстоянии более чем 20 т.п.о. от промотора (Medstrand et al., 2001). Ранее мы показали, что LTR L47334 проявляет энхансерную активность в клетках Tera-1, но не является энхансером в клетках линии NT2/D1 (Domanskii et al., 2002).

Одним из объяснений этого различия между этими близкими по происхождению линиями эмбриональной карциномы человека может быть различие в наборах экспрессируемых в них регуляторных факторов. Для подтверждения этого предположения были произведены эксперименты по сдвигу электрофоретической подвижности комплексов фрагментов ДНК с белками ядерных экстрактов, полученных из клеток этих линий. В качестве ДНК-зондов использовали радиоактивно меченые фрагменты LTR.

Белковые комплексы фрагмента LTR, содержащего энхансер, образующиеся при взаимодействии с белками ядерных экстрактов из клеток Tera-1 и других линий не различались, однако комплексы фрагмента LTR, содержащего промотор, с экстрактом из клеток Tera-1 образовывали дополнительную полосу торможения, свидетельствующую о наличии белков, отсутствующих в клетках линий NT2/D1 и Jurkat. Можно предположить, что



высокая энхансерная активность LTR L47334 в Tera-1 сопряжена с взаимодействием этих белков с LTR. Более того, высокая промоторная активность LTR в клетках линии Tera-1 может быть обусловлена влиянием энхансера LTR на его промотор (рис. 39).

Для доказательства этих предположений мы трансфицировали клетки Jurkat, CHO, HEK-293, HL-60, HeLa, GS, NGP-127 и PANC-1 конструкциями pGL3<sub>prv</sub>, содержащими ранний промотор SV40 и полноразмерный или лишенный HPЭ LTR в качестве энхансера. В качестве контрольных использовали конструкцию pGL3<sub>prv</sub>, в которой экспрессия гена люциферазы находится под контролем сильного промотора вируса SV40 и pGL3<sub>sv</sub>, дополнительно содержащую энхансер SV40.

На рис. 27 представлены результаты, подтверждающие наличие энхансерной активности у полноразмерного и частично делетированного LTR L47334, выраженной в существенном (до восьми раз) повышении уровня экспрессии репортерного гена в клеточной линии Tera-1. Ни в одной другой из использованных линий клеток полноразмерный или лишенный HPЭ LTR не проявляли энхансерной активности. Поскольку ранее мы показали, что наличие HPЭ не снижает энхансерную активность LTR в Tera-1 и при этом он влияет на экспрессию репортерного гена, не будучи включенным в состав транскрипта (Domanskii et al., 2002), можно сделать предположение, что он влияет на способность промотора собирать транскрипционный комплекс и/или на элонгацию транскриптов.

При исследовании энхансерной активности LTR HERV-K мы выявили также тканеспецифичность энхансера SV40. В литературе имеются противоречивые данные о тканеспецифичности энхансера SV40 (Nomiya et al., 1987; Wilson et al., 2003). Активность энхансера SV40 мы оценивали по соотношению люциферазной активности лизатов клеток, трансфицированных плазмидами pGL3<sub>sv</sub> (содержит энхансер и промотор SV40) и pGL3<sub>prv</sub> (содержит промотор SV40, но не содержит энхансера). Выраженная энхансерная активность наблюдалась в линиях клеток Tera-1 и NT2/D1, эпителиоидной карциномы шейки

матки человека HeLa, нейроэпителиомы NGP-127, а также в клетках яичника китайского хомячка CHO. В клетках линий промиелоцитарной лейкемии человека HL-60, острой Т-лимфобластной лейкемии человека Jurkat и клетках почки эмбриона человека HEK-293 активность энхансера SV40 была значительно ниже (рис. 28).

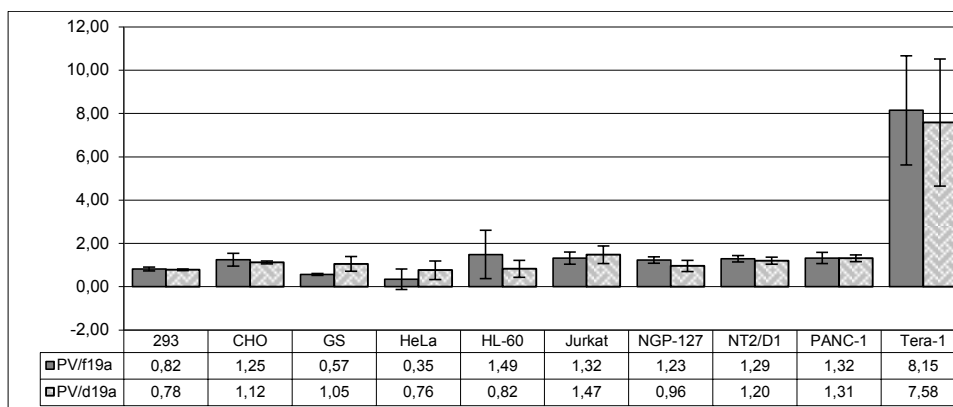


Рис. 27. Сравнение энхансерной активности полноразмерного (f19a) и частично deletированного (d19a) LTR HERV-K. Активность промотора SV40 (без энхансера) принята за 1.

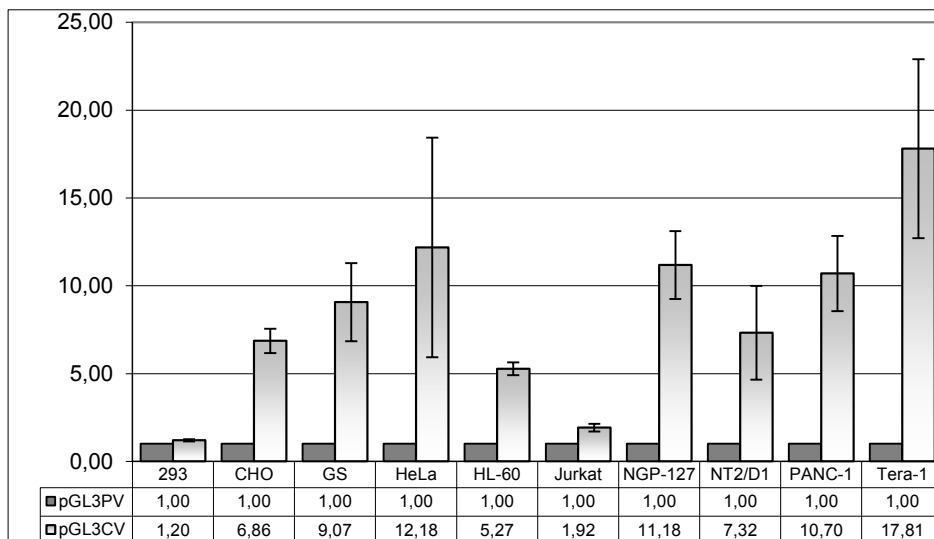


Рис. 28. Относительная активность энхансера SV40 в конструкции pGL3CV в исследованных клеточных линиях. Активность промотора SV40 без энхансера принята за 1.

#### 4.5.5. Клеточные белки, специфически связывающиеся с LTR HERV-K

Фрагменты LTR старого подсемейства HERV-K, картированного на хромосоме 19p12 (Khil et al., 1997a), были амплифицированы с использованием нескольких пар праймеров. Положения праймеров и ПЦР фрагментов в последовательностях LTR показаны на рис.29, последовательности праймеров, представлены в таблице 7. Перекрывающиеся фрагменты LTR радиоактивно метили, инкубировали с ядерным экстрактом клеток Jurkat и наносили на неденатурирующий полиакриламидный гель (рис. 30).

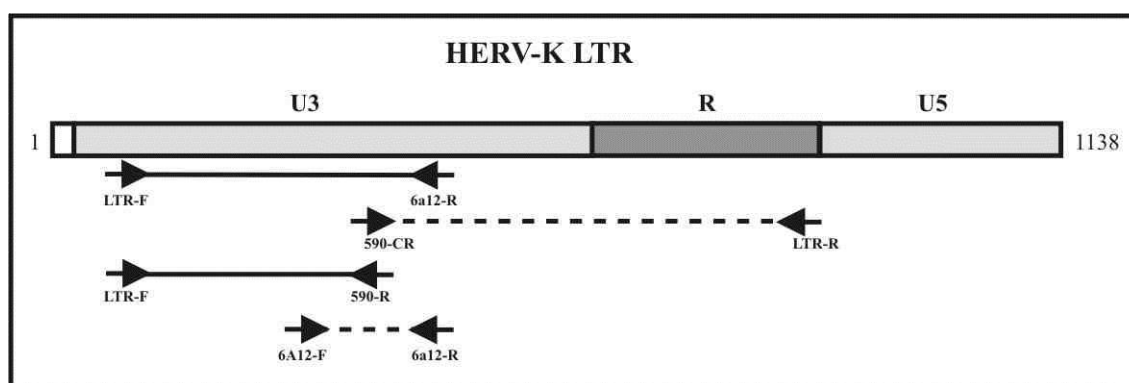


Рис. 29. Структура LTR HERV-K, положение ПЦР-праймеров и полученных фрагментов. Сплошными линиями представлены фрагменты, способные связывать ядерные белки (см. текст).

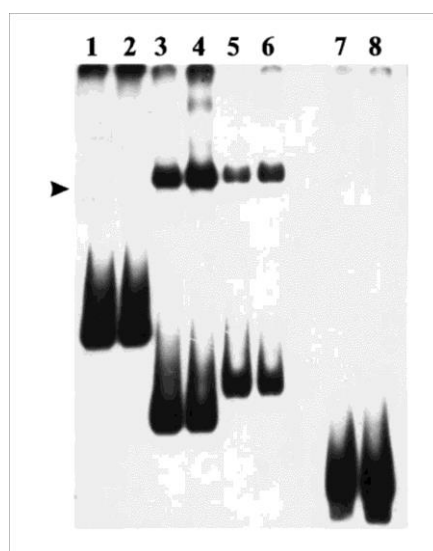


Рис.30. Анализ посредством сдвига электрофоретической подвижности с использованием фрагментов LTR HERV-K в качестве зондов и ядерных экстрактов из клеток линии Jurkat. Дорожки с нечетными и четными номерами содержат по 1 и 3 мкг белка экстракта, соответственно. Дорожки 1, 2 – фрагмент,

полученный с использованием праймеров 590-CR и LTR-R, длина 549 п.о. Дорожки 3, 4 - фрагмент, полученный с использованием праймеров LTR-F и 6A12-R, длина 294 п.о. Дорожки 5, 6 - фрагмент, полученный с использованием праймеров LTR-F и 590-R, длина 333 п.о. Дорожки 7, 8 - фрагмент, полученный с использованием праймеров 6A12-F и 6A12-R, длина 201 п.о. Стрелка указывает на одиночный основной комплекс белок-ДНК.

Из четырех использованных перекрывающихся фрагментов сильные полосы связывания белка (указаны стрелкой на рис. 30) были обнаружены для двух, амплифицированных с парами праймеров LTR-F/6A12-R и LTR-F/590-р. Эти результаты позволили локализовать уникальный сайт связывания белка, расположенный между праймерами LTR-F и 6A12 F во фрагменте длиной 230 п.о. в U3 области LTR.

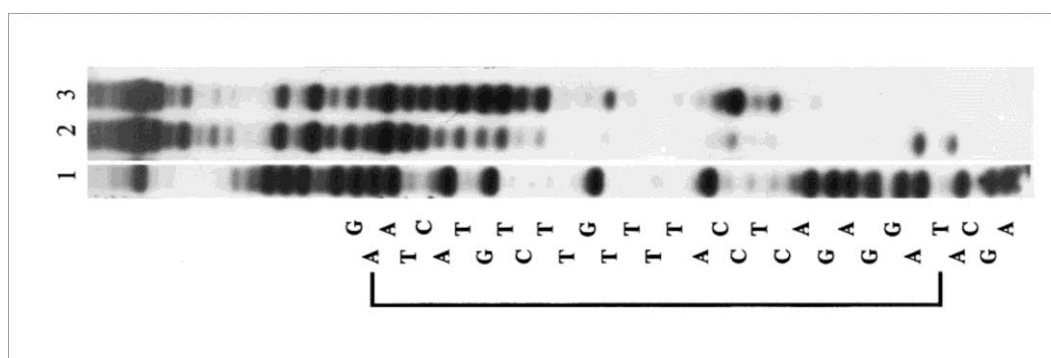


Рис. 31. Защита фрагмента LTR-F/590-R HERV-K LTR от действия ДНКазы I. Дорожка 1, A+G последовательность (маркер). Дорожка 2, к фрагменту ДНК добавлено 50 мкг очищенного белка ядерного экстракта. Дорожка 3, белок не добавлялся.

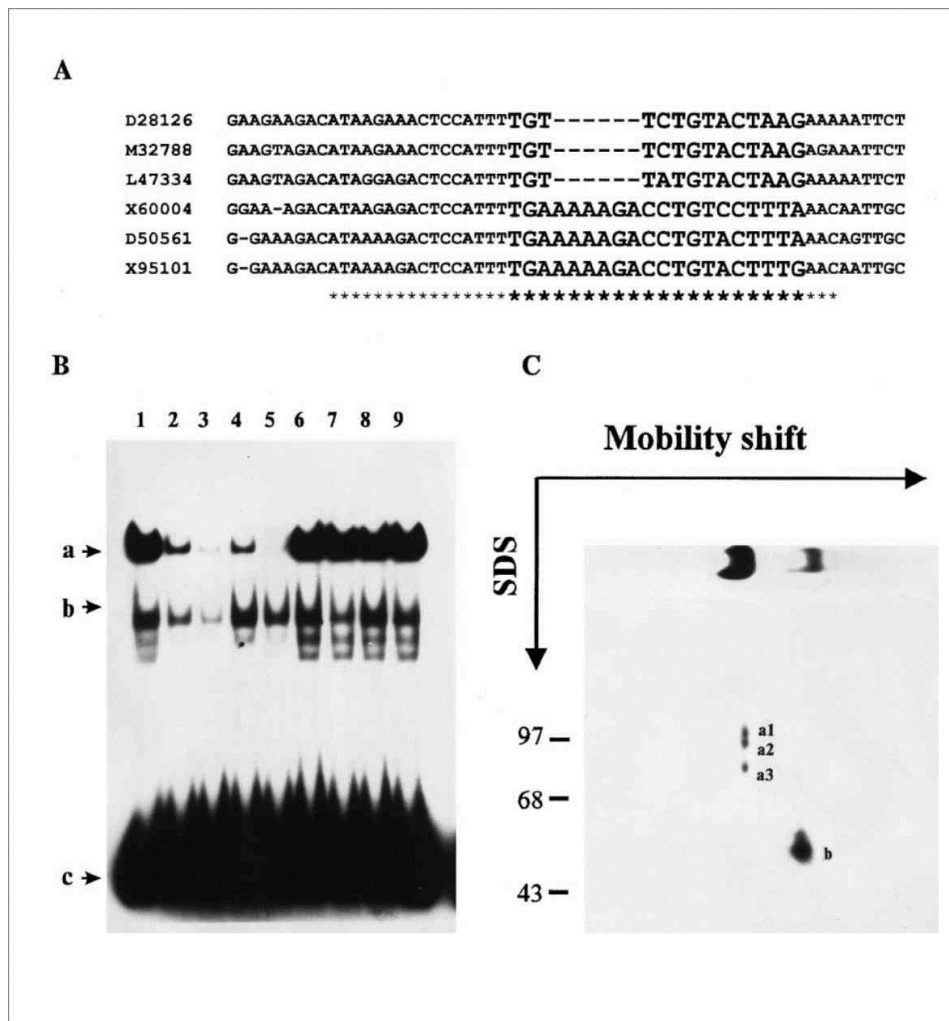


Рис. 32. Идентификация белков, связывающихся с LTR. А. Выравнивание последовательностей LTR в области связывания с белками (отмечена звездочками). GRE и коровые энхансерные элементы, ранее определенные по гомологии последовательностей, увеличены. Слева указаны регистрационные номера из GenBank соответствующих последовательностей. Полноразмерные LTR HERV-K были найдены в GenBank по гомологии с консенсусными последовательностями. Б. Анализ сдвига электрофоретической подвижности и конкуренция с использованием меченого двухцепочечного олигонуклеотида LTR1/LTR1-C (см. табл. 6) в качестве матрицы и ядерного экстракта из клеток линии Jurkat. а, б - основные комплексы белок-ДНК, с - свободный меченый олигонуклеотидный зонд. Дорожка 1 - без добавления конкурента. Дорожки 2-9, сдвиг электрофоретической подвижности с добавлением 10 и 20-кратного молярного избытка немеченых двухцепочечных олигонуклеотидов-конкурентов: дорожки 2,3 - LTR1/LTR1-C; дорожки 4,5 - LTR2/LTR2-C; дорожки 6,7 - HIV/HIV-C; дорожки 8,9 - HIVM/HIVM-C. С. Двумерный электрофорез ДНК-белковых комплексов LTR HERV-K: первое направление, неденатурирующий 10% полиакриламидный гель; второе направление, 12% SDS-полиакриламидный гель после УФ-сшивки комплексов. а1 (ERLBF1), а2 (ERLBF2), а3 (ERLBF3) и б - белки, которые связываются с LTR.

Для дальнейшей характеристики сайта связывания белка был проведен эксперимент по защите от ДНК-азы I или футпринт-анализ (рис. 31). Защищенная последовательность длиной 28 п.о. расположена в 5'-части U3 области, фланкирована сайтами гиперчувствительности к ДНК-азе I и участками, ранее идентифицированными как потенциальный энхансер и сайт связывания глюкокортикоидного рецептора (Ono, 1986; Leib-Mosch et al., 1993). Наиболее эффективно защищена последовательность, расположенная непосредственно с 5'-стороны за этими двумя сайтами.

Для оценки общей консервативности и стабильности этого региона в эволюции было сделано выравнивание полноразмерных последовательностей HERV-K LTR из GenBank. Область связывания белков показана на рис. 32А. Видно, что участок длиной 6 п.о. делегирован у трех LTR молодого подсемейства, но сохраняется у трех LTR старого подсемейства (характеристику подсемейств см. (Khil et al., 1997b)). Как видно из рис. 31, эта делеция лежит в самом центре участка, защищенного белком (белками). Для изучения роли этой делеции в связывании белка были приготовлены два двунитевых олигонуклеотида, представляющие консенсусную последовательность каждой из двух групп (табл. 11), и использованы для анализа методом сдвига электрофоретической подвижности (рис.32Б). Дорожка 1 показывает результат сдвига электрофоретической подвижности при использовании в качестве зонда двухцепочечного олигонуклеотида LTR1/LTR1-С (см. таблицу 11) и ядерного экстракта из клеток Jurkat. Этот олигонуклеотид имеет консенсусную последовательность молодого подсемейства LTR, содержащего AAAAGA (рис. 32А). Помимо двух основных видимых полос торможения а и б (рис. 32Б), детектируются также несколько слабых полос. На других дорожках показаны результаты подобного эксперимента в присутствии 10- и 20-кратного молярного избытка различных немеченых конкурентов, которые либо содержали (олигонуклеотид LTR1/LTR1-С, идентичный зонду) или не содержали (олигонуклеотид LTR2/LTR2-С) сайта, делегированного в LTR молодого

подсемейства (полосы 2, 3 и 3, 4, соответственно). Дорожки 6, 7, и 8, 9 содержат в качестве конкурентов олигонуклеотиды, представляющие соответственно нативную и мутированную энхансерные области вируса иммунодефицита человека типа 1 (Nikolaev, 1996). Видно, что белок (белки) главной полосы торможения (а) легко вытесняется из комплекса с меченым олигонуклеотидом относительно небольшим молярным избытком обоих олигонуклеотидных конкурентов независимо от наличия б п.о.делеции. В то же время последовательность лентивирусного энхансера не конкурирует за связывание белковых факторов, связывающихся с последовательностью LTR. Интенсивность полосы (b) изменялась незначительно при добавлении как олигонуклеотидов HERV-K, так и ВИЧ-1, указывая низкую специфичность связывания белка в этом комплексе. Таким образом, только белок (белки), отвечающие за появление полосы торможения (а) можно считать специфически связывающимися с LTR HERV-K. Эксперименты по сдвигу электрофоретической подвижности с использованием двухцепочечного олигонуклеотида LTR2/LTR2-C в качестве меченого зонда и более короткого LTR1/LTR1-C в качестве конкурента показали, что последний также способен вытеснять белок из комплекса (данные не показаны). Следовательно, несмотря на существенные различия между первичными структурами участка связывания, LTR обоих подсемейств HERV-K способны к образованию специфического комплекса с тем же самым белком (белками) ядерного экстракта клеток Jurkat. В экспериментах по сдвигу электрофоретической подвижности с ядерными экстрактами клеток HeLa и CHO образуются те же самые две основных полосы смещения, со слегка меньшей подвижностью в последнем случае (данные не показаны). Таким образом, белки, которые образуют комплексы а и b с последовательностью HERV-K LTR, присутствуют в различных типах клеток человека и млекопитающих.

Чтобы охарактеризовать ДНК-белковые комплексы более подробно, мы использовали технику ультрафиолетовой сшивки (Vogomolova et al., 1989; Nikolaev, 1996). Комплексы ДНК-белок были разделены методом торможения в геле (как на рис. 9Б), соответствующую полосу вырезали из геля, облучали УФ-

светом и сшитые комплексы белка с мечеными олигонуклеотидами разделяли во втором направлении в соответствии с их молекулярными массами (рис. 32С). Неспецифический комплекс b (рис. 32Б), содержал одиночный белок (b на рис.32) с кажущейся молекулярной массой 56 кДа, в то время как главная полоса торможения содержала три белка с молекулярными массами 98, 91 и 81 кДа.

Все белковые пятна были расположены строго на одной вертикальной линии, что указывает на то, что они были частью одного комплекса ДНК-белок. Все три белка сшиваются с ДНК УФ-светом, т.е. они находятся в тесном контакте с ней. Это объясняет довольно большую протяженность (около 30 п.о.) области, защищенной этими белками от расщепления ДНК-азой I. Мы обозначили белки как ERLBF1, ERLBF2 и ERLBF3 (Endogenous Retrovirus LTR Binding Factors) в порядке убывания их молекулярных масс. Нужно отметить, что комплекс может включать в себя и другие белки, участвующие в связывании с ДНК через белок-белковые взаимодействия.

Поиск в базе данных TRANSFAC (Matys et al., 2006) не выявил белков, способных связывать указанные выше олигонуклеотиды и имеющие молекулярные массы, близкие к ERLBF1, ERLBF2 и ERLBF3.

Далее мы локализовали участок связывания этих факторов с большей точностью. С этой целью был синтезирован набор двунитевых олигонуклеотидов, перекрывающих различные участки внутри последовательности LTR, и их способность связывать комплекс ERLBF была проверена методом EMSA с ядерными экстрактами из различных линий клеток (рис. 33). Как видно из рисунка, короткий олигонуклеотид длиной 21 п.о. (рис. 33В) способен связываться специфически с белками, но вместо одной полосы торможения ДНК-белкового комплекса наблюдаются две.

Для определения белкового состава комплексов с коротким (21 п.о.) олигонуклеотидом, мы провели двумерный анализ EMSA с промежуточным сшиванием ДНК с белками при помощи ультрафиолетового излучения. Из рисунка (рис. 34Б) видно, что один из комплексов с олигонуклеотидом длиной 21 п.о. включает белковые факторы ERLBF1, ERLBF2 и ERLBF3, а другой (с



большой электрофоретической подвижностью) – содержит только факторы ERLBF1 и ERLBF2. Вместе с тем, общий белковый состав комплекса с олигонуклеотидом длиной 21 п.о. не отличается от такового для комплекса, образованного с более длинным олигонуклеотидом (рис. 34Б).

Следует отметить, что состав комплексов области U3 LTR HERV-K сильно зависит от состояния клеток. Известно, что тепловой шок (Moggs and Orphanides, 2003) и митогены, такие как фитогемагглютинин и форболовые эфиры (Gurskaya et al., 1996), оказывают большое влияние на профиль

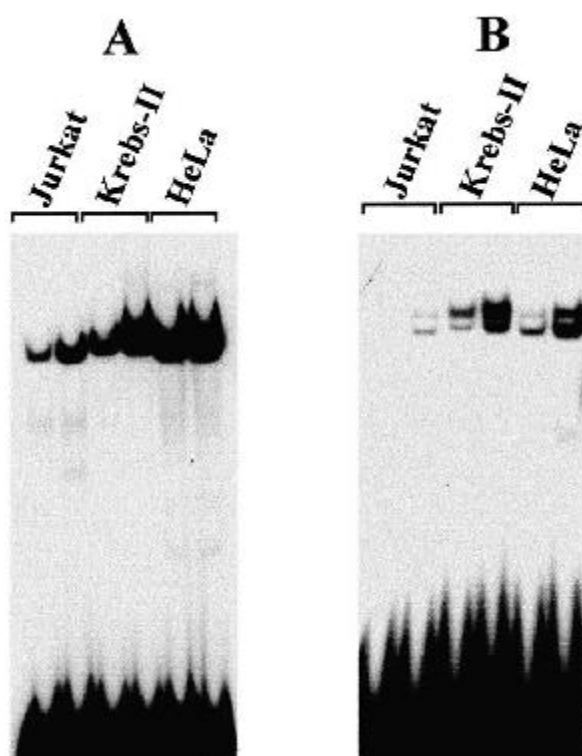


Рис. 33. Электрофореграмма сдвига подвижности (EMSA) ДНК-белковых комплексов, сформированных при добавлении 5 и 10 мкг белка ядерных экстрактов из клеточных линий Jurkat, Krebs-II и HeLa. В качестве зонда использовали двунитевые олигонуклеотиды (А) 30 п.о. и (В) 21 п.о., содержащие сайт связывания ERLBF.

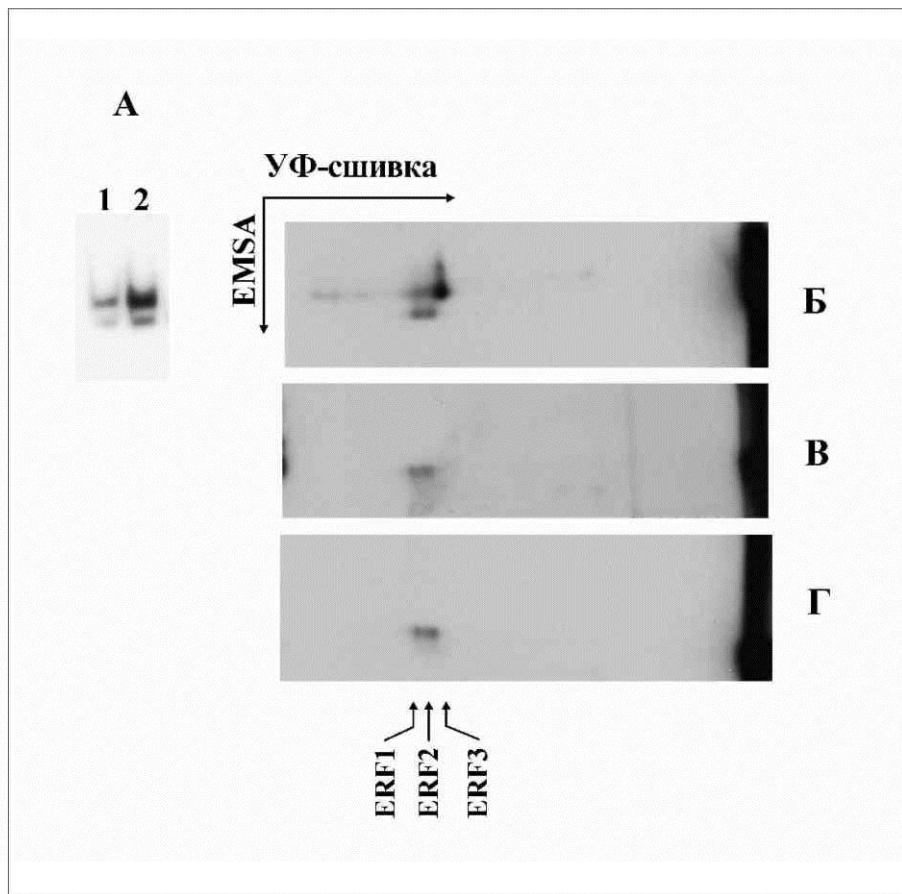


Рис. 34. Электрофореграмма сдвига подвижности (EMSA) с использованием олигонуклеотида длиной 21 п.о. (А), и белковый состав комплексов (Б, В, Г), полученный по результатам двумерного EMSA с ультрафиолетовой сшивкой. Описание в тексте.

экспрессии генов в клетке. Митогены, например, индуцируют продукцию эндогенных ретровирусов (Tacke et al., 2003). Мы исследовали влияние теплового шока и индукции клеток Jurkat митогенами на состав белкового комплекса, связывающегося с 5'-частью области U3 LTR HERV-K (рис. 34В, Г). Из рисунка видно, что в условиях теплового шока и при действии митогенов ДНК-белковый комплекс утрачивает фактор ERLBF3, тогда как факторы ERLBF1 и ERLBF2 остаются связанными с ДНК.

Полученные результаты показывают, что, по крайней мере, некоторые LTR HERV-K сохранили способность специфического взаимодействия с белками клеток, что является важным условием их участия в функциональной регуляции генома.

#### 4.5.6. Связывание белков другими участками LTR

Для изучения связывания ядерных белков использовали три перекрывающихся фрагмента LTR L47334, которые получали при помощи ПЦР-амплификации с парами праймеров, положение которых приведено на рис. 35, а последовательности – в таблице 11. Полученные фрагменты радиоактивно метили (см. Материалы и методы) и связывающиеся с ними белки выявляли при помощи EMSA (рис.36) с ядерными экстрактами из линии клеток Jurkat. Из рисунка видно, что полосы, соответствующие ДНК-белковым комплексам, выявляются для всех трёх фрагментов LTR. Таким образом, все три фрагмента LTR способны связывать белки, отличающиеся по электрофоретической подвижности, при этом наиболее активно комплекс образуется между 5'-концевым фрагментом LTR и факторами ERLBF.

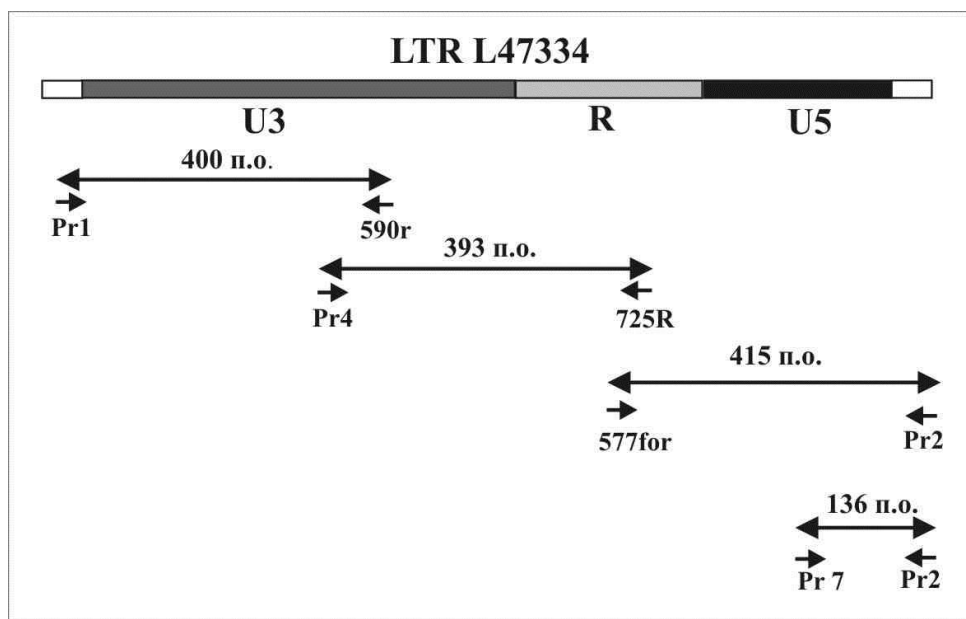


Рис. 35. Структура LTR L47334 и расположение ПЦР-фрагментов, использованных для EMSA. Состав праймеров приведен в таблице 12.

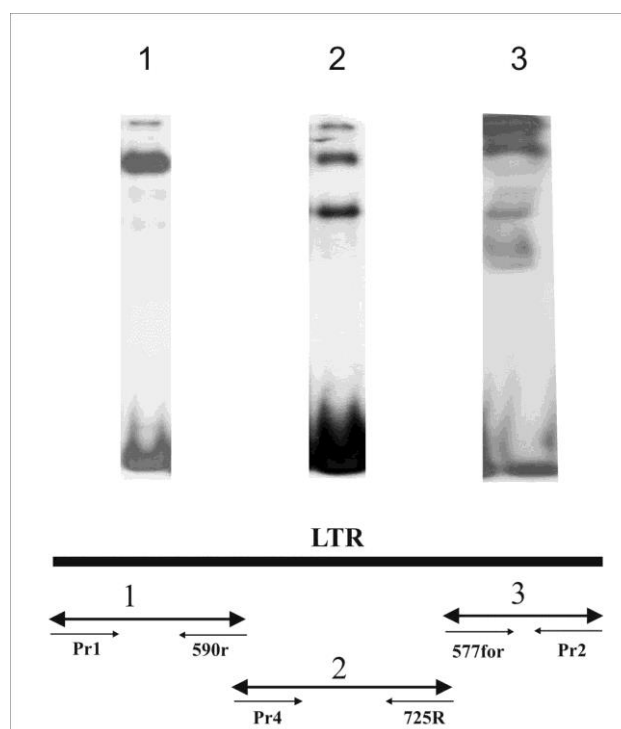


Рис. 36. Эксперименты по сдвигу электрофоретической подвижности (EMSA) с различными фрагментами LTR L47334. Ядерный экстракт из клеток линии Jurkat.

Белки, образующие комплексы с фрагментами LTR, выявляются в ядерных экстрактах из разных клеточных линий. Кроме линии клеток Jurkat, нами были проверены клетки линий эмбриональных карцином Tera-1 и NT2/D1, почки эмбриона человека HEK-293, нейробластомы NGP-127 (рис. 37), а также клеток HeLa и асцитной опухоли мышей Krebs-II (рис. 33, только для комплекса с факторами ERLBF). Наличие белков ERLBF в клетках мышей, геном которых не содержит HERV-K, позволяет предположить, что эти белки связываются с регуляторными последовательностями не только HERV, но и клеточных генов.

Специфичность взаимодействия белковых комплексов с LTR подтверждают эксперименты по конкуренции. Конкурентное связывание представляет собой добавление в реакционную смесь, наряду с радиоактивно меченым специфическим фрагментом, избытка того же самого фрагмента без метки. В результате из-за конкуренции белок взаимодействует с избытком специфического немеченого фрагмента, сигнал на радиоавтографе уменьшается по отношению к контролю (белок, связавшийся с меченым фрагментом). На рис. 38 показано уменьшение и исчезновение сигнала в области комплекса олигонуклеотида с

белками ERLBF по мере увеличения концентрации немеченого специфического олигонуклеотида, что свидетельствует о специфичности связывания белка ERLBF с U3 областью LTR.

#### 4.5.6. Связывание белков с негативным регуляторным элементом (NRE)

Выше было показано, что частичная делеция области U5 приводила к повышению промоторной активности LTR независимо от его ориентации. Подобные негативные регуляторные элементы были описаны для пенистого

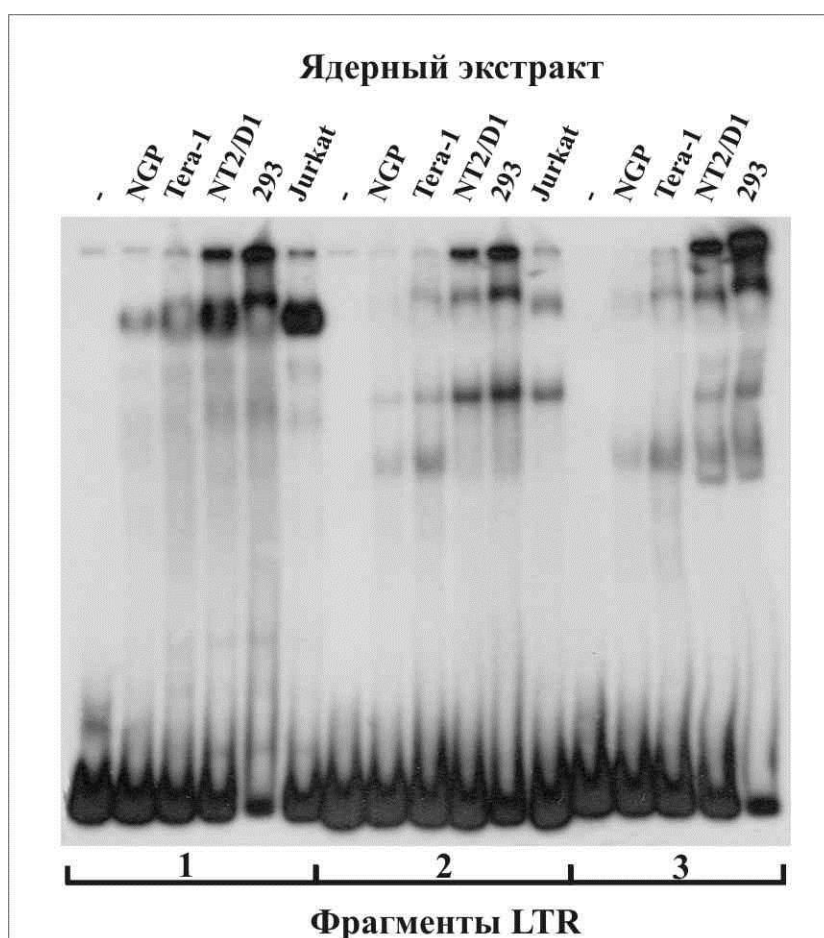


Рис. 37. Сдвиг электрофоретической подвижности при применении различных фрагментов LTR L47334. Расположение фрагментов LTR приведено на рис. 35. Над соответствующими дорожками указаны линии клеток, из которых получали ядерный экстракт.

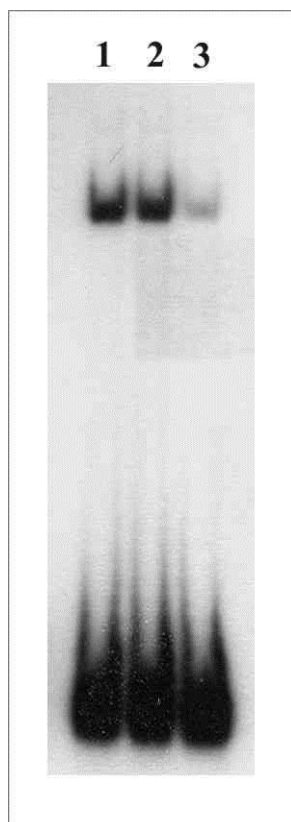


Рис. 38. Конкуренция за связывание с белками ERLBF меченого и немеченого двунитевых олигонуклеотидов. Дор. 1 - исходный ДНК-белковый комплекс, дор. 2 – двукратный избыток, дор. 3 - 5-кратный избыток немеченого олигонуклеотида.

вируса (foamy virus) человека (Seiki et al., 1990), и вируса Т-клеточного лейкоза 1 (HTLV-1) (Seiki et al., 1990, Okumura et al., 1996).

Для проверки способности содержащего НРЭ фрагмента специфически связывать клеточные белки, фрагмент LTR L47334 длиной 136 п.о. ПЦР-амплифицировали с использованием праймеров Pr2 и Pr7 (см. таблицу 12), радиоактивно метили и использовали для EMSA с ядерными экстрактами из нескольких клеточных линий (рис. 39). Видно, что белки, способные связываться с NRE-содержащим фрагментом, обнаруживаются во всех проверенных клеточных линиях. В экспериментах с конкурентным связыванием немеченого фрагмента, содержащего НРЭ, была подтверждена специфичность взаимодействия (не показано).

Количество связавшихся белков было различным для ядерных экстрактов из различных клеток, о чем свидетельствует различная интенсивность

соответствующих ДНК-белковым комплексам полос. При этом интенсивность полос связывания, соответствующих комплексам белковых факторов ERLBF1, 2 и 3 с 5'-участком области U3 LTR, была практически одинаковой при использовании разных ядерных экстрактов (рис. 39).

Таким образом, область LTR L47334, содержащая негативный регуляторный элемент, тканеспецифично связывает клеточные белки, которые, возможно, отвечают за свойства НРЭ.

#### *4.5.8. Идентификация белков, потенциально ответственных за энхансерную активность LTR HERV-K*

Ранее мы показали, что энхансерные свойства LTR HERV-K существенно различаются для различных линий клеток. Так LTR L47334 в линии клеток Tera-1 обладал высокой энхансерной активностью (примерно в 10 раз активировал ранний промотор вируса SV40), а в клеточной линии NT2/D1 энхансерная активность LTR практически отсутствовала (Domanskii et al., 2002). Одним из объяснений этого факта может быть различие наборов клеточных регуляторных

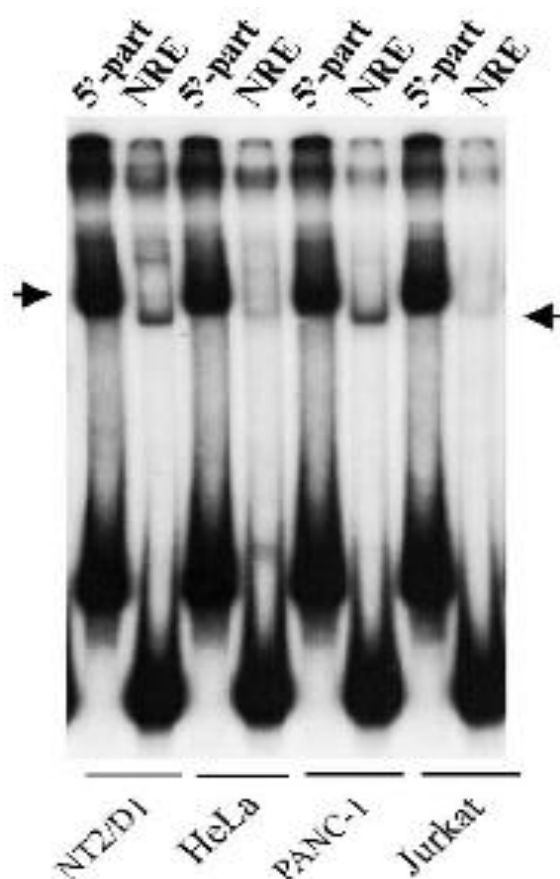


Рис. 39. Сдвиг электрофоретической подвижности (EMSA) с использованием в качестве зонда 5'-участка U3-района LTR L47334 и участка U5-района, содержащего негативный регуляторный элемент (НРЭ, амплифицирован с применением праймеров Pr2 и Pr7, см. табл. 12). Под соответствующими дорожками приведены клеточные линии, использованные для выделения ядерных экстрактов. Стрелками обозначены два основных ДНК-белковых комплекса.

белков, связывающихся с LTR в этих типах клеток. Для проверки этого предположения мы провели эксперименты по EMSA с использованием ядерных экстрактов из клеток Tera-1, NT2/D1 и Jurkat (рис. 40).

Таблица 12. Структуры олигонуклеотидов (праймеров).

Праймер	Позиция	Структура (5'-3')
<b>590r</b> (ss)	Внутренний праймер обратный LTR L47334	GCCATATTTTCAGACTATCACATGG
<b>Pr1</b> (ss)	Фланкирующий праймер прямой LTR L47334	CAGTCTTATCTCCTTTACTGACC
<b>Pr2</b> (ss)	Фланкирующий праймер обратный LTR L47334	CCTCGTGTGTTGTGCTTG
<b>Pr4</b> (ss)	Внутренний праймер прямой LTR L47334	ATTGTCCAAGGTTTCTCCC
<b>577for</b> (ss)	Внутренний праймер прямой LTR L47334	GCCTTAGGGCTGGAGGTG
<b>725R</b> (ss)	Внутренний праймер обратный LTR L47334	CAGCAGACAAACACGTGAAC
<b>Pr7</b> (ss)	Внутренний праймер LTR L47334	GGGTCCCCTTATTTCTTTCTC
<b>C1/C2</b> (ds)	Двухцепочечный олигонуклеотид специфичный к ERLBF.	AGGATCGGGGCGGGGCGATC и CTGATCGCCCCGCCCGATC

На рис. 40А не выявляются различия в связывании факторов ERLBF между ядерными экстрактами из клеток Tera-1, NT2/D1 и Jurkat. Вероятно, тканеспецифичность энхансерной активности LTR не зависит от формирования комплекса с белками ERLBF. Между тем, комплексы, образованные центральной областью LTR с белками ядерных экстрактов линии клеток Tera-1, заметно отличаются от комплексов с ядерными экстрактами линии NT2/D1 или Jurkat (рис. 40В) – появляется полоса торможения ДНК-белкового комплекса с большей электрофоретической подвижностью. Можно сделать предположение, что белок или белки, образующие комплекс с высокой электрофоретической подвижностью,



ответственны за высокую энхансерную активность LTR в клеточной линии Tera-1.

Таким образом, различный набор клеточных белков, связывающихся с центральной областью LTR, может объяснять различие в энхансерной активности между Tera-1 и NT2/D1.

#### 4.5.9. Выделение и идентификация белков ERLBF1, 2 и 3, специфически связывающихся с LTR HERV-K

Белки, специфически связывающиеся с регуляторными элементами LTR человеческих эндогенных ретровирусов семейства HERV-K (см. Обзор литературы), были впервые идентифицированы нами в ядерных экстрактах клеток человека линий Jurkat и CHO, и обозначены как ERLBF1, 2 и 3.

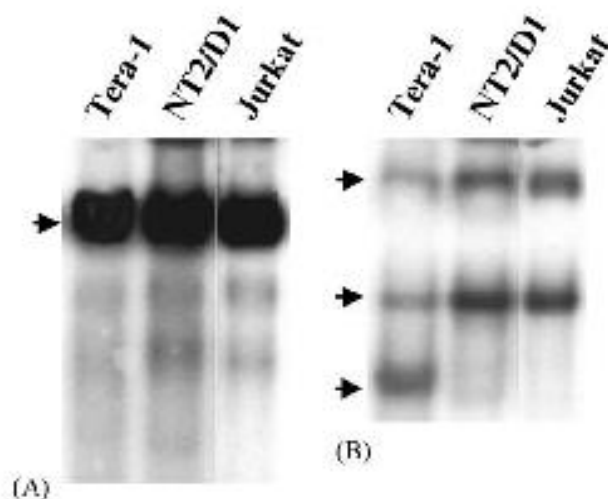


Рис. 40. Сдвиг электрофоретической подвижности (EMSA) с использованием (А) 5'-участка (амплифицирован с применением праймеров Pr1 и 590-R) и (В) центральной части (амплифицирована с применением праймеров Pr4 и 725 R) LTR L47334 (расположение фрагментов приведено на рис. 35, а структура праймеров – в таблице 12). Вверху приведены клеточные линии, использованные для выделения ядерных экстрактов. Стрелками обозначены основные ДНК-белковые комплексы. Несвязавшийся радиоактивный зонд выгоняли из геля для лучшего разделения ДНК-белковых комплексов.

Ядерные экстракты из линий клеток содержат небольшие количества регуляторных белков, и для их очистки требуются десятки миллилитров ядерного

экстракта, что соответствует десяткам литров суспензионной культуры. Ранее группу факторов ERLBF мы выявляли в клетках китайского хомячка, и, так как системы регуляции транскрипции у человека и грызунов весьма похожи, мы предположили, что искомые белки могут содержаться и в опухолевых клетках грызунов. Поэтому в качестве источника ядерных белков далее использовали клетки перевиваемой асцитной карциномы мышей. В результате заражения мышей линии BALB/C получали асцитную жидкость, содержащую клетки эпителиальной асцитной карциномы мышей Krebs-II, из которых выделяли ядерный экстракт. Из рис. 41 видно, что в ядерном экстракте из асцитной опухоли мышей присутствует интересующий нас специфический белковый комплекс в количествах, сравнимых с содержанием его в суспензионных культурах клеток (например, Jurkat).

Основным методом для выделения и очистки ДНК-связывающих белков из ядерных экстрактов клеток является метод ДНК-

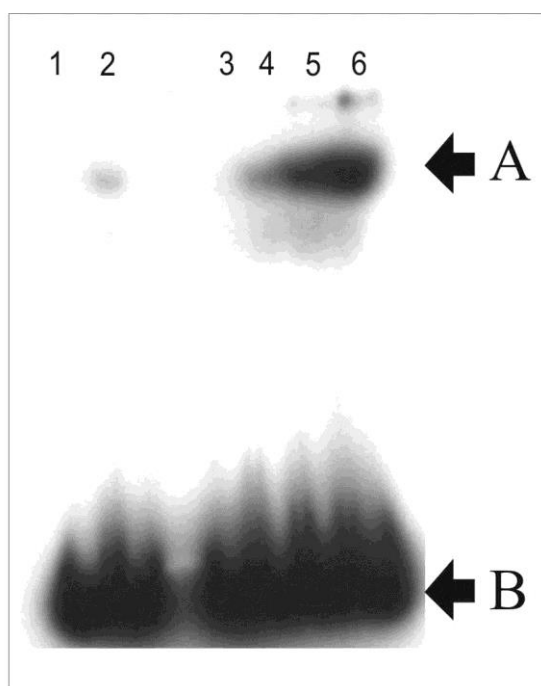


Рис. 41. Сопоставление содержания комплексов олигонуклеотида с белками ERLBF ядерных экстрактов, приготовленных из клеток линии Jurkat и асцитной карциномы мышей Krebs-II. Дорожки 1,2 - ядерный экстракт из клеток линии

Jurkat. Дорожки 3-6 - ядерный экстракт из клеток асцитной эпителиальной карциномы мышей. А - ДНК-белковый комплекс, В - свободный олигонуклеотид.

аффинной хроматографии (Kadonaga and Tjian, 1986). Используется аффинная колонка, в которой к сефарозе пришиты предварительно лигированные олигонуклеотиды, содержащие сайт связывания определенного белка (Arndt-Jovin et al., 1975). На следующей стадии производится солевая элюция с сорбента посторонних белков, образующих с олигонуклеотидами наименее прочные комплексы, а на третьей стадии белки, обладающие большей аффинностью и образующие с олигонуклеотидом прочные комплексы, элюируют с сорбента буфером с высокой концентрацией соли (0,4-0,5 М КСl). В процессе элюции слабо связывающихся белков происходят большие (до 80%) потери специфических белков (таблица 13), их окончательный выход снижается. В целом, метод достаточно трудоёмок, так как требует приготовления специального аффинного носителя для конкретного выделяемого белка (комплекса белков). К тому же, большие потери специфических белков в процессе их выделения требуют значительного увеличения объема исходного материала и приводят к возрастанию расхода сопутствующих материалов на всех этапах получения белков в количествах, необходимых для их надежной идентификации.

Для оптимизации процедуры выделения и выхода факторов ERLBF мы разработали альтернативный метод очистки ДНК-связывающих белков с использованием иммобилизованного на агарозе гепарина. Гепарин довольно часто используется при выделении ДНК-связывающих белков (Genersch et al., 1995; Cheng et al., 1997; Gadgil and Jarrett, 1999). Его повторяющееся олигосахаридное звено можно рассматривать как аналог фосфатного остова ДНК, поэтому большинство белков, специфически связывающихся с ДНК, имеют сродство к гепарину. В то же время, сродство этих белков к ДНК, содержащей последовательность их специфических сайтов связывания, гораздо выше, чем к гепарину, что позволяет быстро и эффективно выделять их при помощи элюции низкосолевым буфером, содержащим такую ДНК (рис. 42). На первом этапе белки ядерного экстракта из клеток асцитной карциномы мышей Krebs-II

наносили на колонку с гепарин-агарозным носителем, как описано выше для первого этапа ДНК-аффинной хроматографии. Затем колонку промывали буфером, содержащим 0,2 М КСl, т.е., в концентрации, при которой специфические белки с колонки не элюируются. Далее белковые факторы ERLBF элюировали тем же буфером с добавлением меченого <sup>32</sup>P двунитевого синтетического олигонуклеотида, содержащего последовательность, соответствующую участку связывания ERLBF. Элюированные с колонки фракции анализировали методом EMSA. На рис. 43 видно, что олигонуклеотид сорбирует белки из комплекса с гепарином, в результате чего на выходе с колонки выявляются высокоаффинные комплексы белков с олигонуклеотидом, совпадающие по электрофоретической подвижности с комплексами, получаемыми при использовании исходного ядерного экстракта. Разработанный нами метод очистки ДНК-связывающих белков дает возможность получать эти белки в виде комплексов с олигонуклеотидом.

Таблица 13. Очистка ДНК-связывающих белков на гепарин-агарозе с помощью аффинной элюции олигонуклеотидом.

Фракция	Белок (мкг)	Белок (%)	Связывающая активность (%)	Кратность очистки
Исходный ядерный экстракт	3200	100	100	1
Белки, элюированные с гепарин-агарозы	270	8,5	92	11
Белки, элюированные с ДНК-аффинной колонки	0,16	0,005	10	2000
Белки, элюированные с гепарин-агарозы двухцепочечным олигонуклеотидом	1,8	0,06	80	1400

Для оптимизации процесса очистки факторов ERLBF были проведены следующие эксперименты:

1. Определение ёмкости гепарин-агарозной колонки. В этой серии экспериментов колонку с гепарин-агарозой (0,5 мл) последовательно нагружали возрастающими количествами ядерного экстракта. Из рис. 44 видно, что насыщение гепарин-

агарозы факторами ERLBF происходит при нанесении на колонку 6,4 мг белка ядерного экстракта. Во всех экспериментах этой серии на колонку наносили не более 4 мг исходного белка.

2. Анализ полноты элюции факторов ERLBF с колонки двунитевым специфическим олигонуклеотидом. После элюции белков посредством олигонуклеотида, колонку промывали 0,5 М КСl. Из рис. 45 видно, что количество белков, элюируемых дополнительно высокосолевым буфером, незначительно.

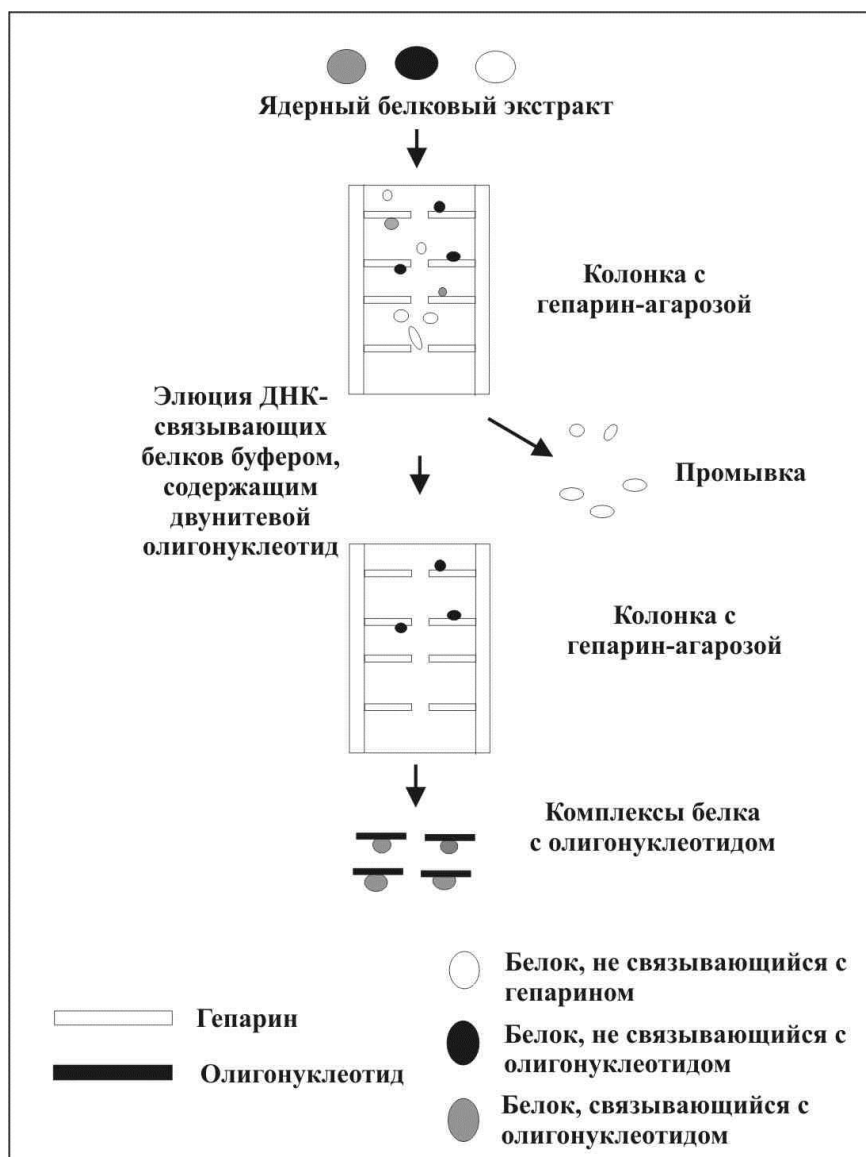


Рис. 42. Схема очистки ДНК-связывающих белков методом аффинной элюции.

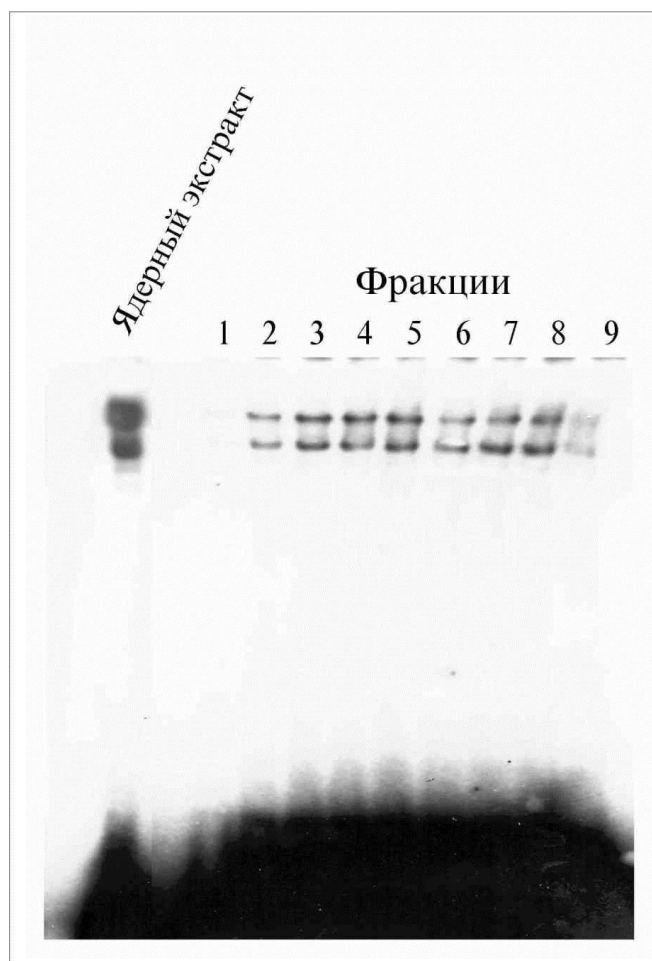


Рис. 43. Анализ при помощи метода сдвига электрофоретической подвижности (EMSA) фракций после аффинной элюции белков с гепарин-агарозной колонки меченым  $^{32}\text{P}$  олигонуклеотидом, содержащим сайты связывания ERLBF.

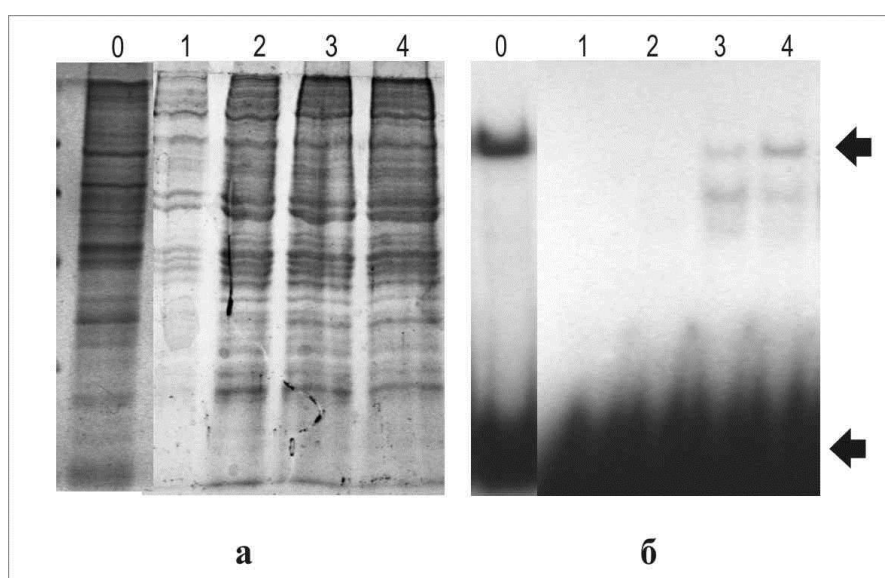


Рис. 44. Определение ёмкости гепарин-агарозной колонки. а - электрофорез белков в полиакриламидном геле с ддс-натрия. б - сдвиг электрофоретической подвижности. Дорожка 0 - исходный ядерный экстракт; дорожки 1-4 – фракции, собранные с колонки. Стрелками обозначены позиции ДНК-белкового комплекса и свободного фрагмента ДНК.

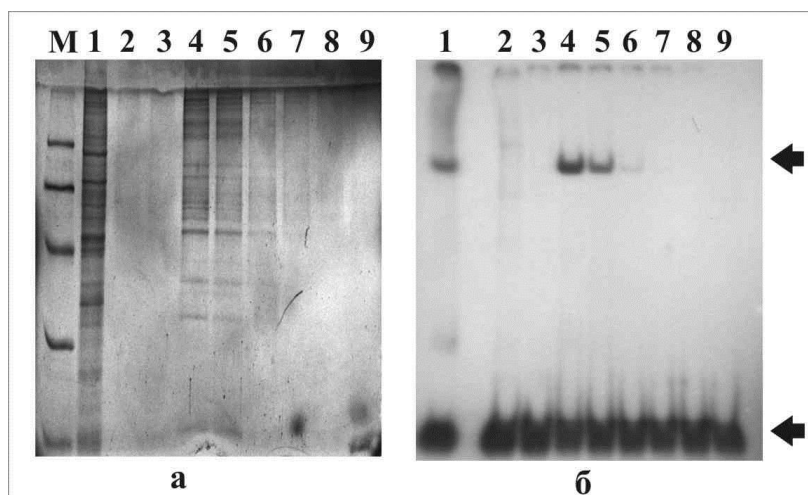


Рис. 45. Промывка гепарин-агарозы 0,5 М КСl после элюции олигонуклеотидом. а - электрофорез белков в полиакриламидном геле с ддс-натрия. б - сдвиг электрофоретической подвижности. Дорожка 1 - исходный ядерный экстракт; дорожки 2-9 – фракции, собранные с колонки. Стрелками обозначены позиции ДНК-белкового комплекса и свободного фрагмента ДНК. М - маркер молекулярных масс.

Выделение факторов ERLBF из 1 мл ядерного экстракта (концентрация тотального белка 2,5 мг/мл) проводили после оптимизации условий с использованием не содержащего радиоактивной метки олигонуклеотида. Элюат концентрировали, переосажая трихлоруксусной кислотой, и анализировали электрофорезом в денатурирующих условиях. На рис. 46 видны три мажорных белка с молекулярными массами от 60 до 70 кДа, не отличающиеся от полученных методом аффинной хроматографии. В таблице 13 приведены результаты очистки факторов ERLBF методом аффинной элюции в сравнении с методом ДНК-аффинной хроматографии. Анализ представленных в таблице результатов позволяет сделать вывод, что использование примененного нами метода позволяет получить из ядерного экстракта с выходом примерно на порядок большим, чем при ДНК-аффинной хроматографии, в одну стадию

высокоочищенные ДНК-связывающие белки при кратности их очистки в 1400 раз.

Метод аффинной элюции обладает следующими преимуществами по сравнению с ДНК-аффинной хроматографией:

1. Простота и высокая скорость выделения - искомый белок может быть выделен в одну стадию за один день;
2. В несколько раз меньшие потери целевого белка;
3. Отсутствие необходимости приготовления специального аффинного носителя для каждого выделяемого белка.

Очевидно, существует возможность выделения нескольких различных белков с различной специфичностью связывания из одной порции ядерного экстракта при их элюции различными олигонуклеотидами с одной и той же колонки.

Белки ERLBF1, 2 и 3, выделенные после аффинной элюции с гепарин-агарозы двунитевым олигонуклеотидом и затем разделенные с помощью ЭФ в ПААГ (рис. 46), анализировали методом MALDI-TOF-MS (Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry) (Shevchenko et al., 1996). Каждый из белков ERLBF(1-3) гидролизovali



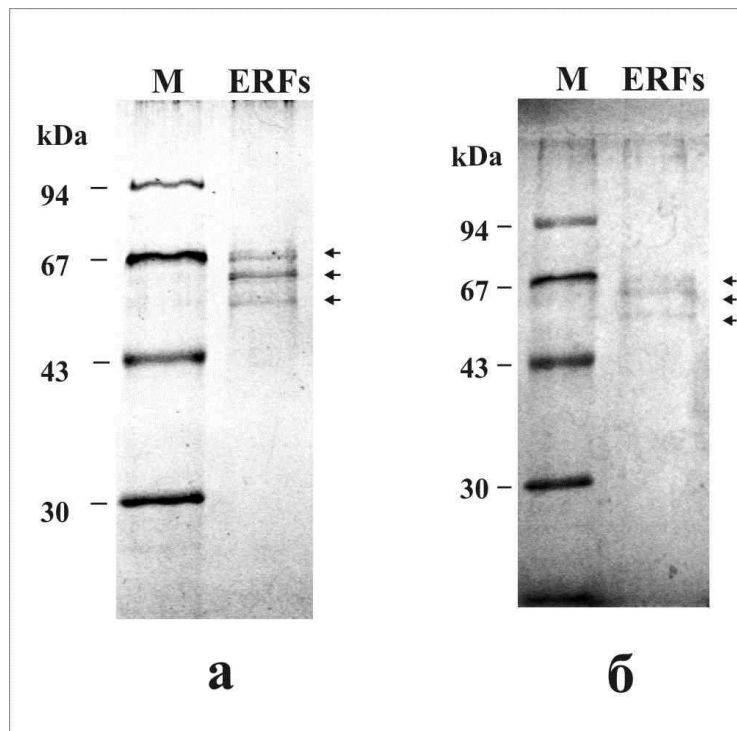


Рис. 46. Электрофорез в полиакриламидном геле белков ERLBF, полученных двумя различными способами. (а) белковые факторы ERLBF, полученные при помощи аффинной элюции (окраска Brilliant Blue R); (б) белковые факторы ERLBF, полученные методом ДНК-аффинной хроматографии (окраска серебряным реагентом). М - маркер молекулярных масс.

трипсином и осуществляли масс-спектрометрический анализ полученных пептидов. Молекулярные массы выявленных пептидов сравнивали с молекулярными массами пептидов, входящих в базу данных MASCOT с последующей идентификацией наиболее вероятных белков, содержащих выявленные пептиды. По результатам анализа ERLBF-1 (Score 115) соответствовал конститутивному белку теплового шока Hsc70 (Heat shock 70 kDa protein 8, isoform 1, Homo sapiens) (рис. 47). MALDI-TOF-MS-анализ белков ERLBF(1-3) проводили 6 раз, при этом для выделения белков использовали ядерный экстракт, полученный в разное время. В итоге определили значительную часть аминокислотной последовательности ERLBF-1, которая на 66% совпадает со структурой белка теплового шока Hsc70 (рис. 48).

MASCOT-анализ пептидов, полученных в результате триптического гидролиза ERLBF-2 и ERLBF-3, не выявил известных белков, включая

гипотетические. Вероятно, ERLBF2 и ERLBF3 могут быть новыми белками, поскольку в настоящее время отсутствуют данные о непосредственном взаимодействии белков теплового шока, в частности Hsc70, с нуклеиновыми кислотами, возможно, что эту функцию в составе комплекса ERLBF(1-3)-ДНК выполняет ERLBF-2.

Дальнейшую идентификацию ERLBF проводили при помощи иммунологического анализа. Поликлональные антитела (IgG фракция) к рекомбинантному Hsc70 (любезно предоставлен д-ром M. Ladjimi, Universite Pierre & Marie Curie, France) были получены Е.В. Свирщевой (ИБХ РАН). EMSA анализ проводили по стандартной методике, но к белку в составе реакционной смеси EMSA добавляли антитела. При специфическом взаимодействии белка с антителами, связывающимися с меченым фрагментом, формируется комплекс ДНК с белком и антителом, обладающий меньшей электрофоретической подвижностью, чем исходный ДНК-белковый комплекс. Дополнительный сдвиг электрофоретической подвижности (supershift) свидетельствует о том, что добавленные антитела специфически связываются с белком, входящим в комплекс с ДНК. В присутствии антител (рис. 49) образуется комплекс ДНК-белок-антитело, который имеет меньшую электрофоретическую подвижность, чем комплекс ДНК-белок.

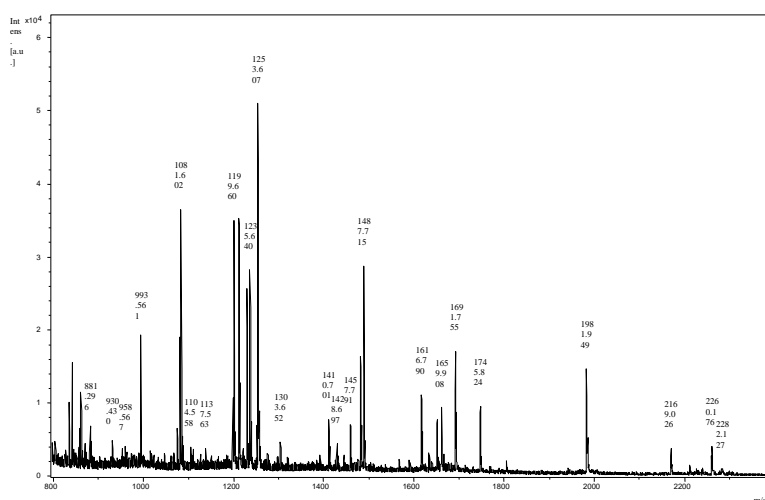


Рис. 47. Масс-спектр фракций триптического гидролизата белкового фактора ERLBF-1.

mskgpavgidlgttyscvgvfqhgkveiiandqgnrttpsyvaftdterligdaaknq  
vamnpntntvfdakrligrrfddavvqsdmkhwpfmvndagrpkvqveykgetksfyp  
eevssmvltkmkeiaeaylgktvtnavvtvpayfndsqrqatkdagtiaglnvlriin  
eptaaaiaygldkkvgaernvlifdlgggtfdvsiltiedgifevkstagdthlgged  
fdnrmvnhfiaefkrkhkkdisenkravrrlrtacerakrtlsstqasieidslyeg  
idfytsitrarfeelnadlfrgtldpvekalrdakldksqihdivlvggstripkiqk  
llqdffngkelnksinpdeavaygaavqaailsgdksenvqdl111ldvtplslgieta  
ggvmtvlikrnttiptktqtfttysdnqpgvliqvyegeramtkdnnllgkfeltgi  
ppaprgvpqievtfdidangilnvsavdkstgkenkititndkgrlskediermvgea  
ekykaedekqrdkvssknslesyafnmkatvedeklggkindedkqkildkcneiinw  
ldknqtaeefehqqkelekvcnpiitklyqsaggmpgg  
mpggfpgggappsggassgptieevd

Рис. 48. Аминокислотная последовательность белка Hsc70 (AАН16660, Heat shock 70kDa protein 8, isoform 1, Homo sapiens). Идентифицированные триптические пептиды подчёркнуты.

Таким образом, поскольку антитела к Hsc70 специфически взаимодействуют со своим антигеном в комплексе с меченым фрагментом ДНК, одним из белков комплекса является белок Hsc70.

Для подтверждения полученных выше результатов мы исследовали возможность взаимодействия белков фракции ERLBF с моноклональными антителами к Hsc70 (любезно предоставлены Б.А.Маргулисом, Институт Цитологии РАН, С-Петербург, Россия). Белки фракции ERLBF разделяли электрофорезом в ПААГ (рис. 50а), переносили их на PVDF-мембрану и инкубировали с антителами к Hsc70 (рис. 50б). На дорожке 3 в элюате белков ядерного экстракта Krebs-II с гепарин-агарозной колонки, наблюдается отчетливый сигнал комплекса на уровне контрольного сигнала на дорожке 2, на которую нанесен маркер из трех белков: Hsc (70 kDa), BSA (67 kDa) и GDG (60 kDa). Сигнал соответствует белку Hsc70. Во фракции белков тотального ядерного экстракта Krebs-II (дорожки 4,5) сигнал комплекса обнаруживается в той же зоне. Таким образом, во фракции белков элюата ядерного экстракта с гепарин-агарозы в зоне подвижности ERLBF1 выявлен белок, гомологичный Hsc70 и образующий

комплекс с его антителами. Данные western-блот анализа с моноклональными антителами к Hsc70 подтверждают результаты EMSA ERLBF и идентификации ERLBF1 методом MALDI-TOF-MS. Следовательно, в составе комплекса белков ERLBF1-3 с ДНК обнаруживается белок, который по триптической пептидной карте и антигенным детерминантам идентифицирован как конститутивный белок теплового шока Hsc70.

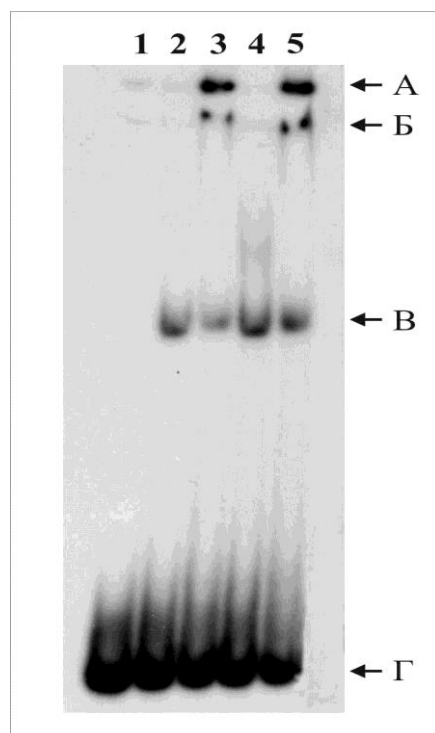


Рис. 49. Эксперимент дополнительного сдвига электрофоретической подвижности в присутствии антител. А, Б, - зоны «торможения» комплексов белок-ДНК-антитело (super-shift); В - зона «торможения» комплекса белок-ДНК; Г - меченый фрагмент ДНК. Дор. 1 - без добавления ядерного экстракта; 2 - 1 мкг ядерного экстракта из клеточной линии Jurkat; 3 – 1 мкг экстракта Jurkat + 15 мкг антител к белку Hsc70; 4 - 1 мкг ядерного экстракта из клеток линии HeLa; 5 – 1 мкг экстракта из клеток линии HeLa + 15 мкг антител к белку Hsc70

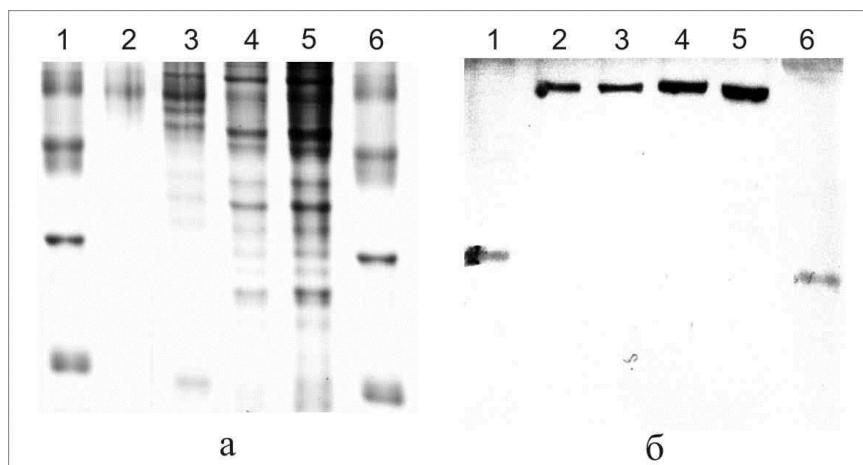


Рис. 50. Western-блот анализ ERLBF с антителами к Hsc70. а - электрофореграмма белков фракции ERLBF в полиакриламидном геле с ддс-натрия; б - реплика с иммобилизованными на PVDF-мембране белками, полученная с полиакриламидного геля (рис. 50а) и инкубированная с моноклональными антителами к Hsc70. В качестве субстрата использовали 4-хлор-1-нафтол. 1,6 - белковые маркеры BioRad; 2 - смесь белков Hsc70 (70 кДа), BSA (67 кДа) и GDG (60 кДа); 3 - ERLBF, элюат с гепарин-агарозы; 4,5 - ядерный экстракт Krebs II (20 и 40 мкг, соотв.).

## ВЫВОДЫ

1. Разработана стратегия идентификации и картирования потенциальных энхансеров в протяженных областях генома. С ее помощью идентифицировано и картировано 15 потенциальных энхансеров в локусе хромосомы 19 человека между генами *FXVD5* и *COX7A1*. Показано, что потенциальные энхансеры специфически взаимодействуют с ядерными белками. 13 из 15 обнаруженных последовательностей проявляют энхансерную активность в системе экспрессии репортерного гена. Таким образом, с помощью разработанной стратегии удается обнаруживать свыше 80% потенциальных энхансерных элементов в протяженных областях генома.

2. С помощью ранее разработанного нами метода позитивно-негативной селекции энхансер-блокирующих элементов генома идентифицировано и картировано 10 новых потенциальных инсуляторов в локусе хромосомы 19 человека между генами *FXVD5* и *COX7A1*. Проведен анализ энхансер-блокирующей активности обнаруженных нами ранее методом двумерного EMSA 10 CTCF-связывающих фрагментов ДНК. Все исследованные CTCF-связывающие последовательности проявляют энхансер-блокирующую активность в данной системе. Возможно, это является общим свойством CTCF-связывающих фрагментов ДНК.

3. Для демонстрации универсальности разработанного нами ранее метода двумерного EMSA выявлены и картированы 77 участков области глобиновых генов кур, способных специфически связываться с фактором транскрипции CTCF. Методом иммунопреципитации хроматина показана повышенная степень связывания трех CTCF-связывающих фрагментов в клетках эритроидного типа по сравнению с лимфоидными клетками.

4. На примере LTR семейства HERV-K (HML-2) исследован потенциальный спектр регуляторных элементов, принадлежащих рассеянному по геному мобильным элементам. Показано наличие промоторной активности LTR

эндогенных ретровирусов человека семейства HERV-K в различных клеточных линиях человека и грызунов и выявлена ее тканеспецифичность. Обнаружено, что находящийся в составе LTR регуляторный элемент, подавляющий активность его промотора, остается активным во всех линиях клеток. Показана также тканеспецифичность энхансерной активности LTR. Таким образом, установлено, что LTR HERV-K обладают уникальным пакетом тканеспецифичных регуляторных элементов. Их обнаружение в области генов должно исследоваться с точки зрения возможного участия в регуляции активности данных генов.

5. Разработан метод выделения ДНК-связывающих белков и с его помощью очищены три ранее неизвестных белка ERLBF1, ERLBF2 и ERLBF3 (Endogenous Retrovirus LTR Binding Factors), образующих специфический комплекс с 5'-областью LTR HERV-K. Один из белков идентифицирован как конститутивно экспрессирующийся белок, сходный с белком теплового шока Hsc70. Полученные результаты показывают, что, по крайней мере, некоторые LTR HERV-K сохранили способность специфического взаимодействия с регуляторными белками клеток, что объясняет их функциональную активность.

## Список литературы

- Abrarova, N.D., Stukacheva, E.A., Pleshkan, V.V., Vinogradova, T.V. and Sverdlov, E.D. Functional analysis of the HERV-K LTR residing in the KIAA1245/NBPF subfamily. *Mol Biol (Mosk)* **44** (2010), pp. 627-34.
- Akopov, S.B., Nikolaev, L.G., Khil, P.P., Lebedev, Y.B. and Sverdlov, E.D. Long terminal repeats of human endogenous retrovirus K family (HERV-K) specifically bind host cell nuclear proteins. *FEBS Lett* **421** (1998), pp. 229-33.
- Akopov, S.B., Nikolaev, L.G., Tyrsin, O., Ruzov, A.S. and Sverdlov, E.D. 14 sequences from Chinese hamster genome preferentially binding to the nuclear matrix. *Bioorg Khim* **23** (1997), pp. 727-31.
- Akopov, S.B., Ruda, V.M., Batrak, V.V., Vetchinova, A.S., Chernov, I.P., Nikolaev, L.G., Bode, J. and Sverdlov, E.D. Identification, genome mapping, and CTCF binding of potential insulators within the FXVD5-COX7A1 locus of human chromosome 19q13.12. *Mamm Genome* **17** (2006), pp. 1042-9.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25** (1997), pp. 3389-402.
- Amouyal, M. Gene insulation. Part II: natural strategies in vertebrates. *Biochem Cell Biol* **88** (2010), pp. 885-98.
- Andersson, G., Svensson, A.C., Setterblad, N. and Rask, L. Retroelements in the human MHC class II region. *Trends Genet* **14** (1998), pp. 109-14.
- Andersson, M.L., Lindeskog, M., Medstrand, P., Westley, B., May, F. and Blomberg, J. Diversity of human endogenous retrovirus class II-like sequences. *J Gen Virol* **80** ( Pt 1) (1999), pp. 255-60.
- Arndt-Jovin, D.J., Jovin, T.M., Bahr, W., Frischauf, A.M. and Marquardt, M. Covalent attachment of DNA to agarose. Improved synthesis and use in affinity chromatography. *Eur J Biochem* **54** (1975), pp. 411-8.
- Arnold, R., Maueler, W., Bassili, G., Lutz, M., Burke, L., Epplen, T.J. and Renkawitz, R. The insulator protein CTCF represses transcription on binding to the (gt)(22)(ga)(15) microsatellite in intron 2 of the HLA-DRB1(\*)0401 gene. *Gene* **253** (2000), pp. 209-14.
- Arnold, C.D., Gerlach, D., Stelzer, C., Boryn, q.M., Rath, M., and Stark, A. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339** (2013), pp. 1074–7.
- Artamonova, II, Gorodentseva, T.N., Lebedev Yu, B. and Sverdlov, E.D. Nonrandom distribution of the endogenous retroviral regulatory elements HERV-K LTR on human chromosome 22. *Dokl Biochem* **372** (2000), pp. 87-9.
- Ashe, H.L., Monks, J., Wijgerde, M., Fraser, P. and Proudfoot, N.J. Intergenic transcription and transinduction of the human beta-globin locus. *Genes Dev* **11** (1997), pp. 2494-509.
- Atchison, M.L. Function of YY1 in Long-Distance DNA Interactions. *Front Immunol* **5** (2014), p. 45.



- Baban, S., Freeman, J.D. and Mager, D.L. Transcripts from a novel human KRAB zinc finger gene contain spliced Alu and endogenous retroviral segments. *Genomics* **33** (1996), pp. 463-72.
- Baer, A., Schubeler, D. and Bode, J. Transcriptional properties of genomic transgene integration sites marked by electroporation or retroviral infection. *Biochemistry* **39** (2000), pp. 7041-9.
- Baltimore, D. RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* **226** (1970), pp. 1209-11.
- Baniahmad, A., Steiner, C., Kohne, A.C. and Renkawitz, R. Modular structure of a chicken lysozyme silencer: involvement of an unusual thyroid hormone receptor binding site. *Cell* **61** (1990), pp. 505-14.
- Bannert, N. and Kurth, R. Retroelements and the human genome: new perspectives on an old relation. *Proc Natl Acad Sci U S A* **101 Suppl 2** (2004), pp. 14572-9.
- Barkess, G. and West, A.G. Chromatin insulator elements: establishing barriers to set heterochromatin boundaries. *Epigenomics* **4** (2012), pp. 67-80.
- Barre-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W. and Montagnier, L. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **220** (1983), pp. 868-71.
- Barrow, J.J., Li, Y., Hossain, M., Huang, S. and Bungert, J. Dissecting the function of the adult beta-globin downstream promoter region using an artificial zinc finger DNA-binding domain. *Nucleic Acids Res* **42** (2014), pp. 4363-74.
- Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I. and Zhao, K. High-resolution profiling of histone methylations in the human genome. *Cell* **129** (2007), pp. 823-37.
- Batlle-Lopez, A., Cortiguera, M.G., Rosa-Garrido, M., Blanco, R., Del Cerro, E., Torrano, V., Wagner, S.D. and Delgado, M.D. Novel CTCF binding at a site in exon1A of BCL6 is associated with active histone marks and a transcriptionally active locus. *Oncogene* (2013).
- Bazzi, H., Fantauzzo, K.A., Richardson, G.D., Jahoda, C.A. and Christiano, A.M. Transcriptional profiling of developing mouse epidermis reveals novel patterns of coordinated gene expression. *Dev Dyn* **236** (2007), pp. 961-70.
- Bell, A.C. and Felsenfeld, G. Stopped at the border: boundaries and insulators. *Curr Opin Genet Dev* **9** (1999), pp. 191-8.
- Bell, A.C. and Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405** (2000), pp. 482-5.
- Bell, A.C., West, A.G. and Felsenfeld, G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98** (1999), pp. 387-96.
- Bell, A.C., West, A.G. and Felsenfeld, G. Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. *Science* **291** (2001), pp. 447-50.
- Belozerov, V.E., Majumder, P., Shen, P. and Cai, H.N. A novel boundary element may facilitate independent gene regulation in the Antennapedia complex of *Drosophila*. *EMBO J* **22** (2003), pp. 3113-21.
- Benabdellah, K., Gutierrez-Guerrero, A., Cobo, M., Munoz, P. and Martin, F. A chimeric HS4-SAR insulator (IS2) that prevents silencing and enhances

- expression of lentiviral vectors in pluripotent stem cells. *PLoS One* **9** (2014), p. e84268.
- Benachou, F., Sperber, G.O., Bongcam-Rudloff, E., Andersson, G., Boeke, J.D. and Blomberg, J. Conserved structure and inferred evolutionary history of long terminal repeats (LTRs). *Mob DNA* **4** (2013), p. 5.
- Bender, M.A., Ragozy, T., Lee, J., Byron, R., Telling, A., Dean, A. and Groudine, M. The hypersensitive sites of the murine beta-globin locus control region act independently to affect nuclear localization and transcriptional elongation. *Blood* **119** (2012), pp. 3820-7.
- Beneke, S. Regulation of chromatin structure by poly(ADP-ribosylation). *Front Genet* **3** (2012), p. 169.
- Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489** (2012), pp. 57-74
- Bi, X. and Broach, J.R. UASrpg can function as a heterochromatin boundary element in yeast. *Genes Dev* **13** (1999), pp. 1089-101.
- Blanton, J., Gaszner, M. and Schedl, P. Protein:protein interactions and the pairing of boundary elements in vivo. *Genes Dev* **17** (2003), pp. 664-75.
- Bogomolova, A.E., Braga, E.A. and Nikolaev, L.G. Protein factors specifically binding to the regulatory elements of non-transcribed spacer of rat ribosomal genes. *Mol Biol (Mosk)* **23** (1989), pp. 1332-9.
- Boller, K., Konig, H., Sauter, M., Mueller-Lantsch, N., Lower, R., Lower, J. and Kurth, R. Evidence that HERV-K is the endogenous retrovirus sequence that codes for the human teratocarcinoma-derived retrovirus HTDV. *Virology* **196** (1993), pp. 349-53.
- Bonner, T.I., O'Connell, C. and Cohen, M. Cloned endogenous retroviral sequences from human DNA. *Proc Natl Acad Sci U S A* **79** (1982), pp. 4709-13.
- Bourque, G., Leong, B., Vega, V.B., Chen, X., Lee, Y.L., Srinivasan, K.G., Chew, J.L., Ruan, Y., Wei, C.L., Ng, H.H. and Liu, E.T. Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res* **18** (2008), pp. 1752-62.
- Bramblett, D., Hsu, C.L., Lozano, M., Earnest, K., Fabritius, C. and Dudley, J. A redundant nuclear protein binding site contributes to negative regulation of the mouse mammary tumor virus long terminal repeat. *J Virol* **69** (1995), pp. 7868-76.
- Brasset, E. and Vaury, C. Insulators are fundamental components of the eukaryotic genomes. *Heredity (Edinb)* **94** (2005), pp. 571-6.
- Brouwer, J.R., Huguet, A., Nicole, A., Munnich, A. and Gourdon, G. Transcriptionally Repressive Chromatin Remodelling and CpG Methylation in the Presence of Expanded CTG-Repeats at the DM1 Locus. *J Nucleic Acids* **2013** (2013), p. 567435.
- Brown, A.J., Goldsworthy, S.M., Barnes, A.A., Eilert, M.M., Tcheang, L., Daniels, D., Muir, A.I., Wigglesworth, M.J., Kinghorn, I., Fraser, N.J., Pike, N.B., Strum, J.C., Steplewski, K.M., Murdock, P.R., Holder, J.C., Marshall, F.H., Szekeres, P.G., Wilson, S., Ignar, D.M., Foord, S.M., Wise, A. and Dowell, S.J. The

- Orphan G protein-coupled receptors GPR41 and GPR43 are activated by propionate and other short chain carboxylic acids. *J Biol Chem* **278** (2003), pp. 11312-9.
- Bulger, M., Schubeler, D., Bender, M.A., Hamilton, J., Farrell, C.M., Hardison, R.C. and Groudine, M. A complex chromatin landscape revealed by patterns of nuclease sensitivity and histone modification within the mouse beta-globin locus. *Mol Cell Biol* **23** (2003), pp. 5234-44.
- Burcin, M., Arnold, R., Lutz, M., Kaiser, B., Runge, D., Lottspeich, F., Filippova, G.N., Lobanenko, V.V., Renkawitz, R. Negative protein 1, which is required for function of the chicken lysozyme gene silencer in conjunction with hormone receptors, is identical to the multivalent zinc finger repressor CTCF *Mol Cell Biol* **17**(1997), pp. 1281-1288
- Burgess-Beusse, B., Farrell, C., Gaszner, M., Litt, M., Mutskov, V., Recillas-Targa, F., Simpson, M., West, A. and Felsenfeld, G. The insulation of genes from external enhancers and silencing chromatin. *Proc Natl Acad Sci U S A* **99 Suppl 4** (2002), pp. 16433-7.
- Burke, L.J., Hollemann, T., Pieler, T. and Renkawitz, R. Molecular cloning and expression of the chromatin insulator protein CTCF in *Xenopus laevis*. *Mech Dev* **113** (2002), pp. 95-8.
- Bushey, A.M., Dorman, E.R. and Corces, V.G. Chromatin insulators: regulatory mechanisms and epigenetic inheritance. *Mol Cell* **32** (2008), pp. 1-9.
- Buzdin, A., Ustyugova, S., Khodosevich, K., Mamedov, I., Lebedev, Y., Hunsmann, G. and Sverdlov, E. Human-specific subfamilies of HERV-K (HML-2) long terminal repeats: three master genes were active simultaneously during branching of hominoid lineages. *Genomics* **81** (2003), pp. 149-56.
- Byun, J., Kim, J.M., Kim, S.H., Yim, J., Robbins, P.D. and Kim, S. A simple and rapid method for the determination of recombinant retrovirus titer by G418 selection. *Gene Ther* **3** (1996), pp. 1018-20.
- Cai, H. and Levine, M. Modulation of enhancer-promoter interactions by insulators in the *Drosophila* embryo. *Nature* **376** (1995), pp. 533-6.
- Cai, H.N. and Levine, M. The gypsy insulator can function as a promoter-specific silencer in the *Drosophila* embryo. *EMBO J* **16** (1997), pp. 1732-41.
- Cai, H.N. and Shen, P. Effects of cis arrangement of chromatin insulators on enhancer-blocking activity. *Science* **291** (2001), pp. 493-5.
- Caiafa, P. and Zlatanova, J. CCCTC-binding factor meets poly(ADP-ribose) polymerase-1. *J Cell Physiol* **219** (2009), pp. 265-70.
- Capelson, M. and Corces, V.G. The ubiquitin ligase dTopors directs the nuclear organization of a chromatin insulator. *Mol Cell* **20** (2005), pp. 105-16.
- Carninci, P. Tagging mammalian transcription complexity. *Trends Genet* **22** (2006), pp. 501-10.
- Casau, A.E., Vaughan, J.E., Lozano, G. and Levine, A.J. Germ cell expression of an isolated human endogenous retroviral long terminal repeat of the HERV-K/HTDV family in transgenic mice. *J Virol* **73** (1999), pp. 9976-83.
- Cavalieri, V., Melfi, R. and Spinelli, G. Promoter activity of the sea urchin (*Paracentrotus lividus*) nucleosomal H3 and H2A and linker H1 {alpha}-histone

- genes is modulated by enhancer and chromatin insulator. *Nucleic Acids Res* **37** (2009), pp. 7407-15.
- Celniker, S.E., Dillon, L.A., Gerstein, M.B., Gunsalus, K.C., Henikoff, S., Karpen, G.H., Kellis, M., Lai, E.C., Lieb, J.D., MacAlpine, D.M., Micklem, G., Piano, F., Snyder, M., Stein, L., White, K.P. and Waterston, R.H. Unlocking the secrets of the genome. *Nature* **459** (2009), pp. 927-30.
- Cena, A., Skoneczny, M., Chelstowska, A., Kowalec, P., Natorff, R. and Kurlandzka, A. Cohesin Irr1/Scc3 is likely to influence transcription in *Saccharomyces cerevisiae* via interaction with Mediator complex. *Acta Biochim Pol* **60** (2013), pp. 233-8.
- Chadwick, B.P. DXZ4 chromatin adopts an opposing conformation to that of the surrounding chromosome and acquires a novel inactive X-specific role involving CTCF and antisense transcripts. *Genome Res* **18** (2008), pp. 1259-69.
- Chai, X., Nagarajan, S., Kim, K., Lee, K. and Choi, J.K. Regulation of the boundaries of accessible chromatin. *PLoS Genet* **9** (2013), p. e1003778.
- Chakraborty, S.A., Simpson, R.T. and Grigoryev, S.A. A single heterochromatin boundary element imposes position-independent antisilencing activity in *Saccharomyces cerevisiae* minichromosomes. *PLoS One* **6** (2011), p. e24835.
- Chalaya, T.V., Akopov, S.B., Nikolaev, L.G. and Sverdlov, E.D. Tissue specificity of methylation of cytosines in regulatory regions of four genes located in the locus FXVD5-COX7A1 of human chromosome 19: correlation with their expression level. *Biochemistry (Mosc)* **71** (2006), pp. 294-9.
- Chao, W., Huynh, K.D., Spencer, R.J., Davidow, L.S. and Lee, J.T. CTCF, a candidate trans-acting factor for X-inactivation choice. *Science* **295** (2002), pp. 345-7.
- Chen, H., Walker, G.E., Taylor, S.I. and McKeon, C. Proximal enhancer of the human insulin receptor gene binds the transcription factor Sp1. *Diabetes* **43** (1994), pp. 884-9.
- Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., Loh, Y.H., Yeo, H.C., Yeo, Z.X., Narang, V., Govindarajan, K.R., Leong, B., Shahab, A., Ruan, Y., Bourque, G., Sung, W.K., Clarke, N.D., Wei, C.L. and Ng, H.H. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133** (2008), pp. 1106-17.
- Cheng, M.C., Wu, S.P., Chen, L.F. and Chen, S.C. Identification and purification of a spinach chloroplast DNA-binding protein that interacts specifically with the plastid psaA-psaB-rps14 promoter region. *Planta* **203** (1997), pp. 373-80.
- Chernov, I.P., Akopov, S.B. and Nikolaev, L.G. Structure and function of nuclear matrix associated regions (S/MARs). *Bioorg Khim* **30** (2004), pp. 3-14.
- Chernukhin, I., Shamsuddin, S., Kang, S.Y., Bergstrom, R., Kwon, Y.W., Yu, W., Whitehead, J., Mukhopadhyay, R., Docquier, F., Farrar, D., Morrison, I., Vigneron, M., Wu, S.Y., Chiang, C.M., Loukinov, D., Lobanekov, V., Ohlsson, R. and Klenova, E. CTCF interacts with and recruits the largest subunit of RNA polymerase II to CTCF target sites genome-wide. *Mol Cell Biol* **27** (2007), pp. 1631-48.
- Chernukhin, I.V., Shamsuddin, S., Robinson, A.F., Carne, A.F., Paul, A., El-Kady, A.I., Lobanekov, V.V. and Klenova, E.M. Physical and functional interaction

- between two pluripotent proteins, the Y-box DNA/RNA-binding factor, YB-1, and the multivalent zinc finger factor, CTCF. *J Biol Chem* **275** (2000), pp. 29915-21.
- Chetverina, D., Savitskaya, E., Maksimenko, O., Melnikova, L., Zaytseva, O., Parshikov, A., Galkin, A.V. and Georgiev, P. Red flag on the white reporter: a versatile insulator abuts the white gene in *Drosophila* and is omnipresent in mini-white constructs. *Nucleic Acids Res* **36** (2008), pp. 929-37.
- Chopra, V.S., Cande, J., Hong, J.W. and Levine, M. Stalled Hox promoters as chromosomal boundaries. *Genes Dev* **23** (2009), pp. 1505-9.
- Christy, R.J. and Huang, R.C. Functional analysis of the long terminal repeats of intracisternal A-particle genes: sequences within the U3 region determine both the efficiency and direction of promoter activity. *Mol Cell Biol* **8** (1988), pp. 1093-102.
- Chung, J.H., Bell, A.C. and Felsenfeld, G. Characterization of the chicken beta-globin insulator. *Proc Natl Acad Sci U S A* **94** (1997), pp. 575-80.
- Chung, J.H., Whiteley, M. and Felsenfeld, G. A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in *Drosophila*. *Cell* **74** (1993), pp. 505-14.
- Cianciolo, G.J., Copeland, T.D., Oroszlan, S. and Snyderman, R. Inhibition of lymphocyte proliferation by a synthetic peptide homologous to retroviral envelope proteins. *Science* **230** (1985), pp. 453-5.
- Cohen, C.J., Lock, W.M. and Mager, D.L. Endogenous retroviral LTRs as promoters for human genes: a critical assessment. *Gene* **448** (2009), pp. 105-14.
- Conrad, B., Weissmahr, R.N., Boni, J., Arcari, R., Schupbach, J. and Mach, B. A human endogenous retroviral superantigen as candidate autoimmune gene in type I diabetes. *Cell* **90** (1997), pp. 303-13.
- Core, L.J. and Lis, J.T. Paused Pol II captures enhancer activity and acts as a potent insulator. *Genes Dev* **23** (2009), pp. 1606-12.
- Costas, J. Evolutionary dynamics of the human endogenous retrovirus family HERV-K inferred from full-length proviral genomes. *J Mol Evol* **53** (2001), pp. 237-43.
- Cuartero, S., Fresan, U., Reina, O., Planet, E. and Espinas, M.L. Ibf1 and Ibf2 are novel CP190-interacting proteins required for insulator function. *EMBO J* **33** (2014), pp. 637-47.
- Cuddapah, S., Jothi, R., Schones, D.E., Roh, T.Y., Cui, K. and Zhao, K. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res* **19** (2009), pp. 24-32.
- Cuvier, O., Hart, C.M. and Laemmli, U.K. Identification of a class of chromatin boundary elements. *Mol Cell Biol* **18** (1998), pp. 7478-86.
- D'Apollito, D., Baiamonte, E., Bagliesi, M., Di Marzo, R., Calzolari, R., Ferro, L., Franco, V., Spinelli, G., Maggio, A. and Acuto, S. The sea urchin *sns5* insulator protects retroviral vectors from chromosomal position effects by maintaining active chromatin structure. *Mol Ther* **17** (2009), pp. 1434-41.
- Darrow, E.M. and Chadwick, B.P. A novel tRNA variable number tandem repeat at human chromosome 1q23.3 is implicated as a boundary element based on

- conservation of a CTCF motif in mouse. *Nucleic Acids Res* **42** (2014), pp. 6421-35.
- de Laat, W., and Duboule, D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* **502** (2013), pp. 499–506.
- De La Rosa-Velazquez, I.A., Rincon-Arano, H., Benitez-Bribiesca, L. and Recillas-Targa, F. Epigenetic regulation of the human retinoblastoma tumor suppressor gene promoter by CTCF. *Cancer Res* **67** (2007), pp. 2577-85.
- de Parseval, N., Alkabbani, H. and Heidmann, T. The long terminal repeats of the HERV-H human endogenous retrovirus contain binding sites for transcriptional regulation by the Myb protein. *J Gen Virol* **80** ( Pt 4) (1999), pp. 841-5.
- de Parseval, N., Lazar, V., Casella, J.F., Benit, L. and Heidmann, T. Survey of human genes of retroviral origin: identification and transcriptome of the genes with coding capacity for complete envelope proteins. *J Virol* **77** (2003), pp. 10414-22.
- Defossez, P.A., Kelly, K.F., Filion, G.J., Perez-Torrado, R., Magdinier, F., Menoni, H., Nordgaard, C.L., Daniel, J.M. and Gilson, E. The human enhancer blocker CTC-binding factor interacts with the transcription factor Kaiso. *J Biol Chem* **280** (2005), pp. 43017-23.
- Deininger, P.L., Batzer, M.A., Hutchison, C.A., 3rd and Edgell, M.H. Master genes in mammalian repetitive DNA amplification. *Trends Genet* **8** (1992), pp. 307-11.
- DeMare, L.E., Leng, J., Cotney, J., Reilly, S.K., Yin, J., Sarro, R. and Noonan, J.P. The genomic landscape of cohesin-associated chromatin interactions. *Genome Res* **23** (2013), pp. 1224-34.
- Deng, Z., Wang, Z., Stong, N., Plasschaert, R., Moczan, A., Chen, H.S., Hu, S., Wikramasinghe, P., Davuluri, R.V., Bartolomei, M.S., Riethman, H. and Lieberman, P.M. A role for CTCF and cohesin in subtelomere chromatin organization, TERRA transcription, and telomere end protection. *EMBO J* **31** (2012), pp. 4165-78.
- Di Cristofano, A., Strazzullo, M., Longo, L. and La Mantia, G. Characterization and genomic mapping of the ZNF80 locus: expression of this zinc-finger gene is driven by a solitary LTR of ERV9 endogenous retroviral family. *Nucleic Acids Res* **23** (1995a), pp. 2823-30.
- Di Cristofano, A., Strazzullo, M., Parisi, T. and La Mantia, G. Mobilization of an ERV9 human endogenous retroviral element during primate evolution. *Virology* **213** (1995b), pp. 271-5.
- Di Simone, P., Di Leonardo, A., Costanzo, G., Melfi, R. and Spinelli, G. The sea urchin insulator blocks CMV enhancer following integration in human cells. *Biochem Biophys Res Commun* **284** (2001), pp. 987-92.
- Didych, D.A., Akopov, S.B., Snezhkov, E.V., Skaptsova, N.V., Nikolaev, L.G. and Sverdlov, E.D. Identification and mapping of ten new potential insulators in the FXVD5-COX7A1 region of human chromosome 19q13.12. *Biochemistry (Mosc)* **74** (2009), pp. 728-33.
- Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S. and Ren, B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485** (2012), pp. 376-80.

- Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Röder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Bar NS, Batut P, Bell K, Bell I, Chakraborty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Falconnet E, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena H, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Luo OJ, Park E, Persaud K, Preall JB, Ribeca P, Risk B, Robyr D, Sammeth M, Schaffer L, See LH, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Wrobel J, Yu Y, Ruan X, Hayashizaki Y, Harrow J, Gerstein M, Hubbard T, Reymond A, Antonarakis SE, Hannon G, Giddings MC, Ruan Y, Wold B, Carninci P, Guigó R, Gingeras TR. Landscape of transcription in human cells. *Nature* **489** (2012), pp. 101-8.
- Docquier, F., Farrar, D., D'Arcy, V., Chernukhin, I., Robinson, A.F., Loukinov, D., Vatolin, S., Pack, S., Mackay, A., Harris, R.A., Dorricott, H., O'Hare, M.J., Lobanenko, V. and Klenova, E. Heightened expression of CTCF in breast cancer cells is associated with resistance to apoptosis. *Cancer Res* **65** (2005), pp. 5112-22.
- Domanskii, A.N., Akopov, S.B., Lebedev Iu, B., Nikolaev, L.G. and Sverdlov, E.D. Enhancer activity of solitary long terminal repeat of the human endogenous retrovirus of the HERV-K family. *Bioorg Khim* **28** (2002), pp. 341-5.
- Domansky, A.N., Kopantzev, E.P., Snezhkov, E.V., Lebedev, Y.B., Leib-Mosch, C. and Sverdlov, E.D. Solitary HERV-K LTRs possess bi-directional promoter activity and contain a negative regulatory element in the U5 region. *FEBS Lett* **472** (2000), pp. 191-5.
- Donohoe, M.E., Zhang, L.F., Xu, N., Shi, Y. and Lee, J.T. Identification of a Ctf cofactor, Yy1, for the X chromosome binary switch. *Mol Cell* **25** (2007), pp. 43-56.
- Dorsett, D. Distance-independent inactivation of an enhancer by the suppressor of Hairy-wing DNA-binding protein of Drosophila. *Genetics* **134** (1993), pp. 1135-44.
- Eisfeld, K., Candau, R., Truss, M. and Beato, M. Binding of NF1 to the MMTV promoter in nucleosomes: influence of rotational phasing, translational positioning and histone H1. *Nucleic Acids Res* **25** (1997), pp. 3733-42.
- Elkahloun, A.G., Bittner, M., Hoskins, K., Gemmill, R. and Meltzer, P.S. Molecular cytogenetic characterization and physical mapping of 12q13-15 amplification in human cancers. *Genes Chromosomes Cancer* **17** (1996), pp. 205-14.
- Engel, N., Raval, A.K., Thorvaldsen, J.L. and Bartolomei, S.M. Three-dimensional conformation at the H19/Igf2 locus supports a model of enhancer tracking. *Hum Mol Genet* **17** (2008), pp. 3021-9.
- Ernst, J. and Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol* **28** (2010), pp. 817-25.

- Ernst, J., Kheradpour, P., Mikkelson, T.S., Shores, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., Ku, M., Durham, T., Kellis, M. and Bernstein, B.E. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473** (2011), pp. 43-9.
- Erokhin, M., Davydova, A., Kyrchanova, O., Parshikov, A., Georgiev, P. and Chetverina, D. Insulators form gene loops by interacting with promoters in *Drosophila*. *Development* **138** (2011), pp. 4097-106.
- Fahlbusch, F.B., Ruebner, M., Volkert, G., Offergeld, R., Hartner, A., Menendez-Castro, C., Strick, R., Rauh, M., Rascher, W. and Dotsch, J. Corticotropin-releasing hormone stimulates expression of leptin, 11beta-HSD2 and syncytin-1 in primary human trophoblasts. *Reprod Biol Endocrinol* **10** (2012), p. 80.
- Fang, Y., Gao, J., Qi, L. and Li, N. CTCF-regulating endocrine function of pancreatic islet cells in transgenic mice. *Horm Metab Res* **46** (2014), pp. 419-23.
- Farrell, C.M., West, A.G. and Felsenfeld, G. Conserved CTCF insulator elements flank the mouse and human beta-globin loci. *Mol Cell Biol* **22** (2002), pp. 3820-31.
- Fedoriw, A.M., Stein, P., Svoboda, P., Schultz, R.M. and Bartolomei, M.S. Transgenic RNAi reveals essential function for CTCF in H19 gene imprinting. *Science* **303** (2004), pp. 238-40.
- Fei, C., Atterby, C., Edqvist, P.H., Ponten, F., Zhang, W.W., Larsson, E. and Ryan, F.P. Detection of the human endogenous retrovirus ERV3-encoded Env-protein in human tissues using antibody-based proteomics. *J R Soc Med* **107** (2013), pp. 22-9.
- Feuchter, A. and Mager, D. Functional heterogeneity of a large family of human LTR-like promoters and enhancers. *Nucleic Acids Res* **18** (1990), pp. 1261-70.
- Filippova, G.N. Genetics and epigenetics of the multifunctional protein CTCF. *Curr Top Dev Biol* **80** (2008), pp. 337-60.
- Filippova, G.N., Cheng, M.K., Moore, J.M., Truong, J.P., Hu, Y.J., Nguyen, D.K., Tsuchiya, K.D. and Distech, C.M. Boundaries between chromosomal domains of X inactivation and escape bind CTCF and lack CpG methylation during early development. *Dev Cell* **8** (2005), pp. 31-42.
- Filippova, G.N., Fagerlie, S., Klenova, E.M., Myers, C., Dehner, Y., Goodwin, G., Neiman, P.E., Collins, S.J. and Lobanekov, V.V. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol Cell Biol* **16** (1996), pp. 2802-13.
- Filippova, G.N., Lindblom, A., Meincke, L.J., Klenova, E.M., Neiman, P.E., Collins, S.J., Doggett, N.A. and Lobanekov, V.V. A widely expressed transcription factor with multiple DNA sequence specificity, CTCF, is localized at chromosome segment 16q22.1 within one of the smallest regions of overlap for common deletions in breast and prostate cancers. *Genes Chromosomes Cancer* **22** (1998), pp. 26-36.
- Filippova, G.N., Qi, C.F., Ulmer, J.E., Moore, J.M., Ward, M.D., Hu, Y.J., Loukinov, D.I., Pugacheva, E.M., Klenova, E.M., Grundy, P.E., Feinberg, A.P., Cleton-Jansen, A.M., Moerland, E.W., Cornelisse, C.J., Suzuki, H., Komiyama, A., Lindblom, A., Dorion-Bonnet, F., Neiman, P.E., Morse, H.C., 3rd, Collins, S.J.



- and Lobanenkov, V.V. Tumor-associated zinc finger mutations in the CTCF transcription factor selectively alter its DNA-binding specificity. *Cancer Res* **62** (2002), pp. 48-52.
- Finnegan, D.J. Retrotransposons. *Curr Biol* **22** (2012), pp. R432-7.
- Fiorentino, F.P. and Giordano, A. The tumor suppressor role of CTCF. *J Cell Physiol* **227** (2011), pp. 479-92.
- Fong, S.E., Pallansch, L.A., Mikovits, J.A., Lackman-Smith, C.S., Ruscetti, F.W. and Gonda, M.A. cis-acting regulatory elements in the bovine immunodeficiency virus long terminal repeat. *Virology* **209** (1995), pp. 604-14.
- Friedli, M., Turelli, P., Kapopoulou, A., Rauwel, B., Castro-Diaz, N., Rowe, H.M., Ecco, G., Unzu, C., Planet, E., Lombardo, A., Mangeat, B., Wildhaber, B.E., Naldini, L. and Trono, D. Loss of transcriptional control over endogenous retroelements during reprogramming to pluripotency. *Genome Res* (2014).
- Fu, Y., Sinha, M., Peterson, C.L. and Weng, Z. The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet* **4** (2008), p. e1000138.
- Fuchs, N.V., Kraft, M., Tondera, C., Hanschmann, K.M., Lower, J. and Lower, R. Expression of the human endogenous retrovirus (HERV) group HML-2/HERV-K does not depend on canonical promoter elements but is regulated by transcription factors Sp1 and Sp3. *J Virol* **85** (2011), pp. 3436-48.
- Fujioka, M., Sun, G. and Jaynes, J.B. The Drosophila eve insulator Homie promotes eve expression and protects the adjacent gene from repression by polycomb spreading. *PLoS Genet* **9** (2013), p. e1003883.
- Furlan-Magaril, M., Rebollar, E., Guerrero, G., Fernandez, A., Molto, E., Gonzalez-Buendia, E., Cantero, M., Montoliu, L. and Recillas-Targa, F. An insulator embedded in the chicken alpha-globin locus regulates chromatin domain configuration and differential gene expression. *Nucleic Acids Res* **39** (2010), pp. 89-103.
- Gadgil, H. and Jarrett, H.W. Heparin elution of transcription factors from DNA-Sepharose columns. *J Chromatogr A* **848** (1999), pp. 131-8.
- Gallagher, P.G., Steiner, L.A., Liem, R.I., Owen, A.N., Cline, A.P., Seidel, N.E., Garrett, L.J. and Bodine, D.M. Mutation of a barrier insulator in the human ankyrin-1 gene is associated with hereditary spherocytosis. *J Clin Invest* **120** (2010), pp. 4453-65.
- Gaszner, M. and Felsenfeld, G. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet* **7** (2006), pp. 703-13.
- Gaszner, M., Vazquez, J. and Schedl, P. The Zw5 protein, a component of the scs chromatin domain boundary, is able to block enhancer-promoter interaction. *Genes Dev* **13** (1999), pp. 2098-107.
- Gaussin, A., Modlich, U., Bauche, C., Niederlander, N.J., Schambach, A., Duros, C., Artus, A., Baum, C., Cohen-Haguenaer, O. and Mermoud, N. CTF/NF1 transcription factors act as potent genetic insulators for integrating gene transfer vectors. *Gene Ther* **19** (2011), pp. 15-24.

- Gdula, D.A., Gerasimova, T.I. and Corces, V.G. Genetic and molecular analysis of the gypsy chromatin insulator of *Drosophila*. *Proc Natl Acad Sci U S A* **93** (1996), pp. 9378-83.
- Genersch, E., Eckerskorn, C., Lottspeich, F., Herzog, C., Kuhn, K. and Poschl, E. Purification of the sequence-specific transcription factor CTCBF, involved in the control of human collagen IV genes: subunits with homology to Ku antigen. *EMBO J* **14** (1995), pp. 791-800.
- Gerasimova, T.I., Byrd, K. and Corces, V.G. A chromatin insulator determines the nuclear localization of DNA. *Mol Cell* **6** (2000), pp. 1025-35.
- Gerasimova, T.I., Lei, E.P., Bushey, A.M. and Corces, V.G. Coordinated control of dCTCF and gypsy chromatin insulators in *Drosophila*. *Mol Cell* **28** (2007), pp. 761-72.
- Geyer, P.K. and Corces, V.G. DNA position-specific repression of transcription by a *Drosophila* zinc finger protein. *Genes Dev* **6** (1992), pp. 1865-73.
- Geyer, P.K., Green, M.M. and Corces, V.G. Tissue-specific transcriptional enhancers may act in trans on the gene located in the homologous chromosome: the molecular basis of transvection in *Drosophila*. *EMBO J* **9** (1990), pp. 2247-56.
- Geyer, P.K., Spana, C. and Corces, V.G. On the molecular mechanism of gypsy-induced mutations at the yellow locus of *Drosophila melanogaster*. *EMBO J* **5** (1986), pp. 2657-62.
- Ghirlando, R., Giles, K., Gowher, H., Xiao, T., Xu, Z., Yao, H. and Felsenfeld, G. Chromatin domains, insulators, and the regulation of gene expression. *Biochim Biophys Acta* **1819** (2012), pp. 644-51.
- Ghosh, D., Gerasimova, T.I. and Corces, V.G. Interactions between the Su(Hw) and Mod(mdg4) proteins required for gypsy insulator function. *EMBO J* **20** (2001), pp. 2518-27.
- Glazkov, M.V. Loop-domain organization of genes in eukaryotic chromosomes. *Mol Biol (Mosk)* **29** (1995), pp. 965-82.
- Gogvadze, E. and Buzdin, A. Retroelements and their impact on genome evolution and functioning. *Cell Mol Life Sci* **66** (2009), pp. 3727-42.
- Golovnin, A.K., Dvoretzky, E.V., Kostyuchenko, M.V., Shamsutdinov, M.F., Georgiev, P.G. and Melnikova, L.S. MOD(MDG4)-64.2 protein, isoform of MOD(MDG4) loci, directly interacts with the Tweedle protein family of *Drosophila melanogaster*. *Dokl Biochem Biophys* **452** (2013), pp. 225-8.
- Gomes, N.P. and Espinosa, J.M. Gene-specific repression of the p53 target gene PUMA via intragenic CTCF-Cohesin binding. *Genes Dev* **24** (2010), pp. 1022-34.
- Gonzalez-Hernandez, M.J., Swanson, M.D., Contreras-Galindo, R., Cookinham, S., King, S.R., Noel, R.J., Jr., Kaplan, M.H. and Markovitz, D.M. Expression of human endogenous retrovirus type K (HML-2) is activated by the Tat protein of HIV-1. *J Virol* **86** (2012), pp. 7790-805.
- Grandchamp, N., Henriot, D., Philippe, S., Amar, L., Ursulet, S., Serguera, C., Mallet, J. and Sarkis, C. Influence of insulators on transgene expression from integrating and non-integrating lentiviral vectors. *Genet Vaccines Ther* **9** (2011), p. 1.

- Gray, C.E. and Coates, C.J. Cloning and characterization of cDNAs encoding putative CTCFs in the mosquitoes, *Aedes aegypti* and *Anopheles gambiae*. *BMC Mol Biol* **6** (2005), p. 16.
- Grdisa, M. and White, M.K. Erythrocytic differentiation and glyceraldehyde-3-phosphate dehydrogenase expression are regulated by protein phosphorylation and cAMP in HD3 cells. *Int J Biochem Cell Biol* **32** (2000), pp. 589-95.
- Green, A.R., Krivinskas, S., Young, P., Rakha, E.A., Paish, E.C., Powe, D.G. and Ellis, I.O. Loss of expression of chromosome 16q genes DPEP1 and CTCF in lobular carcinoma in situ of the breast. *Breast Cancer Res Treat* **113** (2009), pp. 59-66.
- Grosveld, F., Antoniou, M., van Assendelft, G.B., de Boer, E., Hurst, J., Kollias, G., MacFarlane, F. and Wrighton, N. The regulation of expression of human beta-globin genes. *Prog Clin Biol Res* **251** (1987), pp. 133-44.
- Gruzdeva, N., Kyrchanova, O., Parshikov, A., Kullyev, A. and Georgiev, P. The Mcp element from the bithorax complex contains an insulator that is capable of pairwise interactions and can facilitate enhancer-promoter communication. *Mol Cell Biol* **25** (2005), pp. 3682-9.
- Gualtieri, A., Andreola, F., Sciamanna, I., Sinibaldi-Vallebona, P., Serafino, A. and Spadafora, C. Increased expression and copy number amplification of LINE-1 and SINE B1 retrotransposable elements in murine mammary carcinoma progression. *Oncotarget* **4** (2013), pp. 1882-93.
- Guastafierro, T., Cecchinelli, B., Zampieri, M., Reale, A., Riggio, G., Sthandier, O., Zupi, G., Calabrese, L. and Caiafa, P. CCCTC-binding factor activates PARP-1 affecting DNA methylation machinery. *J Biol Chem* **283** (2008), pp. 21873-80.
- Guerrero, G., Delgado-Olguin, P., Escamilla-Del-Arenal, M., Furlan-Magaril, M., Rebollar, E., De La Rosa-Velazquez, I.A., Soto-Reyes, E., Rincon-Arango, H., Valdes-Quezada, C., Valadez-Graham, V. and Recillas-Targa, F. Globin genes transcriptional switching, chromatin structure and linked lessons to epigenetics in cancer: a comparative overview. *Comp Biochem Physiol A Mol Integr Physiol* **147** (2007), pp. 750-60.
- Gurskaya, N.G., Diatchenko, L., Chenchik, A., Siebert, P.D., Khaspekov, G.L., Lukyanov, K.A., Vagner, L.L., Ermolaeva, O.D., Lukyanov, S.A. and Sverdlov, E.D. Equalizing cDNA subtraction based on selective suppression of polymerase chain reaction: cloning of Jurkat cell transcripts induced by phytohemagglutinin and phorbol 12-myristate 13-acetate. *Anal Biochem* **240** (1996), pp. 90-7.
- Gurudatta, B.V., Ramos, E. and Corces, V.G. The BEAF insulator regulates genes involved in cell polarity and neoplastic growth. *Dev Biol* **369** (2012), pp. 124-32.
- Hancks, D.C. and Kazazian, H.H., Jr. Active human retrotransposons: variation and disease. *Curr Opin Genet Dev* **22** (2012), pp. 191-203.
- Hanke, K., Chudak, C., Kurth, R. and Bannert, N. The Rec protein of HERV-K(HML-2) upregulates androgen receptor activity by binding to the human small glutamine-rich tetratricopeptide repeat protein (hSGT). *Int J Cancer* **132** (2012), pp. 556-67.
- Harju, S., McQueen, K.J. and Peterson, K.R. Chromatin structure and control of beta-like globin gene switching. *Exp Biol Med (Maywood)* **227** (2002), pp. 683-700.

- Hark, A.T., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorse, J.M. and Tilghman, S.M. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* **405** (2000), pp. 486-9.
- Harris, J.R. Placental endogenous retrovirus (ERV): structural, functional, and evolutionary significance. *Bioessays* **20** (1998), pp. 307-16.
- Hart, C.M., Zhao, K. and Laemmli, U.K. The scs' boundary element: characterization of boundary element-associated factors. *Mol Cell Biol* **17** (1997), pp. 999-1009.
- Hayes, M., Whitesell, M. and Brown, M.A. Pathological and evolutionary implications of retroviruses as mobile genetic elements. *Genes (Basel)* **4** (2013), pp. 573-82.
- Heath, H., Ribeiro de Almeida, C., Sleutels, F., Dingjan, G., van de Nobelen, S., Jonkers, I., Ling, K.W., Gribnau, J., Renkawitz, R., Grosveld, F., Hendriks, R.W. and Galjart, N. CTCF regulates cell cycle progression of alphabeta T cells in the thymus. *EMBO J* **27** (2008), pp. 2839-50.
- Heng, H.H., Krawetz, S.A., Lu, W., Bremer, S., Liu, G. and Ye, C.J. Re-defining the chromatin loop domain. *Cytogenet Cell Genet* **93** (2001), pp. 155-61.
- Hens, K., Feuz, J.D., Isakova, A., Iagovitina, A., Massouras, A., Bryois, J., Callaerts, P., Celniker, S.E. and Deplancke, B. Automated protein-DNA interaction screening of Drosophila regulatory elements. *Nat Methods* **8** (2011), pp. 1065-70.
- Herbst, H., Kuhler-Obbarius, C., Lauke, H., Sauter, M., Mueller-Lantzsch, N., Harms, D. and Loning, T. Human endogenous retrovirus (HERV)-K transcripts in gonadoblastomas and gonadoblastoma-derived germ cell tumours. *Virchows Arch* **434** (1999), pp. 11-5.
- Herbst, H., Sauter, M., Kuhler-Obbarius, C., Loning, T. and Mueller-Lantzsch, N. Human endogenous retrovirus (HERV)-K transcripts in germ cell and trophoblastic tumours. *APMIS* **106** (1998), pp. 216-20.
- Herold, M., Bartkuhn, M. and Renkawitz, R. CTCF: insights into insulator function during development. *Development* **139** (2012), pp. 1045-57.
- Herrmann, M., Selige, J., Raffael, S., Sachs, G., Brambilla, A. and Klein, T. Systematic expression profiling of the gastric H<sup>+</sup>/K<sup>+</sup> ATPase in human tissue. *Scand J Gastroenterol* **42** (2007), pp. 1275-88.
- Ho, Y., Tadevosyan, A., Liebhaber, S.A. and Cooke, N.E. The juxtaposition of a promoter with a locus control region transcriptional domain activates gene expression. *EMBO Rep* **9** (2008), pp. 891-8.
- Hohenadl, C., Leib-Mosch, C., Hehlmann, R. and Erfle, V. Biological significance of human endogenous retroviral sequences. *J Acquir Immune Defic Syndr Hum Retrovirol* **13 Suppl 1** (1996), pp. S268-73.
- Hohn, O., Hanke, K. and Bannert, N. HERV-K(HML-2), the Best Preserved Family of HERVs: Endogenization, Expression, and Implications in Health and Disease. *Front Oncol* **3** (2013), p. 246.
- Holwerda, S.J. and de Laat, W. CTCF: the protein, the binding partners, the binding sites and their chromatin loops. *Philos Trans R Soc Lond B Biol Sci* **368** (2013), p. 20120369.
- Hore, T.A., Deakin, J.E. and Marshall Graves, J.A. The evolution of epigenetic regulators CTCF and BORIS/CTCFL in amniotes. *PLoS Genet* **4** (2008), p. e1000169.

- Horie, M., Honda, T., Suzuki, Y., Kobayashi, Y., Daito, T., Oshida, T., Ikuta, K., Jern, P., Gojobori, T., Coffin, J.M. and Tomonaga, K. Endogenous non-retroviral RNA virus elements in mammalian genomes. *Nature* **463** (2010), pp. 84-7.
- Hou, C., Zhao, H., Tanimoto, K. and Dean, A. CTCF-dependent enhancer-blocking by alternative chromatin loop formation. *Proc Natl Acad Sci U S A* **105** (2008), pp. 20398-403.
- Hsu, T.W., Sabran, J.L., Mark, G.E., Guntaka, R.V. and Taylor, J.M. Analysis of unintegrated avian RNA tumor virus double-stranded DNA intermediates. *J Virol* **28** (1978), pp. 810-8.
- Hughes, J.F. and Coffin, J.M. Human endogenous retrovirus K solo-LTR formation and insertional polymorphisms: implications for human and viral evolution. *Proc Natl Acad Sci U S A* **101** (2004), pp. 1668-72.
- Illarionova, A.E., Vinogradova, T.V. and Sverdlov, E.D. Only those genes of the KIAA1245 gene subfamily that contain HERV(K) LTRs in their introns are transcriptionally active. *Virology* **358** (2007), pp. 39-47.
- Indraccolo, S., Gunzburg, W.H., Leib-Mosch, C., Erfle, V. and Salmons, B. Identification of three human sequences with viral superantigen-specific primers. *Mamm Genome* **6** (1995), pp. 339-44.
- Ishihara, K., Oshimura, M. and Nakao, M. CTCF-dependent chromatin insulator is linked to epigenetic remodeling. *Mol Cell* **23** (2006), pp. 733-42.
- Kadonaga, J.T. and Tjian, R. Affinity purification of sequence-specific DNA binding proteins. *Proc Natl Acad Sci U S A* **83** (1986), pp. 5889-93.
- Kanduri, M., Kanduri, C., Mariano, P., Vostrov, A.A., Quitschke, W., Lobanenko, V. and Ohlsson, R. Multiple nucleosome positioning sites regulate the CTCF-mediated insulator function of the H19 imprinting control region. *Mol Cell Biol* **22** (2002), pp. 3339-44.
- Kang, H. and Lieberman, P.M. Mechanism of glycyrrhizic acid inhibition of Kaposi's sarcoma-associated herpesvirus: disruption of CTCF-cohesin-mediated RNA polymerase II pausing and sister chromatid cohesion. *J Virol* **85** (2011), pp. 11159-69.
- Kapitonov, V.V. and Jurka, J. The long terminal repeat of an endogenous retrovirus induces alternative splicing and encodes an additional carboxy-terminal sequence in the human leptin receptor. *J Mol Evol* **48** (1999), pp. 248-51.
- Kato, N., Pfeifer-Ohlsson, S., Kato, M., Larsson, E., Rydnert, J., Ohlsson, R. and Cohen, M. Tissue-specific expression of human provirus ERV3 mRNA in human placenta: two of the three ERV3 mRNAs contain human cellular sequences. *J Virol* **61** (1987), pp. 2182-91.
- Kato, N., Shimotohno, K., VanLeeuwen, D. and Cohen, M. Human proviral mRNAs down regulated in choriocarcinoma encode a zinc finger protein related to Kruppel. *Mol Cell Biol* **10** (1990), pp. 4401-5.
- Katoh, I. Impacts of endogenous retroviruses on tumorigenesis, immunity, stem cells, and research safety. *Front Oncol* **4** (2014), p. 66.
- Katoh, I. and Kurata, S. Association of endogenous retroviruses and long terminal repeats with human disorders. *Front Oncol* **3** (2013), p. 234.

- Kejnovsky, E. and Lexa, M. Quadruplex-forming DNA sequences spread by retrotransposons may serve as genome regulators. *Mob Genet Elements* **4** (2014), p. e28084.
- Kellum, R. and Schedl, P. A position-effect assay for boundaries of higher order chromosomal domains. *Cell* **64** (1991), pp. 941-50.
- Kellum, R. and Schedl, P. A group of scs elements function as domain boundaries in an enhancer-blocking assay. *Mol Cell Biol* **12** (1992), pp. 2424-31.
- Kemp, C.J., Moore, J.M., Moser, R., Bernard, B., Teater, M., Smith, L.E., Rabaia, N.A., Gurley, K.E., Guinney, J., Busch, S.E., Shaknovich, R., Lobanekov, V.V., Liggitt, D., Shmulevich, I., Melnick, A. and Filippova, G.N. CTCF haploinsufficiency destabilizes DNA methylation and predisposes to cancer. *Cell Rep* **7** (2014), pp. 1020-9.
- Kent, W.J. BLAT--the BLAST-like alignment tool. *Genome Res* **12** (2002), pp. 656-64.
- Kessler, A.F., Wiesner, M., Denner, J., Kammerer, U., Vince, G.H., Linsenmann, T., Lohr, M., Ernestus, R.I. and Hagemann, C. Expression-analysis of the human endogenous retrovirus HERV-K in human astrocytic tumors. *BMC Res Notes* **7** (2014), p. 159.
- Khil, P.P., Kostina, M.B., Azhikina, T.L., Kolesnik, T.B., Lebedev Iu, B. and Sverdlov, E.D. Structural characteristics of four long terminal repeats (LTR) of human endogenous retroviruses and features of their integration sites. *Bioorg Khim* **23** (1997a), pp. 434-40.
- Khil, P.P., Lebedev Iu, B. and Sverdlov, E.D. Subfamilies of long terminal repeats (LTR) of human endogenous retroviruses of the HERV-K type. *Dokl Akad Nauk* **356** (1997b), pp. 833-7.
- Khil, P.P., Lebedev Iu, B. and Sverdlov, E.D. New putative gene from human retrovirus-containing locus on chromosome 19. *Bioorg Khim* **24** (1998), pp. 72-4.
- Kim, J. Multiple YY1 and CTCF binding sites in imprinting control regions. *Epigenetics* **3** (2008), pp. 115-8.
- Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanekov, V.V. and Ren, B. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128** (2007), pp. 1231-45.
- Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D. and Ren, B. A high-resolution map of active promoters in the human genome. *Nature* **436** (2005), pp. 876-80.
- Kim, T.W., Wu, K., Xu, J.L., McAuliffe, G., Tanzi, R.E., Wasco, W. and Black, I.B. Selective localization of amyloid precursor-like protein 1 in the cerebral cortex postsynaptic density. *Brain Res Mol Brain Res* **32** (1995), pp. 36-44.
- Kingsman, S.M. and Kingsman, A.J. The regulation of human immunodeficiency virus type-1 gene expression. *Eur J Biochem* **240** (1996), pp. 491-507.
- Kinsella, T.M. and Nolan, G.P. Episomal vectors rapidly and stably produce high-titer recombinant retrovirus. *Hum Gene Ther* **7** (1996), pp. 1405-13.
- Kirkland, J.G., Raab, J.R. and Kamakaka, R.T. TFIIC bound DNA elements in nuclear organization and insulation. *Biochim Biophys Acta* **1829** (2012), pp. 418-24.

- Kitchen, N.S. and Schoenherr, C.J. Sumoylation modulates a domain in CTCF that activates transcription and decondenses chromatin. *J Cell Biochem* **111** (2010), pp. 665-75.
- Klenova, E.M., Chernukhin, I.V., El-Kady, A., Lee, R.E., Pugacheva, E.M., Loukinov, D.I., Goodwin, G.H., Delgado, D., Filippova, G.N., Leon, J., Morse, H.C., 3rd, Neiman, P.E. and Lobanekov, V.V. Functional phosphorylation sites in the C-terminal region of the multivalent multifunctional transcriptional factor CTCF. *Mol Cell Biol* **21** (2001), pp. 2221-34.
- Klenova, E.M., Fagerlie, S., Filippova, G.N., Kretzner, L., Goodwin, G.H., Loring, G., Neiman, P.E. and Lobanekov, V.V. Characterization of the chicken CTCF genomic locus, and initial study of the cell cycle-regulated promoter of the gene. *J Biol Chem* **273** (1998), pp. 26571-9.
- Klenova, E.M., Nicolas, R.H., U, S., Carne, A.F., Lee, R.E., Lobanekov, V.V. and Goodwin, G.H. Molecular weight abnormalities of the CTCF transcription factor: CTCF migrates aberrantly in SDS-PAGE and the size of the expressed protein is affected by the UTRs and sequences within the coding region of the CTCF gene. *Nucleic Acids Res* **25** (1997), pp. 466-74.
- Klochkov, D., Rincon-Arango, H., Ioudinkova, E.S., Valadez-Graham, V., Gavrilov, A., Recillas-Targa, F. and Razin, S.V. A CTCF-dependent silencer located in the differentially methylated area may regulate expression of a housekeeping gene overlapping a tissue-specific gene domain. *Mol Cell Biol* **26** (2006), pp. 1589-97.
- Knossl, M., Lower, R. and Lower, J. Expression of the human endogenous retrovirus HTDV/HERV-K is enhanced by cellular transcription factor YY1. *J Virol* **73** (1999), pp. 1254-61.
- Kolovos, P., Knoch, T.A., Grosveld, F.G., Cook, P.R. and Papantonis, A. Enhancers and silencers: an integrated and simple model for their function. *Epigenetics Chromatin* **5** (2012), p. 1.
- Kong, S., Bohl, D., Li, C. and Tuan, D. Transcription of the HS2 enhancer toward a cis-linked gene is independent of the orientation, position, and distance of the enhancer relative to the gene. *Mol Cell Biol* **17** (1997), pp. 3955-65.
- Kongsuwan, K., Allen, J. and Adams, J.M. Expression of Hox-2.4 homeobox gene directed by proviral insertion in a myeloid leukemia. *Nucleic Acids Res* **17** (1989), pp. 1881-92.
- Kowalski, P.E. and Mager, D.L. A human endogenous retrovirus suppresses translation of an associated fusion transcript, PLA2L. *J Virol* **72** (1998), pp. 6164-8.
- Kraus, B., Boller, K., Reuter, A. and Schnierle, B.S. Characterization of the human endogenous retrovirus K Gag protein: identification of protease cleavage sites. *Retrovirology* **8** (2011), p. 21.
- Kuhn, E.J. and Geyer, P.K. Genomic insulators: connecting properties to mechanism. *Curr Opin Cell Biol* **15** (2003), pp. 259-65.
- Kuhn, E.J., Viering, M.M., Rhodes, K.M. and Geyer, P.K. A test of insulator interactions in *Drosophila*. *EMBO J* **22** (2003), pp. 2463-71.
- Kuhn, R.M., Karolchik, D., Zweig, A.S., Trumbower, H., Thomas, D.J., Thakkapallayil, A., Sugnet, C.W., Stanke, M., Smith, K.E., Siepel, A., Rosenbloom, K.R., Rhead, B., Raney, B.J., Pohl, A., Pedersen, J.S., Hsu, F., Hinrichs, A.S., Harte, R.A.,

- Diekhans, M., Clawson, H., Bejerano, G., Barber, G.P., Baertsch, R., Haussler, D. and Kent, W.J. The UCSC genome browser database: update 2007. *Nucleic Acids Res* **35** (2007), pp. D668-73.
- Kumar, P., Tripathi, S. and Pandey, K.N. Histone deacetylase inhibitors modulate the transcriptional regulation of guanylyl cyclase/natriuretic peptide receptor-a gene: interactive roles of modified histones, histone acetyltransferase, p300, AND Sp1. *J Biol Chem* **289** (2014), pp. 6991-7002.
- Kurth, R., Lower, J., Lower, R., Boller, K. and Frank, H. Retroviruses in human tumors. *Arch Geschwulstforsch* **53** (1983), pp. 289-99.
- Kurukuti, S., Tiwari, V.K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., Lobanenko, V., Reik, W. and Ohlsson, R. CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc Natl Acad Sci U S A* **103** (2006), pp. 10684-9.
- Kyrchanova, O., Toshchakov, S., Parshikov, A. and Georgiev, P. Study of the functional interaction between Mcp insulators from the Drosophila bithorax complex: effects of insulator pairing on enhancer-promoter communication. *Mol Cell Biol* **27** (2007), pp. 3035-43.
- La Mantia, G., Majello, B., Di Cristofano, A., Strazzullo, M., Minchiotti, G. and Lania, L. Identification of regulatory elements within the minimal promoter region of the human endogenous ERV9 proviruses: accurate transcription initiation is controlled by an Inr-like element. *Nucleic Acids Res* **20** (1992), pp. 4129-36.
- Laemmli, U.K. and Johnson, R.A. Maturation of the head of bacteriophage T4. II. Head-related, aberrant tau-particles. *J Mol Biol* **80** (1973), pp. 601-11.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., et al. Initial sequencing and analysis of the human genome. *Nature* **409** (2001), pp. 860-921.
- Landry, J.R. and Mager, D.L. Functional analysis of the endogenous retroviral promoter of the human endothelin B receptor gene. *J Virol* **77** (2003), pp. 7459-66.



- Lapuk, A.V., Khil, P.P., Lavrentieva, I.V., Lebedev, Y.B. and Sverdlov, E.D. A human endogenous retrovirus-like (HERV) LTR formed more than 10 million years ago due to an insertion of HERV-H LTR into the 5' LTR of HERV-K is situated on human chromosomes 10, 19 and Y. *J Gen Virol* **80** ( Pt 4) (1999), pp. 835-9.
- Lavrentieva, I., Khil, P., Vinogradova, T., Akhmedov, A., Lapuk, A., Shakhova, O., Lebedev, Y., Monastyrskaya, G. and Sverdlov, E.D. Subfamilies and nearest-neighbour dendrogram for the LTRs of human endogenous retroviruses HERV-K mapped on human chromosome 19: physical neighbourhood does not correlate with identity level. *Hum Genet* **102** (1998), pp. 107-16.
- Lebedev, Y.B., Belonovitch, O.S., Zybrowa, N.V., Khil, P.P., Kurdyukov, S.G., Vinogradova, T.V., Hunsmann, G. and Sverdlov, E.D. Differences in HERV-K LTR insertions in orthologous loci of humans and great apes. *Gene* **247** (2000), pp. 265-77.
- Lebedev, Y.B., Volik, S.V., Obradovic, D., Ermolaeva, O.D., Ashworth, L.K., Lennon, G.G. and Sverdlov, E.D. Physical mapping of sequences homologous to an endogenous retrovirus LTR on human chromosome 19. *Mol Gen Genet* **247** (1995), pp. 742-8.
- Lee, Y.N. and Bieniasz, P.D. Reconstitution of an infectious human endogenous retrovirus. *PLoS Pathog* **3** (2007), p. e10.
- Leib-Mosch, C., Bachmann, M., Brack-Werner, R., Werner, T., Erfle, V. and Hehlmann, R. Expression and biological significance of human endogenous retroviral sequences. *Leukemia* **6 Suppl 3** (1992), pp. 72S-75S.
- Leib-Mosch, C., Haltmeier, M., Werner, T., Geigl, E.M., Brack-Werner, R., Francke, U., Erfle, V. and Hehlmann, R. Genomic distribution and transcription of solitary HERV-K LTRs. *Genomics* **18** (1993), pp. 261-9.
- Leib-Mosch, C. and Seifarth, W. Evolution and biological significance of human retroelements. *Virus Genes* **11** (1995), pp. 133-45.
- Leis, J., Baltimore, D., Bishop, J.M., Coffin, J., Fleissner, E., Goff, S.P., Oroszlan, S., Robinson, H., Skalka, A.M., Temin, H.M. and et al. Standardized and simplified nomenclature for proteins common to all retroviruses. *J Virol* **62** (1988), pp. 1808-9.
- Levin, H.L. and Moran, J.V. Dynamic interactions between transposable elements and their hosts. *Nat Rev Genet* **12** (2011), pp. 615-27.
- Levine M., Cattoglio C., Tjian R. Looping back to leap forward: transcription enters a new era. *Cell* **157** (2014), pp. 13-25.
- Li, D.J., Verma, D., Mosbrugger, T. and Swaminathan, S. CTCF and Rad21 act as host cell restriction factors for Kaposi's sarcoma-associated herpesvirus (KSHV) lytic replication by modulating viral gene transcription. *PLoS Pathog* **10** (2014), p. e1003880.
- Li, Q., Harju, S. and Peterson, K.R. Locus control regions: coming of age at a decade plus. *Trends Genet* **15** (1999), pp. 403-8.
- Li, T., Hu, J.F., Qiu, X., Ling, J., Chen, H., Wang, S., Hou, A., Vu, T.H. and Hoffman, A.R. CTCF regulates allelic expression of *Igf2* by orchestrating a promoter-polycomb repressive complex 2 intrachromosomal loop. *Mol Cell Biol* **28** (2008), pp. 6473-82.

- Li, T. and Lu, L. Epidermal growth factor-induced proliferation requires down-regulation of Pax6 in corneal epithelial cells. *J Biol Chem* **280** (2005), pp. 12988-95.
- Li, T., Lu, Z. and Lu, L. Regulation of eye development by transcription control of CCCTC binding factor (CTCF). *J Biol Chem* **279** (2004), pp. 27575-83.
- Li, T., Lu, Z. and Lu, L. Pax6 regulation in retinal cells by CCCTC binding factor. *Invest Ophthalmol Vis Sci* **47** (2006), pp. 5218-26.
- Libby, R.T., Hagerman, K.A., Pineda, V.V., Lau, R., Cho, D.H., Baccam, S.L., Axford, M.M., Cleary, J.D., Moore, J.M., Sopher, B.L., Tapscott, S.J., Filippova, G.N., Pearson, C.E. and La Spada, A.R. CTCF cis-regulates trinucleotide repeat instability in an epigenetic manner: a novel basis for mutational hot spot determination. *PLoS Genet* **4** (2008), p. e1000257.
- Lindeskog, M., Mager, D.L. and Blomberg, J. Isolation of a human endogenous retroviral HERV-H element with an open env reading frame. *Virology* **258** (1999), pp. 441-50.
- Lindeskog, M., Medstrand, P., Cunningham, A.A. and Blomberg, J. Coamplification and dispersion of adjacent human endogenous retroviral HERV-H and HERV-E elements; presence of spliced hybrid transcripts in normal leukocytes. *Virology* **244** (1998), pp. 219-29.
- Ling, J., Baibakov, B., Pi, W., Emerson, B.M. and Tuan, D. The HS2 enhancer of the beta-globin locus control region initiates synthesis of non-coding, polyadenylated RNAs independent of a cis-linked globin promoter. *J Mol Biol* **350** (2005), pp. 883-96.
- Ling, J.Q., Li, T., Hu, J.F., Vu, T.H., Chen, H.L., Qiu, X.W., Cherry, A.M. and Hoffman, A.R. CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science* **312** (2006), pp. 269-72.
- Litt, M.D., Simpson, M., Recillas-Targa, F., Prioleau, M.N. and Felsenfeld, G. Transitions in histone acetylation reveal boundaries of three separately regulated neighboring loci. *EMBO J* **20** (2001), pp. 2224-35.
- Lobanenkova, V.V. and Gudvin, G.G. CCCTC-binding protein: a new nuclear protein factor which interaction with 5'-flanking sequence of chicken c-myc oncogene correlates with repression of the gene. *Dokl Akad Nauk SSSR* **309** (1989), pp. 741-5.
- Lobanenkova, V.V., Nicolas, R.H., Adler, V.V., Paterson, H., Klenova, E.M., Polotskaja, A.V. and Goodwin, G.H. A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. *Oncogene* **5** (1990), pp. 1743-53.
- Lower, R., Lower, J. and Kurth, R. The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences. *Proc Natl Acad Sci U S A* **93** (1996), pp. 5177-84.
- Lower, R., Lower, J., Tondera-Koch, C. and Kurth, R. A general method for the identification of transcribed retrovirus sequences (R-U5 PCR) reveals the expression of the human endogenous retrovirus loci HERV-H and HERV-K in teratocarcinoma cells. *Virology* **192** (1993), pp. 501-11.

- Lu, L. and Tower, J. A transcriptional insulator element, the su(Hw) binding site, protects a chromosomal DNA replication origin from position effects. *Mol Cell Biol* **17** (1997), pp. 2202-6.
- MacPherson, M.J., Beatty, L.G., Zhou, W., Du, M. and Sadowski, P.D. The CTCF insulator protein is posttranslationally modified by SUMO. *Mol Cell Biol* **29** (2009), pp. 714-25.
- Mager, D.L. and Freeman, J.D. HERV-H endogenous retroviruses: presence in the New World branch but amplification in the Old World primate lineage. *Virology* **213** (1995), pp. 395-404.
- Mager, D.L., Hunter, D.G., Schertzer, M. and Freeman, J.D. Endogenous retroviruses provide the primary polyadenylation signal for two new human genes (HHLA2 and HHLA3). *Genomics* **59** (1999), pp. 255-63.
- Magin-Lachmann, C., Hahn, S., Strobel, H., Held, U., Lower, J. and Lower, R. Rec (formerly Corf) function requires interaction with a complex, folded RNA structure within its responsive element rather than binding to a discrete specific binding site. *J Virol* **75** (2001), pp. 10359-71.
- Magin, C., Lower, R. and Lower, J. cORF and RcRE, the Rev/Rex and RRE/RxRE homologues of the human endogenous retrovirus family HTDV/HERV-K. *J Virol* **73** (1999), pp. 9496-507.
- Majocchi, S., Aritonovska, E. and Mermod, N. Epigenetic regulatory elements associate with specific histone modifications to prevent silencing of telomeric genes. *Nucleic Acids Res* **42** (2013), pp. 193-204.
- Majumder, P., Gomez, J.A. and Boss, J.M. The human major histocompatibility complex class II HLA-DRB1 and HLA-DQA1 genes are separated by a CTCF-binding enhancer-blocking element. *J Biol Chem* **281** (2006), pp. 18435-43.
- Maksimenko, O. and Georgiev, P. Mechanisms and proteins involved in long-distance interactions. *Front Genet* **5** (2014), p. 28.
- Maksimenko, O., Golovnin, A. and Georgiev, P. Enhancer-promoter communication is regulated by insulator pairing in a Drosophila model bigenic locus. *Mol Cell Biol* **28** (2008), pp. 5469-77.
- Manavathi, B., Lo, D., Bugide, S., Dey, O., Imren, S., Weiss, M.J. and Humphries, R.K. Functional regulation of pre-B-cell leukemia homeobox interacting protein 1 (PBXIP1/HPIP) in erythroid differentiation. *J Biol Chem* **287** (2011), pp. 5600-14.
- Manghera, M. and Douville, R.N. Endogenous retrovirus-K promoter: a landing strip for inflammatory transcription factors? *Retrovirology* **10** (2013), p. 16.
- Martin, D., Pantoja, C., Fernandez Minan, A., Valdes-Quezada, C., Molto, E., Matesanz, F., Bogdanovic, O., de la Calle-Mustienes, E., Dominguez, O., Taher, L., Furlan-Magaril, M., Alcina, A., Canon, S., Fedetz, M., Blasco, M.A., Pereira, P.S., Ovcharenko, I., Recillas-Targa, F., Montoliu, L., Manzanares, M., Guigo, R., Serrano, M., Casares, F. and Gomez-Skarmeta, J.L. Genome-wide CTCF distribution in vertebrates defines equivalent sites that aid the identification of disease-associated genes. *Nat Struct Mol Biol* **18** (2011), pp. 708-14.
- Martin, M. A.Bryan, T.Rasheed, S.Khan, A. S. Identification and cloning of

- retroviral sequences present in human DNA *Proc Natl Acad Sci USA* **78** (1981), pp. 4892-4896.
- Martin, J., Herniou, E., Cook, J., Waugh O'Neill, R. and Tristem, M. Human endogenous retrovirus type I-related viruses have an apparently widespread distribution within vertebrates. *J Virol* **71** (1997), pp. 437-43.
- Mattern, K.A., Humbel, B.M., Muijsers, A.O., de Jong, L. and van Driel, R. hnRNP proteins and B23 are the major proteins of the internal nuclear matrix of HeLa S3 cells. *J Cell Biochem* **62** (1996), pp. 275-89.
- Matys, V., Kel-Margoulis, O.V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., Voss, N., Stegmaier, P., Lewicki-Potapov, B., Saxel, H., Kel, A.E. and Wingender, E. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* **34** (2006), pp. D108-10.
- Mayer, W.E., O'HUigin, C. and Klein, J. Resolution of the HLA-DRB6 puzzle: a case of grafting a de novo-generated exon on an existing gene. *Proc Natl Acad Sci U S A* **90** (1993), pp. 10720-4.
- Medstrand, P. and Blomberg, J. Characterization of novel reverse transcriptase encoding human endogenous retroviral sequences similar to type A and type B retroviruses: differential transcription in normal human tissues. *J Virol* **67** (1993), pp. 6778-87.
- Medstrand, P., Landry, J.R. and Mager, D.L. Long terminal repeats are used as alternative promoters for the endothelin B receptor and apolipoprotein C-I genes in humans. *J Biol Chem* **276** (2001), pp. 1896-903.
- Medstrand, P., Lindeskog, M. and Blomberg, J. Expression of human endogenous retroviral sequences in peripheral blood mononuclear cells of healthy individuals. *J Gen Virol* **73** ( Pt 9) (1992), pp. 2463-6.
- Medstrand, P. and Mager, D.L. Human-specific integrations of the HERV-K endogenous retrovirus family. *J Virol* **72** (1998), pp. 9782-7.
- Medstrand, P., Mager, D.L., Yin, H., Dietrich, U. and Blomberg, J. Structure and genomic organization of a novel human endogenous retrovirus family: HERV-K (HML-6). *J Gen Virol* **78** ( Pt 7) (1997), pp. 1731-44.
- Melfi, R., Palla, F., Di Simone, P., Alessandro, C., Cali, L., Anello, L. and Spinelli, G. Functional characterization of the enhancer blocking element of the sea urchin early histone gene cluster reveals insulator properties and three essential cis-acting sequences. *J Mol Biol* **304** (2000), pp. 753-63.
- Meneghini, M.D., Wu, M. and Madhani, H.D. Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin. *Cell* **112** (2003), pp. 725-36.
- Mighell, A.J., Markham, A.F. and Robinson, P.A. Alu sequences. *FEBS Lett* **417** (1997), pp. 1-5.
- Milman, G. and Herzberg, M. Efficient DNA transfection and rapid assay for thymidine kinase activity and viral antigenic determinants. *Somatic Cell Genet* **7** (1981), pp. 161-70.
- Moggs, J.G. and Orphanides, G. Genomic analysis of stress response genes. *Toxicol Lett* **140-141** (2003), pp. 149-53.

- Molto, E., Fernandez, A. and Montoliu, L. Boundaries in vertebrate genomes: different solutions to adequately insulate gene expression domains. *Brief Funct Genomic Proteomic* **8** (2009), pp. 283-96.
- Mongelard, F. and Corces, V.G. Two insulators are not better than one. *Nat Struct Biol* **8** (2001), pp. 192-4.
- Moon, H., Filippova, G., Loukinov, D., Pugacheva, E., Chen, Q., Smith, S.T., Munhall, A., Grewe, B., Bartkuhn, M., Arnold, R., Burke, L.J., Renkawitz-Pohl, R., Ohlsson, R., Zhou, J., Renkawitz, R. and Lobanenko, V. CTCF is conserved from *Drosophila* to humans and confers enhancer blocking of the Fab-8 insulator. *EMBO Rep* **6** (2005), pp. 165-70.
- Moore, J.M., Rabaia, N.A., Smith, L.E., Fagerlie, S., Gurley, K., Loukinov, D., Disteche, C.M., Collins, S.J., Kemp, C.J., Lobanenko, V.V. and Filippova, G.N. Loss of maternal CTCF is associated with peri-implantation lethality of *Ctcf* null embryos. *PLoS One* **7** (2012), p. e34915.
- Morimura, N., Inoue, T., Katayama, K. and Aruga, J. Comparative analysis of structure, expression and PSD95-binding capacity of *Lrln*, a novel family of neuronal transmembrane proteins. *Gene* **380** (2006), pp. 72-83.
- Morozov, V.A., Dao Thi, V.L. and Denner, J. The transmembrane protein of the human endogenous retrovirus--K (HERV-K) modulates cytokine release and gene expression. *PLoS One* **8** (2013), p. e70399.
- Mukhopadhyay, R., Yu, W., Whitehead, J., Xu, J., Lezcano, M., Pack, S., Kanduri, C., Kanduri, M., Ginja, V., Vostrov, A., Quitschke, W., Chernukhin, I., Klenova, E., Lobanenko, V. and Ohlsson, R. The binding sites for the chromatin insulator protein CTCF map to DNA methylation-free domains genome-wide. *Genome Res* **14** (2004), pp. 1594-602.
- Muravyova, E., Golovnin, A., Gracheva, E., Parshikov, A., Belenkaya, T., Pirrotta, V. and Georgiev, P. Loss of insulator activity by paired *Su(Hw)* chromatin insulators. *Science* **291** (2001), pp. 495-8.
- Murrell, A., Heeson, S. and Reik, W. Interaction between differentially methylated regions partitions the imprinted genes *Igf2* and *H19* into parent-specific chromatin loops. *Nat Genet* **36** (2004), pp. 889-93.
- Mutskov, V.J., Farrell, C.M., Wade, P.A., Wolffe, A.P. and Felsenfeld, G. The barrier function of an insulator couples high histone acetylation levels with specific protection of promoter DNA from methylation. *Genes Dev* **16** (2002), pp. 1540-54.
- Nabirochkin, S., Ossokina, M. and Heidmann, T. A nuclear matrix/scaffold attachment region co-localizes with the gypsy retrotransposon insulator sequence. *J Biol Chem* **273** (1998), pp. 2473-9.
- Namciu, S.J., Blochlinger, K.B. and Fournier, R.E. Human matrix attachment regions insulate transgene expression from chromosomal position effects in *Drosophila melanogaster*. *Mol Cell Biol* **18** (1998), pp. 2382-91.
- Nasmyth, K. Cohesin: a catenase with separate entry and exit gates? *Nat Cell Biol* **13** (2011), pp. 1170-7.
- Nativio, R., Sparago, A., Ito, Y., Weksberg, R., Riccio, A. and Murrell, A. Disruption of genomic neighbourhood at the imprinted *IGF2-H19* locus in Beckwith-

- Wiedemann syndrome and Silver-Russell syndrome. *Hum Mol Genet* **20** (2011), pp. 1363-74.
- Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B.R., Mirny, L.A. and Dekker, J. Organization of the mitotic chromosome. *Science* **342** (2013), pp. 948-53.
- Nelson, D.T., Goodchild, N.L. and Mager, D.L. Gain of Sp1 sites and loss of repressor sequences associated with a young, transcriptionally active subset of HERV-H endogenous long terminal repeats. *Virology* **220** (1996), pp. 213-8.
- Nelson, P.N., Carnegie, P.R., Martin, J., Davari Ejtehadi, H., Hooley, P., Roden, D., Rowland-Jones, S., Warren, P., Astley, J. and Murray, P.G. Demystified. Human endogenous retroviruses. *Mol Pathol* **56** (2003), pp. 11-8.
- Ni, X., Zhang, Y.E., Negre, N., Chen, S., Long, M. and White, K.P. Adaptive evolution and the birth of CTCF binding sites in the Drosophila genome. *PLoS Biol* **10** (2012), p. e1001420.
- Nikolaev, L.G. Identification and isolation of proteins, recognizing the sequence of the human immunodeficiency virus (HIV-1) enhancer. *Mol Biol (Mosk)* **30** (1996), pp. 714-20.
- Nikolaev, L.G., Glotov, B.O., Belyavsky, A.V., Grachev, S.A. and Levin, A.V. Identification of sequence-specific DNA-binding factors by label transfer: application to the adenovirus-2 major late promoter. *Nucleic Acids Res* **16** (1988), pp. 519-35.
- Nomiyama, H., Fromental, C., Xiao, J.H. and Chambon, P. Cell-specific activity of the constituent elements of the simian virus 40 enhancer. *Proc Natl Acad Sci U S A* **84** (1987), pp. 7881-5.
- Norris, J., Fan, D., Aleman, C., Marks, J.R., Futreal, P.A., Wiseman, R.W., Iglehart, J.D., Deininger, P.L. and McDonnell, D.P. Identification of a new subclass of Alu DNA repeats which can function as estrogen receptor-dependent transcriptional enhancers. *J Biol Chem* **270** (1995), pp. 22777-82.
- Ogbourne, S. and Antalis, T.M. Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *Biochem J* **331** ( Pt 1) (1998), pp. 1-14.
- Ohlsson, R., Lobanenkov, V. and Klenova, E. Does CTCF mediate between nuclear organization and gene expression? *Bioessays* **32** (2009), pp. 37-50.
- Ohtsuki, S. and Levine, M. GAGA mediates the enhancer blocking activity of the eve promoter in the Drosophila embryo. *Genes Dev* **12** (1998), pp. 3325-30.
- Okamura, E., Matsuzaki, H., Fukamizu, A. and Tanimoto, K. The chicken HS4 insulator element does not protect the H19 ICR from differential DNA methylation in yeast artificial chromosome transgenic mouse. *PLoS One* **8** (2013), p. e73925.
- Okumura, K., Sakaguchi, G., Takagi, S., Naito, K., Mimori, T. and Igarashi, H. Sp1 family proteins recognize the U5 repressive element of the long terminal repeat of human T cell leukemia virus type I through binding to the CACCC core motif. *J Biol Chem* **271** (1996), pp. 12944-50.
- Okuwaki, M. The structure and functions of NPM1/Nucleophsmin/B23, a multifunctional nucleolar acidic protein. *J Biochem* **143** (2008), pp. 441-8.

- Ong, C.T. and Corces, V.G. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet* **15** (2014), pp. 234-46.
- Onn, I., Heidinger-Pauli, J.M., Guacci, V., Unal, E. and Koshland, D.E. Sister chromatid cohesion: a simple concept with a complex reality. *Annu Rev Cell Dev Biol* **24** (2008), pp. 105-29.
- Ono, M. Molecular cloning and long terminal repeat sequences of human endogenous retrovirus genes related to types A and B retrovirus genes. *J Virol* **58** (1986), pp. 937-44.
- Orlando, V. Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. *Trends Biochem Sci* **25** (2000), pp. 99-104.
- Osumi, N., Shinohara, H., Numayama-Tsuruta, K. and Maekawa, M. Concise review: Pax6 transcription factor contributes to both embryonic and adult neurogenesis as a multifunctional regulator. *Stem Cells* **26** (2008), pp. 1663-72.
- Ouboussad, L., Kreuz, S. and Lefevre, P.F. CTCF depletion alters chromatin structure and transcription of myeloid-specific factors. *J Mol Cell Biol* **5** (2013), pp. 308-22.
- Pai, C.Y., Lei, E.P., Ghosh, D. and Corces, V.G. The centrosomal protein CP190 is a component of the gypsy chromatin insulator. *Mol Cell* **16** (2004), pp. 737-48.
- Palla, F., Melfi, R., Anello, L., Di Bernardo, M. and Spinelli, G. Enhancer blocking activity located near the 3' end of the sea urchin early H2A histone gene. *Proc Natl Acad Sci U S A* **94** (1997), pp. 2272-7.
- Pant, V., Kurukuti, S., Pugacheva, E., Shamsuddin, S., Mariano, P., Renkawitz, R., Klenova, E., Lobanekov, V. and Ohlsson, R. Mutation of a single CTCF target site within the H19 imprinting control region leads to loss of Igf2 imprinting and complex patterns of de novo methylation upon maternal inheritance. *Mol Cell Biol* **24** (2004), pp. 3497-504.
- Pantel, J., Machinis, K., Sobrier, M.L., Duquesnoy, P., Goossens, M. and Amselem, S. Species-specific alternative splice mimicry at the growth hormone receptor locus revealed by the lineage of retroelements during primate evolution. *J Biol Chem* **275** (2000), pp. 18664-9.
- Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., Cobb, B.S., Yokomori, K., Dillon, N., Aragon, L., Fisher, A.G. and Merckenschlager, M. Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* **132** (2008), pp. 422-33.
- Parnell, T.J. and Geyer, P.K. Differences in insulator properties revealed by enhancer blocking assays on episomes. *EMBO J* **19** (2000), pp. 5864-74.
- Peeters, A., Lambert, P.F. and Deacon, N.J. A fourth Sp1 site in the human immunodeficiency virus type 1 long terminal repeat is essential for negative-sense transcription. *J Virol* **70** (1996), pp. 6665-72.
- Pereira, L.A., Bentley, K., Peeters, A., Churchill, M.J. and Deacon, N.J. A compilation of cellular transcription factor interactions with the HIV-1 LTR promoter. *Nucleic Acids Res* **28** (2000), pp. 663-8.
- Peters, J.M., Tedeschi, A. and Schmitz, J. The cohesin complex and its roles in chromosome biology. *Genes Dev* **22** (2008), pp. 3089-114.

- Pikaart, M.J., Recillas-Targa, F. and Felsenfeld, G. Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators. *Genes Dev* **12** (1998), pp. 2852-62.
- Poiesz, B.J., Ruscetti, F.W., Gazdar, A.F., Bunn, P.A., Minna, J.D. and Gallo, R.C. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proc Natl Acad Sci U S A* **77** (1980), pp. 7415-9.
- Prendergast, J.G., Chambers, E.V. and Semple, C.A. Sequence level mechanisms of human epigenome evolution. *Genome Biol Evol* (2014).
- Prickett, A.R., Barkas, N., McCole, R.B., Hughes, S., Amante, S.M., Schulz, R. and Oakey, R.J. Genome-wide and parental allele-specific analysis of CTCF and cohesin DNA binding in mouse brain reveals a tissue-specific binding pattern and an association with imprinted differentially methylated regions. *Genome Res* **23** (2013), pp. 1624-35.
- Prioleau, M.N., Nony, P., Simpson, M. and Felsenfeld, G. An insulator element and condensed chromatin region separate the chicken beta-globin locus from an independently regulated erythroid-specific folate receptor gene. *EMBO J* **18** (1999), pp. 4035-48.
- Pugacheva, E.M., Kwon, Y.W., Hukriede, N.A., Pack, S., Flanagan, P.T., Ahn, J.C., Park, J.A., Choi, K.S., Kim, K.W., Loukinov, D., Dawid, I.B. and Lobanenko, V.V. Cloning and characterization of zebrafish CTCF: Developmental expression patterns, regulation of the promoter region, and evolutionary aspects of gene organization. *Gene* **375** (2006), pp. 26-36.
- Qin, X., Zhou, H., Liu, L. and Gerson, S.L. Transgenic expression of human MGMT blocks the hypersensitivity of PMS2-deficient mice to low dose MNU thymic lymphomagenesis. *Carcinogenesis* **20** (1999), pp. 1667-73.
- Rabbi, M.F., Saifuddin, M., Gu, D.S., Kagnoff, M.F. and Roebuck, K.A. U5 region of the human immunodeficiency virus type 1 long terminal repeat contains TRE-like cAMP-responsive elements that bind both AP-1 and CREB/ATF proteins. *Virology* **233** (1997), pp. 235-45.
- Rakha, E.A., Pinder, S.E., Paish, C.E. and Ellis, I.O. Expression of the transcription factor CTCF in invasive breast cancer: a candidate gene located at 16q22.1. *Br J Cancer* **91** (2004), pp. 1591-6.
- Ramezani, A., Hawley, T.S. and Hawley, R.G. Combinatorial incorporation of enhancer-blocking components of the chicken beta-globin 5'HS4 and human T-cell receptor alpha/delta BEAD-1 insulators in self-inactivating retroviral vectors reduces their genotoxic potential. *Stem Cells* **26** (2008), pp. 3257-66.
- Razin, S.V. and Ioudinkova, E.S. Mechanisms controlling activation of the alpha-globin gene domain in chicken erythroid cells. *Biochemistry (Mosc)* **72** (2007), pp. 467-70.
- Reale, A., Matteis, G.D., Galleazzi, G., Zampieri, M. and Caiafa, P. Modulation of DNMT1 activity by ADP-ribose polymers. *Oncogene* **24** (2005), pp. 13-9.
- Recillas-Targa, F., Bell, A.C. and Felsenfeld, G. Positional enhancer-blocking activity of the chicken beta-globin insulator in transiently transfected cells. *Proc Natl Acad Sci U S A* **96** (1999), pp. 14354-9.



- Recillas-Targa, F., Pikaart, M.J., Burgess-Beusse, B., Bell, A.C., Litt, M.D., West, A.G., Gaszner, M. and Felsenfeld, G. Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proc Natl Acad Sci U S A* **99** (2002), pp. 6883-8.
- Renda, M., Baglivo, I., Burgess-Beusse, B., Esposito, S., Fattorusso, R., Felsenfeld, G. and Pedone, P.V. Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. *J Biol Chem* **282** (2007), pp. 33336-45.
- Roberts, J., Scott, A.C., Howard, M.R., Breen, G., Bubb, V.J., Klenova, E. and Quinn, J.P. Differential regulation of the serotonin transporter gene by lithium is mediated by transcription factors, CCCTC binding protein and Y-box binding protein 1, through the polymorphic intron 2 variable number tandem repeat. *J Neurosci* **27** (2007), pp. 2793-801.
- Robinett, C.C., O'Connor, A. and Dunaway, M. The repeat organizer, a specialized insulator element within the intergenic spacer of the *Xenopus* rRNA genes. *Mol Cell Biol* **17** (1997), pp. 2866-75.
- Rosenbloom, K.R., Dreszer, T.R., Pheasant, M., Barber, G.P., Meyer, L.R., Pohl, A., Raney, B.J., Wang, T., Hinrichs, A.S., Zweig, A.S., Fujita, P.A., Learned, K., Rhead, B., Smith, K.E., Kuhn, R.M., Karolchik, D., Haussler, D. and Kent, W.J. ENCODE whole-genome data in the UCSC Genome Browser. *Nucleic Acids Res* **38** (2009), pp. D620-5.
- Rousseau, M., Crutchley, J.L., Miura, H., Suderman, M., Blanchette, M. and Dostie, J. Hox in motion: tracking HoxA cluster conformation during differentiation. *Nucleic Acids Res* **42** (2013), pp. 1524-40.
- Rubio, E.D., Reiss, D.J., Welch, P.L., Disteche, C.M., Filippova, G.N., Baliga, N.S., Aebersold, R., Ranish, J.A. and Krumm, A. CTCF physically links cohesin to chromatin. *Proc Natl Acad Sci U S A* **105** (2008), pp. 8309-14.
- Ruda, V.M., Akopov, S.B., Trubetskoy, D.O., Manuylov, N.L., Vetchinova, A.S., Zavalova, L.L., Nikolaev, L.G. and Sverdlov, E.D. Tissue specificity of enhancer and promoter activities of a HERV-K(HML-2) LTR. *Virus Res* **104** (2004), pp. 11-6.
- Ruiz-Narvaez, E.A. and Campos, H. Evolutionary rate heterogeneity of Alu repeats upstream of the APOA5 gene: do they regulate APOA5 expression? *J Hum Genet* **53** (2008), pp. 247-53.
- Ruprecht, K., Ferreira, H., Flockerzi, A., Wahl, S., Sauter, M., Mayer, J. and Mueller-Lantsch, N. Human endogenous retrovirus family HERV-K(HML-2) RNA transcripts are selectively packaged into retroviral particles produced by the human germ cell tumor line Tera-1 and originate mainly from a provirus on chromosome 22q11.21. *J Virol* **82** (2008), pp. 10008-16.
- Sambrook, J. and Russell, D.W., Molecular Cloning. A laboratory Manual., 3rd Edition ed. CSHL Press, Cold Spring Harbor (2001).
- Santoni, F.A., Guerra, J. and Luban, J. HERV-H RNA is abundant in human embryonic stem cells and a precise marker for pluripotency. *Retrovirology* **9** (2012), p. 111.

- Sass, A.V., Ruda, V.M., Akopov, S.B., Snezhkov, E.V., Nikolaev, L.G. and Sverdlov, E.D. Regulatory potential of S/MAR elements in transient expression. *Bioorg Khim* **31** (2005), pp. 77-81.
- Scherdin, U., Rhodes, K. and Breindl, M. Transcriptionally active genome regions are preferred targets for retrovirus integration. *J Virol* **64** (1990), pp. 907-12.
- Schmidt, D., Schwalie, P.C., Wilson, M.D., Ballester, B., Goncalves, A., Kutter, C., Brown, G.D., Marshall, A., Flicek, P. and Odom, D.T. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **148** (2012), pp. 335-48.
- Schmitt, K., Reichrath, J., Roesch, A., Meese, E. and Mayer, J. Transcriptional profiling of human endogenous retrovirus group HERV-K(HML-2) loci in melanoma. *Genome Biol Evol* **5** (2013), pp. 307-28.
- Schon, U., Seifarth, W., Baust, C., Hohenadl, C., Erfle, V. and Leib-Mosch, C. Cell type-specific expression and promoter activity of human endogenous retroviral long terminal repeats. *Virology* **279** (2001), pp. 280-91.
- Schoorlemmer, J., Perez-Palacios, R., Climent, M., Guallar, D. and Muniesa, P. Regulation of Mouse Retroelement MuERV-L/MERVL Expression by REX1 and Epigenetic Control of Stem Cell Potency. *Front Oncol* **4** (2014), p. 14.
- Schulte, A.M. and Wellstein, A. Structure and phylogenetic analysis of an endogenous retrovirus inserted into the human growth factor pleiotrophin. *J Virol* **72** (1998), pp. 6065-72.
- Schwalie, P.C., Ward, M.C., Cain, C.E., Faure, A.J., Gilad, Y., Odom, D.T. and Flicek, P. Co-binding by YY1 identifies the transcriptionally active, highly conserved set of CTCF-bound regions in primate genomes. *Genome Biol* **14** (2014), p. R148.
- Schwartz, Y.B., Linder-Basso, D., Kharchenko, P.V., Tolstorukov, M.Y., Kim, M., Li, H.B., Gorchakov, A.A., Minoda, A., Shanower, G., Alekseyenko, A.A., Riddle, N.C., Jung, Y.L., Gu, T., Plachetka, A., Elgin, S.C., Kuroda, M.I., Park, P.J., Savitsky, M., Karpen, G.H. and Pirrotta, V. Nature and function of insulator protein binding sites in the Drosophila genome. *Genome Res* **22** (2012), pp. 2188-98.
- Scott, K.C., Taubman, A.D. and Geyer, P.K. Enhancer blocking by the Drosophila gypsy insulator depends upon insulator anatomy and enhancer strength. *Genetics* **153** (1999), pp. 787-98.
- Seifarth, W., Baust, C., Murr, A., Skladny, H., Krieg-Schneider, F., Blusch, J., Werner, T., Hehlmann, R. and Leib-Mosch, C. Proviral structure, chromosomal location, and expression of HERV-K-T47D, a novel human endogenous retrovirus derived from T47D particles. *J Virol* **72** (1998), pp. 8384-91.
- Seiki, M., Hikikoshi, A. and Yoshida, M. The U5 sequence is a cis-acting repressive element for genomic RNA expression of human T cell leukemia virus type I. *Virology* **176** (1990), pp. 81-6.
- Seo, H.W., Kim, T.M., Choi, J.W., Han, B.K., Song, G. and Han, J.Y. Evaluation of combinatorial cis-regulatory elements for stable gene expression in chicken cells. *BMC Biotechnol* **10** (2010), p. 69.
- Shank, P.R., Hughes, S.H., Kung, H.J., Majors, J.E., Quintrell, N., Guntaka, R.V., Bishop, J.M. and Varmus, H.E. Mapping unintegrated avian sarcoma virus DNA:

- termini of linear DNA bear 300 nucleotides present once or twice in two species of circular DNA. *Cell* **15** (1978), pp. 1383-95.
- Shevchenko, A., Wilm, M., Vorm, O. and Mann, M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* **68** (1996), pp. 850-8.
- Shields, R. The emperor's new clothes revisited. *Trends Genet* **22** (2006), p. 463.
- Shin, W., Lee, J., Son, S.Y., Ahn, K., Kim, H.S. and Han, K. Human-specific HERV-K insertion causes genomic variations in the human genome. *PLoS One* **8** (2013), p. e60605.
- Shrimali, S., Srivastava, S., Varma, G., Grinberg, A., Pfeifer, K. and Srivastava, M. An ectopic CTCF-dependent transcriptional insulator influences the choice of Vbeta gene segments for VDJ recombination at TCRbeta locus. *Nucleic Acids Res* **40** (2012), pp. 7753-65.
- Simonis, M., Kooren, J. and de Laat, W. An evaluation of 3C-based methods to capture DNA interactions. *Nat Methods* **4** (2007), pp. 895-901.
- Singh, P., Lee, D.H. and Szabo, P.E. More than insulator: multiple roles of CTCF at the H19-Igf2 imprinted domain. *Front Genet* **3** (2012), p. 214.
- Singh, V. and Srivastava, M. Enhancer blocking activity of the insulator at H19-ICR is independent of chromatin barrier establishment. *Mol Cell Biol* **28** (2008), pp. 3767-75.
- Sjottem, E., Anderssen, S. and Johansen, T. The promoter activity of long terminal repeats of the HERV-H family of human retrovirus-like elements is critically dependent on Sp1 family proteins interacting with a GC/GT box located immediately 3' to the TATA box. *J Virol* **70** (1996), pp. 188-98.
- Sompayrac, L.M. and Danna, K.J. Efficient infection of monkey cells with DNA of simian virus 40. *Proc Natl Acad Sci U S A* **78** (1981), pp. 7575-8.
- Spencer, R.J., del Rosario, B.C., Pinter, S.F., Lessing, D., Sadreyev, R.I. and Lee, J.T. A boundary element between Tsix and Xist binds the chromatin insulator Ctcf and contributes to initiation of X-chromosome inactivation. *Genetics* **189** (2011), pp. 441-54.
- Splinter, E., Heath, H., Kooren, J., Palstra, R.J., Klous, P., Grosveld, F., Galjart, N. and de Laat, W. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev* **20** (2006), pp. 2349-54.
- Stedman, W., Kang, H., Lin, S., Kissil, J.L., Bartolomei, M.S. and Lieberman, P.M. Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. *EMBO J* **27** (2008), pp. 654-66.
- Strazzullo, M., Majello, B., Lania, L. and La Mantia, G. Mutational analysis of the human endogenous ERV9 proviruses promoter region. *Virology* **200** (1994), pp. 686-95.
- Strazzullo, M., Parisi, T., Di Cristofano, A., Rocchi, M. and La Mantia, G. Characterization and genomic mapping of chimeric ERV9 endogenous retroviruses-host gene transcripts. *Gene* **206** (1998), pp. 77-83.
- Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., Cooke, M.P., Walker, J.R. and Hogenesch, J.B. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci U S A* **101** (2004), pp. 6062-7.

- Subramanian, R.P., Wildschutte, J.H., Russo, C. and Coffin, J.M. Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. *Retrovirology* **8** (2011), p. 90.
- Sverdlov, E.D. Perpetually mobile footprints of ancient infections in human genome. *FEBS Lett* **428** (1998), pp. 1-6.
- Sverdlov, E.D. Retroviruses and primate evolution. *Bioessays* **22** (2000), pp. 161-71.
- Tacke, S.J., Specke, V. and Denner, J. Differences in release and determination of subtype of porcine endogenous retroviruses produced by stimulated normal pig blood cells. *Intervirology* **46** (2003), pp. 17-24.
- Takaki, R., Watson, S.R. and Lanier, L.L. DAP12: an adapter protein with dual functionality. *Immunol Rev* **214** (2006), pp. 118-29.
- Taslim, C., Chen, Z., Huang, K., Huang, T.H., Wang, Q. and Lin, S. Integrated analysis identifies a class of androgen-responsive genes regulated by short combinatorial long-range mechanism facilitated by CTCF. *Nucleic Acids Res* **40** (2012), pp. 4754-64.
- Temin, H.M. The provirus hypothesis: speculations on the significance of RNA-directed DNA synthesis for normal development and for carcinogenesis. *J Natl Cancer Inst* **46** (1971), pp. 3-7.
- Temin, H.M. Origin of retroviruses from cellular moveable genetic elements. *Cell* **21** (1980), pp. 599-600.
- Temin, H.M. Structure, variation and synthesis of retrovirus long terminal repeat. *Cell* **27** (1981), pp. 1-3.
- Temin, H.M. and Mizutani, S. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* **226** (1970), pp. 1211-3.
- Thomas, B.J., Rubio, E.D., Krumm, N., Broin, P.O., Bomsztyk, K., Welcsh, P., Grealley, J.M., Golden, A.A. and Krumm, A. Allele-specific transcriptional elongation regulates monoallelic expression of the IGF2BP1 gene. *Epigenetics Chromatin* **4** (2011), p. 14.
- Thomas, J.H. and Schneider, S. Coevolution of retroelements and tandem zinc finger genes. *Genome Res* **21** (2011), pp. 1800-12.
- Thomas, M.J. and Seto, E. Unlocking the mechanisms of transcription factor YY1: are chromatin modifying enzymes the key? *Gene* **236** (1999), pp. 197-208.
- Tonjes, R.R., Czauderna, F. and Kurth, R. Genome-wide screening, cloning, chromosomal assignment, and expression of full-length human endogenous retrovirus type K. *J Virol* **73** (1999), pp. 9187-95.
- Torrano, V., Chernukhin, I., Docquier, F., D'Arcy, V., Leon, J., Klenova, E. and Delgado, M.D. CTCF regulates growth and erythroid differentiation of human myeloid leukemia cells. *J Biol Chem* **280** (2005), pp. 28152-61.
- Towler, E.M., Gulnik, S.V., Bhat, T.N., Xie, D., Gustschina, E., Sumpter, T.R., Robertson, N., Jones, C., Sauter, M., Mueller-Lantsch, N., Debouck, C. and Erickson, J.W. Functional characterization of the protease of human endogenous retrovirus, K10: can it complement HIV-1 protease? *Biochemistry* **37** (1998), pp. 17137-44.

- Tristem, M. Identification and characterization of novel human endogenous retrovirus families by phylogenetic screening of the human genome mapping project database. *J Virol* **74** (2000), pp. 3715-30.
- Tuan, D., Kong, S. and Hu, K. Transcription of the hypersensitive site HS2 enhancer in erythroid cells. *Proc Natl Acad Sci U S A* **89** (1992), pp. 11219-23.
- Tugnet, N., Rylance, P., Roden, D., Trela, M. and Nelson, P. Human Endogenous Retroviruses (HERVs) and Autoimmune Rheumatic Disease: Is There a Link? *Open Rheumatol J* **7** (2013), pp. 13-21.
- Turelli, P., Castro-Diaz, N., Marzetta, F., Kapopoulou, A., Raclot, C., Duc, J., Tieng, V., Quenneville, S. and Trono, D. Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. *Genome Res* (2014).
- Tybulewicz, V.L., Crawford, C.E., Jackson, P.K., Bronson, R.T. and Mulligan, R.C. Neonatal lethality and lymphopenia in mice with a homozygous disruption of the c-abl proto-oncogene. *Cell* **65** (1991), pp. 1153-63.
- Urnovitz, H.B. and Murphy, W.H. Human endogenous retroviruses: nature, occurrence, and clinical implications in human disease. *Clin Microbiol Rev* **9** (1996), pp. 72-99.
- Valadez-Graham, V., Razin, S.V. and Recillas-Targa, F. CTCF-dependent enhancer blockers at the upstream region of the chicken alpha-globin gene domain. *Nucleic Acids Res* **32** (2004), pp. 1354-62.
- Valenzuela, L. and Kamakaka, R.T. Chromatin insulators. *Annu Rev Genet* **40** (2006), pp. 107-38.
- Varmus, H.E. Form and function of retroviral proviruses. *Science* **216** (1982), pp. 812-20.
- Vetchinova, A.S., Akopov, S.B., Chernov, I.P., Nikolaev, L.G. and Sverdlov, E.D. Two-dimensional electrophoretic mobility shift assay: identification and mapping of transcription factor CTCF target sequences within an FXVD5-COX7A1 region of human chromosome 19. *Anal Biochem* **354** (2006), pp. 85-93.
- Vinogradova, T., Volik, S., Lebedev, Y., Shevchenko, Y., Lavrentyeva, I., Khil, P., Grzeschik, K.H., Ashworth, L.K. and Sverdlov, E. Positioning of 72 potentially full size LTRs of human endogenous retroviruses HERV-K on the human chromosome 19 map. Occurrences of the LTRs in human gene sites. *Gene* **199** (1997), pp. 255-64.
- Vinogradova, T.V., Leppik, L.P., Nikolaev, L.G., Akopov, S.B., Kleiman, A.M., Senyuta, N.B. and Sverdlov, E.D. Solitary human endogenous retroviruses-K LTRs retain transcriptional activity in vivo, the mode of which is different in different cell types. *Virology* **290** (2001), pp. 83-90.
- Vydra, J., Vytasek, R., Sainerova, H., Mandys, V., Sovova, V., Hejnar, J., Sloncova, E., Pouckova, P., Bubenik, J., Novaak, J. and et al. A new cell line, GS, derived from a human renal cell carcinoma. *Folia Biol (Praha)* **34** (1988), pp. 308-15.
- Walters, M.C., Fiering, S., Bouhassira, E.E., Scalzo, D., Goeke, S., Magis, W., Garrick, D., Whitelaw, E. and Martin, D.I. The chicken beta-globin 5'HS4 boundary element blocks enhancer-mediated suppression of silencing. *Mol Cell Biol* **19** (1999), pp. 3714-26.

- Wan, L.B., Pan, H., Hannenhalli, S., Cheng, Y., Ma, J., Fedoriw, A., Lobanenkov, V., Latham, K.E., Schultz, R.M. and Bartolomei, M.S. Maternal depletion of CTCF reveals multiple functions during oocyte and preimplantation embryo development. *Development* **135** (2008), pp. 2729-38.
- Wang, H., Maurano, M.T., Qu, H., Varley, K.E., Gertz, J., Pauli, F., Lee, K., Canfield, T., Weaver, M., Sandstrom, R., Thurman, R.E., Kaul, R., Myers, R.M. and Stamatoyannopoulos, J.A. Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res* **22** (2012), pp. 1680-8.
- Wang, J., Lunnyak, V.V. and Jordan, I.K. Genome-wide prediction and analysis of human chromatin boundary elements. *Nucleic Acids Res* **40** (2011), pp. 511-29.
- Weiss, R.A. and Stoye, J.P. Virology. Our viral inheritance. *Science* **340** (2013), pp. 820-1.
- Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., Yahata, K., Imamoto, F., Aburatani, H., Nakao, M., Imamoto, N., Maeshima, K., Shirahige, K. and Peters, J.M. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451** (2008), pp. 796-801.
- West, A.G. and Fraser, P. Remote control of gene transcription. *Hum Mol Genet* **14 Spec No 1** (2005), pp. R101-11.
- West, A.G., Gaszner, M. and Felsenfeld, G. Insulators: many functions, many mechanisms. *Genes Dev* **16** (2002), pp. 271-88.
- Whitfield, T.W., Wang, J., Collins, P.J., Partridge, E.C., Aldred, S.F., Trinklein, N.D., Myers, R.M. and Weng, Z. Functional analysis of transcription factor binding sites in human promoters. *Genome Biol* **13** (2012), p. R50.
- Williams, A. and Flavell, R.A. The role of CTCF in regulating nuclear organization. *J Exp Med* **205** (2008), pp. 747-50.
- Willoughby, D.A., Vilalta, A. and Oshima, R.G. An Alu element from the K18 gene confers position-independent expression in transgenic mice. *J Biol Chem* **275** (2000), pp. 759-68.
- Wilson, C.A., Laeeq, S., Ritzhaupt, A., Colon-Moran, W. and Yoshimura, F.K. Sequence analysis of porcine endogenous retrovirus long terminal repeats and identification of transcriptional regulatory regions. *J Virol* **77** (2003), pp. 142-9.
- Wu, J., Song, Y., Bakker, A.B., Bauer, S., Spies, T., Lanier, L.L. and Phillips, J.H. An activating immunoreceptor complex formed by NKG2D and DAP10. *Science* **285** (1999), pp. 730-2.
- Xi, H., Shulha, H.P., Lin, J.M., Vales, T.R., Fu, Y., Bodine, D.M., McKay, R.D., Chenoweth, J.G., Tesar, P.J., Furey, T.S., Ren, B., Weng, Z. and Crawford, G.E. Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. *PLoS Genet* **3** (2007), p. e136.
- Xie, X., Mikkelsen, T.S., Gnirke, A., Lindblad-Toh, K., Kellis, M. and Lander, E.S. Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. *Proc Natl Acad Sci U S A* **104** (2007), pp. 7145-50.
- Xu, Q., Li, M., Adams, J. and Cai, H.N. Nuclear location of a chromatin insulator in *Drosophila melanogaster*. *J Cell Sci* **117** (2004), pp. 1025-32.

- Yang, P., Zemba, M., Aboud, M., Flugel, R.M. and Lochelt, M. Deletion analysis of both the long terminal repeat and the internal promoters of the human foamy virus. *Virus Genes* **15** (1997), pp. 17-23.
- Yannaki, E., Tubb, J., Aker, M., Stamatoyannopoulos, G. and Emery, D.W. Topological constraints governing the use of the chicken HS4 chromatin insulator in oncoretrovirus vectors. *Mol Ther* **5** (2002), pp. 589-98.
- Yoon, Y.S., Jeong, S., Rong, Q., Park, K.Y., Chung, J.H. and Pfeifer, K. Analysis of the H19ICR insulator. *Mol Cell Biol* **27** (2007), pp. 3499-510.
- Yu, W., Ginja, V., Pant, V., Chernukhin, I., Whitehead, J., Docquier, F., Farrar, D., Tavoosidana, G., Mukhopadhyay, R., Kanduri, C., Oshimura, M., Feinberg, A.P., Lobanenko, V., Klenova, E. and Ohlsson, R. Poly(ADP-ribosyl)ation regulates CTCF-dependent chromatin insulation. *Nat Genet* **36** (2004), pp. 1105-10.
- Yusufzai, T.M., Tagami, H., Nakatani, Y. and Felsenfeld, G. CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol Cell* **13** (2004), pp. 291-8.
- Zampieri, M., Ciccarone, F., Palermo, R., Cialfi, S., Passananti, C., Chiaretti, S., Nocchia, D., Talora, C., Screpanti, I. and Caiafa, P. The epigenetic factor BORIS/CTCF regulates the NOTCH3 gene expression in cancer cells. *Biochim Biophys Acta* (2014).
- Zampieri, M., Guastafierro, T., Calabrese, R., Ciccarone, F., Bacalini, M.G., Reale, A., Perilli, M., Passananti, C. and Caiafa, P. ADP-ribose polymers localized on Ctfp-Parp1-Dnmt1 complex prevent methylation of Ctf target sites. *Biochem J* **441** (2011), pp. 645-52.
- Zhan, H.C., Liu, D.P. and Liang, C.C. Insulator: from chromatin domain boundary to gene regulation. *Hum Genet* **109** (2001), pp. 471-8.
- Zhang, H., Niu, B., Hu, J.F., Ge, S., Wang, H., Li, T., Ling, J., Steelman, B.N., Qian, G. and Hoffman, A.R. Interruption of intrachromosomal looping by CCCTC binding factor decoy proteins abrogates genomic imprinting of human insulin-like growth factor II. *J Cell Biol* **193** (2011), pp. 475-87.
- Zhang, R., Burke, L.J., Rasko, J.E., Lobanenko, V. and Renkawitz, R. Dynamic association of the mammalian insulator protein CTCF with centrosomes and the midbody. *Exp Cell Res* **294** (2004), pp. 86-93.
- Zhong, X.P. and Krangel, M.S. An enhancer-blocking element between alpha and delta gene segments within the human T cell receptor alpha/delta locus. *Proc Natl Acad Sci U S A* **94** (1997), pp. 5219-24.
- Zhou, J., Barolo, S., Szymanski, P. and Levine, M. The Fab-7 element of the bithorax complex attenuates enhancer-promoter interactions in the *Drosophila* embryo. *Genes Dev* **10** (1996), pp. 3195-201.
- Zhu, Q., Gregg, K., Lozano, M., Liu, J. and Dudley, J.P. CDP is a repressor of mouse mammary tumor virus expression in the mammary gland. *J Virol* **74** (2000), pp. 6348-57.
- Zhu, X., Ling, J., Zhang, L., Pi, W., Wu, M. and Tuan, D. A facilitated tracking and transcription mechanism of long-range enhancer function. *Nucleic Acids Res* **35** (2007), pp. 5532-44.

- Ziebarth, J.D., Bhattacharya, A. and Cui, Y. CTCFBSDB 2.0: a database for CTCF-binding sites and genome organization. *Nucleic Acids Res* **41** (2012), pp. D188-94.
- Zlatanova, J. and Caiafa, P. CCCTC-binding factor: to loop or to bridge. *Cell Mol Life Sci* **66** (2009), pp. 1647-60.
- Zsiros, J., Jebbink, M.F., Voute, P.A. and Berkhout, B. Identification of novel human endogenous retroviral sequences belonging to the HERV-K family. *AIDS Res Hum Retroviruses* **14** (1998), pp. 1093-8.



## Приложение 1

Последовательности олигонуклеотидов, использованных в работе.

Библиотечный праймер 27Pr		ACTGAGCTCGAGTATCCATGAACA
Адаптер <i>Sau3A</i>		GATCTGTTCATGG
Адаптер <i>Csp6-I</i>		TATGTTCATGG
5'-фланк LTR	Pr1	CAGTCTTATCTCCTTTACTGACC
3'-фланк LTR	Pr2	CCTCGTGTTTGTGCTTG
в LTR	Pr3	AAAGACACAGAGACAAAGTATAGA
Праймеры для плазмид на основе pGL3		
	GL2	CTTTATGTTTTTGGCGTCTTCCA
	RV3	CTAGCAAATAGGCTGTCCC
	RV4	GACGATAGTCATGCCCCGCG
Праймеры для плазмид на основе pGTN8		
	FRToutL	GGATTTCCAAGTCTCCAGGGGAT
	FRToutR	GGATTTCCAAGTCTCCAGGGGAT
Внутренние праймеры LTRL47334		
	590r(ss)	GCCATATTTTCAGACTATCACATGG
	Pr4 (ss)	ATTGTCCAAGGTTTCTCCC
	577for (ss)	GCCTTAGGGCTGGAGGTG
	725R (ss)	CAGCAGACAAACACGTGAAC
	Pr7 (ss)	GGGTCCCCTTATTTCTTTCTC
Фланкирующие праймеры LTR L47334		
	Pr1 (ss)	CAGTCTTATCTCCTTTACTGACC
	Pr2 (ss)	CCTCGTGTTTGTGCTTG
Двухцепочечный олигонуклеотид специфичный к ERLBF (C1/C2)		
	C1	AGGATCGGGGCGGGGCGATC
	C2	CTGATCGCCCCGCCCCGATC
Промотор <i>c-myc</i>		
<i>c-myc</i> for		GGGATCGCGCTGAGTATAAA
<i>c-myc</i> rev		GGATCTCCCTTCCCAGGAC
Последовательности олигонуклеотидов, предназначенных для введения полилинкера (MCS) по сайту Cla I		
MCS for		CGATCTAGACTCGAGAATTC
MCS rev		ATCGGAATTCTCGAGTCTAG'
Библиотечный праймер LP		
1L		CTTGAACCTCCTCGTTCGAC
1R		TGGCTACCCGTGATATTGCT
2L		GGATTTCCAAGTCTCCAGGGGAT
2R		ACCTCCCACCGTACACGCCT
CTCF-1	1for	GTTTGTGACCTGTGCCCTTT
	1rev	GAGGCCCGACTCTTAACTC
CTCF-2	2for	AGGTCCCTTCTCTCCCTGCT
CTCF-3	3for	AATGATATTCCTCACGGCACT
CTCF-4	4for	GCTCTGGGAAGAAACCACAG

	4rev	GCAGGAGCAAGGTGAGATG
CTCF-5	5for	GCTTTCTGACCCGCCTTT
CTCF-6	6for	ATAGAGAAGCAGGGGGTGTG
	6rev	TGCTGTTCCGTAATAACTTGCT
CTCF-7	7for	САСТААТGAGAGACGCTGAGGA
	7rev	GCTTCTGGAGGGTGTTTCTG
CTCF-8	8for	CCTTGACTTTGATGTAGAAAGGAA
	8rev	CACCGTCCTCTGCCAACT
CTCF-9	9rev	AAGGCACTGGCATCCTGTCT
CTCF-10	10for	GTACAGCCCTGGAGCAAGGAC
sns	snsL	ACTCGCAAACCTCAACACCT
	snsR	CAAAACTGGAATGGGGAAGA
Плазмидные праймеры	3L	CCGGACGAACTAAACCTGAC
	3R	TGTAGGTA CTCTGTTCTCACCTTC

## Благодарности

Хочу выразить искреннюю признательность научному консультанту академику Е.Д. Свердлову за советы и помощь на всех этапах работы. Сердечно благодарю в.н.с. Г.С. Монастырскую за ценные консультации, помощь и повседневную заботу. Искренне благодарю сотрудников лаборатории структуры и функций генов человека ИБХ РАН и аспирантов разных лет, в непосредственном сотрудничестве с которыми были получены основные результаты работы - Л.Г. Николаева, И.П. Чернова, Е.В. Снежкова, Л.Л. Завалову, Д.О. Трубецкого, Т.В. Виноградову, Е.П. Копанцева, А.В. Сасса, В.М. Руда, А.С. Ветчинову, Д.А. Дидыча, Т.Л. Ажикину, Ю.Б. Лебедева, Е.С. Котову, Н.А. Смирнова, С.С. Буланенкову, М.Б. Костину, Цогтхишиг Цэвэгийн. Выражаю искреннюю благодарность Е.А. Стукачевой и Е.В. Свирцевской за творческое сотрудничество. Огромное спасибо всем сотрудникам лаборатории структуры и функций генов человека за дружеское и внимательное отношение. Без их советов и дружеского участия в работе ее выполнение было бы невозможно. За проведенные совместные исследования выражаю искреннюю благодарность О.В. Яровой и С.В. Разину (ИБГ РАН), Юргену Боду (DFG, Германия), а также Мари Хенриксон (Каролинский Институт, Швеция).

На разных этапах работа финансировалась грантами РФФИ, программой поддержки научных школ, российской программой "Геном человека", программой по молекулярной и клеточной биологии РАН, за что выражаю им благодарность.