

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

---

*На правах рукописи*

Смирнов Николай Андреевич

**ИССЛЕДОВАНИЕ АКТИВНОСТИ ПОТЕНЦИАЛЬНЫХ  
ИНСУЛЯТОРНЫХ И ЭНХАНСЕРНЫХ ЭЛЕМЕНТОВ ГЕНОМА  
ЧЕЛОВЕКА**

03.01.03 – Молекулярная биология

Диссертация на соискание ученой степени

кандидата биологических наук

Научный руководитель:  
доктор биологических наук

Акопов С.Б.

Москва – 2017

# СОДЕРЖАНИЕ

<b>СПИСОК СОКРАЩЕНИЙ</b> .....	4
<b>1. ВВЕДЕНИЕ</b> .....	7
<b>2. ОБЗОР ЛИТЕРАТУРЫ</b> .....	11
<b>2.1. ИНСУЛЯТОРЫ</b> .....	11
2.1.1.    СТCF – многофункциональный фактор .....	12
2.1.2.    Картирование сайтов связывания инсуляторных белков позвоночных.....	14
2.1.3.    СТCF и белки когезинового комплекса .....	15
2.1.4.    5'-сHS4 инсулятор из бета-глобинового локуса кур.....	17
2.1.5.    Механизмы функционирования инсуляторов .....	19
2.1.6.    Функциональные методы изучения инсуляторов .....	22
<b>2.2. ЭНХАНСЕРЫ</b> .....	23
2.2.1.    Структура энхансеров.....	24
2.2.2.    Функциональные модели активации транскрипции энхансерами .....	29
2.2.3.    Механизмы действия энхансеров .....	30
2.2.4.    Регуляция активности энхансеров.....	33
2.2.5.    Идентификация и функциональный анализ энхансеров .....	36
<b>3. МАТЕРИАЛЫ И МЕТОДЫ</b> .....	42
<b>3.1. МАТЕРИАЛЫ</b> .....	42
<b>3.2. МЕТОДЫ</b> .....	51
3.2.1.    Стандартные методики. ....	51
3.2.2.    ПЦР-амплификация фрагментов ДНК. ....	51
3.2.3.    Трансформация клеток <i>E. coli</i> . ....	53
3.2.4.    5'-RACE.....	54
3.2.5.    Подготовка плазмид для секвенирования.....	55
3.2.6.    Электрофорез в агарозном геле .....	55
3.2.7.    Электрофорез в полиакриламидном геле (очистка меченой ДНК) .....	55
3.2.8.    Связывание меченых фрагментов ДНК с белками <i>in vitro</i> .....	56
3.2.9.    Сдвиг электрофоретической подвижности в полиакриламидном геле (EMSA).....	56
3.2.10.   Использованные культуры клеток.....	57
3.2.11.   Трансфекция клеток с использованием реагента Lipofectamine 2000.....	58
3.2.12.   Получение клеточного лизата и измерение активности люциферазы .....	59
3.2.13.   3С анализ.....	59
3.2.14.   Определение количества трансфицированных плазмид в ядрах.....	62
<b>4. РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ</b> .....	63

<b>4.1. ФУНКЦИОНАЛЬНАЯ ДИСЕКЦИЯ ЭНХАНСЕРНОГО ЭЛЕМЕНТА, РАСПОЛОЖЕННОГО ВО ВТОРОМ ИНТРОНЕ ГЕНА <i>U2AF1L4</i> ЧЕЛОВЕКА.....</b>	<b>63</b>
4.1.1. Анализ активности фрагментов энхансера 12 .....	64
4.1.2. Анализ функциональной консервативности энхансера 12.....	67
<b>4.2. ИЗУЧЕНИЕ ДВУНАПРАВЛЕННОГО ПРОМОТОРА ГЕНОВ <i>PSENNEN</i> И <i>U2AF1L4</i> ..</b>	<b>73</b>
4.2.1. Точки инициации транскрипции и тканевая специфичность транскрипции генов <i>PSENNEN</i> и <i>U2AF1L4</i> .....	73
4.2.2. Анализ активности промоторной области генов <i>PSENNEN</i> и <i>U2AF1L4</i> .....	76
4.2.3. Функциональный анализ консервативных областей промотора генов <i>PSENNEN</i> и <i>U2AF1L4</i> .....	79
<b>4.3. ОЦЕНКА ЭНХАНСЕР-БЛОКИРУЮЩЕЙ АКТИВНОСТИ ИНСУЛЯТОРОВ КАРТИРОВАННЫХ В ЛОКУСЕ <i>FXYD5-COX7A1</i> ХРОМОСОМЫ 19 ЧЕЛОВЕКА .....</b>	<b>84</b>
4.3.1. Конструирование плазмид.....	84
4.3.2. Выбор фрагментов ДНК .....	85
4.3.3. Энхансерная активность геномных фрагментов .....	89
4.3.4. Энхансер-блокирующая активность инсулятора сHS4 из бета-глобинового локуса кур .....	90
4.3.5. Энхансер-блокирующая активность геномных фрагментов.....	92
4.3.6. Энхансер-блокирующая активность фрагментов, отобранных при помощи негативно-позитивной селекции.....	94
<b>4.4. АКТИВАЦИЯ ПРОМОТОРА ЭНХАНСЕРОМ В <i>ТРАНС</i> ПОЛОЖЕНИИ .....</b>	<b>96</b>
4.4.1. Конструирование плазмид.....	96
4.4.2. Активация промотора энхансером в <i>транс</i> .....	98
4.4.3. Промоторная и клеточная специфичность.....	99
4.4.4. Взаимодействие промотора с энхансером .....	101
<b>5. ЗАКЛЮЧЕНИЕ И ВЫВОДЫ.....</b>	<b>105</b>
<b>6. СПИСОК ЦИТИРОВАННОЙ ЛИТЕРАТУРЫ.....</b>	<b>106</b>

## СПИСОК СОКРАЩЕНИЙ

- 3C** (chromosome conformation capture) -фиксация конформации хромосомы;
- 5'-RACE** (Rapid amplification of cDNA ends) – метод быстрой амплификации 5' концов кДНК;
- AP-1** (Activator protein 1) – активаторный белок 1;
- AP-2** (Activating Protein 2) – активаторный белок 2;
- ATF** (Activating transcription factor) – активаторный транскрипционный фактор;
- BAF155** (BRG1-Associated Factor 155) – BRG1 ассоциированный фактор с молекулярной массой 155 кДа;
- BAF170** (BRG1-Associated Factor 170) – BRG1 ассоциированный фактор с молекулярной массой 170 кДа;
- ChIP** (Chromatin Immunoprecipitation) – иммунопреципитация хроматина;
- ChIP-chip** (Chromatin Immunoprecipitation on chip) – иммунопреципитация хроматина с использованием чипов для анализа иммунопреципитированной ДНК;
- ChIP-seq** (Chromatin immunoprecipitation sequencing) – иммунопреципитация хроматина с последующим масштабным секвенированием;
- ChIA-PET** (Chromatin Interaction Analysis by Paired-End Tag Sequencing) – метод выявления хроматиновых взаимодействий;
- CMV** (Cytomegalovirus) – цитомегаловирус;
- CNEs** (Conserved non-coding elements) – консервативные некодирующие элементы;
- CREB** (cAMP response element binding protein) – цАМФ-зависимый транскрипционный фактор;
- CBP** (CREB-binding protein) – CREB-связывающий белок;
- CTCF** (CCCTC binding factor) – фактор, связывающийся с последовательностью CCCTC;
- EKLF** (Erythroid Krüppel-like factor) – эритроидный Круппель-подобный фактор;

**EMSA** (Electrophoretic Mobility Shift Assay) – метод сдвига электрофоретической подвижности;

**ENCODE** (The Encyclopedia of DNA Elements) – энциклопедия регуляторных ДНК элементов;

**ETS** (Extra TFIIIC sites) – дополнительные сайты связывания фактора TFIIIC;

**FBS** (Fetal bovine serum) – эмбриональная телячья сыворотка;

**GATA1** (GATA-binding factor 1) – фактор, связывающийся с последовательностью GATA;

**GABP** (GA-binding protein transcription factor) – транскрипционный фактор, связывающийся с последовательностью GA;

**HEPES** – 4-(2-гидроксиэтил)-1-пиперазин этансульфоновая кислота;

**HNF-4 $\alpha$**  (hepatocyte nuclear factor 4 $\alpha$ ) – гепатацитарный ядерный фактор 4 $\alpha$ ;

**HS** (Hyper sensitive) – гиперчувствительный;

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

**ICR** (imprinting control region) – регуляторный элемент, контролирующий геномный импринтинг;

**Ini1** (Integrase Interactor 1) – интерактор 1 интегразы;

**IPTG** – изопропил  $\beta$ -D-1-тиогалактопиранозид;

**IRF** (interferon-regulatory factor) – регуляторный фактор интерферона;

**LCR** (Locus Control Region) – локус-контролирующая область;

**MIR** (Mammalian-wide interspersed repeat) – одно из семейств SINE;

**NF- $\kappa$ B** (Nuclear factor-kappaB) – ядерный фактор каппа B;

**NRF1** (Nuclear respiratory factor 1) – ядерный белок клеточного дыхания 1;

**PBS** (Phosphate Buffered Saline) – фосфатно-солевой буфер;

**PIC** (preinitiation complex) – преинициаторный комплекс;

**PolII** (RNA polymerase II) – РНК полимераза II;

**PRD** (positive regulatory domains) – позитивный регуляторный домен;

**RLT** (Random ligation template) - смесь продуктов лигирования;

**SATB-1** (Special adenine-thymine (AT)-rich DNA-binding protein) – фактор, связывающийся с последовательностью богатой АТ;

**SDS** (Sodium dodecyl sulfate) – натрия додецилсульфат;

**SINE** (Short interspersed element) – короткий диспергированный повтор;

**SP 1** (Specificity protein 1) – специфичный белок 1;

**SV40** (Semian Virus 40) – обезьяний вирус 40;

**TAF** (Transcription Activation Factor) – фактор активации транскрипции;

**TAE** – Трис-ацетатный буфер;

**TBE** – Трис-боратный буфер с добавлением ЭДТА;

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

**TEMED** – N,N,N',N'-тетраметилэтилендиамин;

**TE** – трис-ЭДТА буфер;

**TFIID** (Transcription factor II D) – транскрипционный фактор II D;

**TFIIIC** (Transcription Factor for polymerase III C) – транскрипционный фактор III C;

**TSS** (Transcription start site) - точка инициации транскрипции;

**VEZF1** (Vascular Endothelial Zinc Finger 1) – белок с цинковыми пальцами 1 эндотелия сосудов;

**UAS** (Upstream Activating Sequence) – дистальная активирующая последовательность;

**X-gal** – 5-бром-4-хлоро-3-индолил-бета-D-галактопиранозид;

**ZNF143** (Zinc Finger Protein 143) – белок с цинковыми пальцами 143;

**ДТТ** - дитиотрейтол;

**мяРНК** – Малые ядерные РНК;

**ОТ-ПЦР**– полимеразная цепная реакция с обратной транскрипцией;

**ПЦР** – полимеразная цепная реакция;

**п.н.** - пар нуклеотидов;

**т.п.н.** – тысяч пар нуклеотидов;

**ЭДТА** – этилендиаминтетрауксусная кислота;

**ЯЭ** – ядерный экстракт.

## 1. ВВЕДЕНИЕ

Регуляция транскрипции является процессом, играющим основополагающую роль в развитии организмов, так как функциональность генов определяется не только их продуктами, но, также, специфическим паттерном их экспрессии. Гены транскрибируются на определенном уровне для того чтобы обеспечить надлежащее протекание онтогенетических процессов. Большое количество элементов вовлечено в процесс регуляции как на транскрипционном, так и на пост-транскрипционном уровне. Элементы, которые вовлечены в регуляцию генов на транскрипционном уровне, включают в себя промоторы, энхансеры, сайленсеры, инсуляторы. В настоящее время установлено, что сложные функциональные и пространственные взаимосвязи между геномными регуляторными элементами, в конечном счете, определяют функциональную идентичность клетки [1, 2].

Исследования генома, проводившиеся в последние 20 лет, в основном были нацелены на интенсивное изучение и идентификацию белок-кодирующих генов и только отчасти на изучение сложной регуляторной сети, обеспечивающей их экспрессию. Было определено, что лишь около 1,2 % генома человека кодирует белки [3]. Использование методов сравнительной геномики показало, что большая часть консервативных и недавно приобретенных в ходе эволюции участков генома представлено некодирующими последовательностями, играющими структурную и/или функциональную роль [4-8]. Недавние полногеномные исследования показали, что некодирующие элементы также играют важнейшую роль в предрасположенности к возникновению заболеваний [9, 10].

Несмотря на важную роль, которую играют регуляторные элементы в экспрессии генов, степень изученности их функций в данный момент является недостаточной и большая часть регуляторных элементов не охарактеризована. Основной причиной этого является то, что количество регуляторных

элементов в несколько раз превышает количество генов, а также, то, что появление методик для высокоэффективного полногеномного картирования регуляторных элементов стало возможным относительно недавно, в то время как средне- и высокоэффективные подходы для картирования генов, и промоторов являются отработанными и широко распространенными.

К сожалению, функциональный анализ регуляторных элементов на уровне генома сопряжен с рядом технических трудностей и несмотря на то, что использование полногеномных методов позволило накопить огромные массивы данных, следует учитывать, что картирование огромного числа функциональных элементов, выполненное с применением методик, основанных на секвенировании нового поколения, сопряжено с большим количеством ошибок. Поэтому исчерпывающий функциональный анализ регуляторных элементов на уровне отдельных сопряженных сегментов с последующим интегрированием полученных данных представляется нам более предпочтительным.

Целью диссертационной работы является функциональный анализ *cis*-регуляторных элементов (энхансеры, инсуляторы) в локусе *FXYD5-COX7A1* хромосомы 19 человека. Регуляторные элементы для анализа были выбраны на основании данных ChIP-seq проекта ENCODE (Encyclopedia of DNA Elements), а также из элементов, идентифицированных ранее с использованием метода позитивно-негативной селекции [11, 12]. Был проведен также детальный функциональный анализ двунаправленного промотора генов *PSENFEN* и *U2AF1L4* и энхансера 12, расположенного во втором интроне гена *U2AF1L4*, ранее идентифицированного в нашей лаборатории [13]. Проанализировать эффект *транс*-активации промотора энхансером с использованием системы экспрессии репортерного гена.

В ходе работы были поставлены и решены следующие задачи:

Используя методы сравнительной геномики, выявить консервативные области в энхансере 12 и разделить последовательность энхансера на перекрывающиеся фрагменты для их дальнейшего изучения. Для



идентификации области, отвечающей за энхансерную активность, провести функциональный анализ полученных перекрывающихся фрагментов энхансера 12, используя систему экспрессии репортерного гена. С помощью метода сдвига электрофоретической подвижности EMSA (electrophoretic mobility shift assay) провести функциональный анализ ортологичных последовательностей приматов для подтверждения функциональной роли консервативных областей энхансера 12.

Используя методику 5'-RACE (Rapid amplification of cDNA ends), определить точки начала транскрипции промотора генов *PSENE1* и *U2AF1L4*, расположенного в непосредственной близости от энхансера 12. Используя метод экспрессии репортерного гена, провести анализ активности промотора генов *PSENE1* и *U2AF1L4*.

С помощью количественной ОТ-ПЦР (полимеразная цепная реакция с обратной транскрипцией) определить уровень и тканеспецифичность экспрессии генов *PSENE1* и *U2AF1L4*. Определить консервативные области промотора генов *PSENE1* и *U2AF1L4*, используя методы сравнительной геномики, и с помощью сайт-специфического мутагенеза выявить области, являющиеся функциональными.

Создать систему, предназначенную для анализа энхансер-блокирующей активности фрагментов ДНК *in vitro*, в том числе плазмиду pGL4EPV, содержащую в регуляторной области гена люциферазы светлячка энхансер и промотор вируса SV40.

Клонировать в плазмиду pGL4PV, содержащую в регуляторной области гена люциферазы светлячка промотор вируса SV40, десять последовательностей потенциальных энхансеров, выбранных на основании данных ChIP-seq проекта ENCODE. Полученными векторными конструкциями провести транзистентные трансфекции клеточных линий HeLa и HepG2 с последующим анализом энхансерной активности анализируемых фрагментов.

Клонировать в вектор pGL4EPV между энхансером и минимальным промотором SV40 последовательности четырех потенциальных инсуляторов, обнаруженных нами ранее методом позитивно-негативной селекции в локусе 19q13.12, и 5 последовательностей, выбранных на основании данных ChIP-seq проекта ENCODE, а также контрольную последовательность 5'-HS4 инсулятора кур. Полученными векторными конструкциями провести транзистентные трансфекции клеточных линий HeLa, CHO и HepG2 с последующим анализом энхансер-блокирующей активности анализируемых фрагментов.

Создать систему, предназначенную для анализа промоторной и клеточной специфичности активации промотора энхансером в *транс*-положении при транзистентных трансфекциях. С использованием полученной системы изучить способность клеточных и вирусных энхансеров активировать различные промоторы в *транс*-положении. Используя метод фиксации конформации хромомсом 3C (chromosome conformation capture), продемонстрировать физическое сближение промотора с энхансером при *транс*-активации.

## 2. ОБЗОР ЛИТЕРАТУРЫ

### 2.1. ИНСУЛЯТОРЫ

Инсуляторы – это последовательности ДНК, нарушающие взаимодействие энхансера с промотором или выполняющие барьерную функцию. Некоторые авторы расширяют определение инсуляторов и рассматривают их как ДНК-белковые комплексы, опосредующие взаимодействие различных участков генома [14, 15]. Параллельное исследование инсуляторов в дрожжах, плодовой мушке *Drosophila Melanogaster* и позвоночных показало их высококонсервативную роль в регуляции экспрессии генов и пространственной организации генома [16].

Основным инсуляторным белком позвоночных, отвечающим за активность инсуляторов, в настоящее время считается CTCF (CCCTC binding factor) [17]. Инсуляторная функция CTCF было доказана в модельных генетических системах, содержащих репортерный ген [18], и в нативном геномном окружении [19, 20]. Относительно недавно для млекопитающих также была показана инсуляторная активность генов тРНК и ТФПС-связывающих сайтов [21], в которых транскрипционный фактор ТФПС часто колокализуется с белками когезинового комплекса и инсуляторным белком CTCF [22]. Картирование сайтов связывания ТФПС, не относящихся к генам тРНК, так называемых ETC (Extra ТФПС sites) сайтов в клетках линии K562 показало корреляцию с сайтами связывания CTCF, данный факт подтверждает регуляторную роль ТФПС [23].

Молекулярные механизмы, при помощи которых инсуляторы осуществляют присущее им многообразие функций остаются до конца не выясненными. Считается, что данные функции обусловлены специфическими белками, которые связываются с инсуляторными последовательностями и с дополнительными факторами. Образующиеся в результате комплексы взаимодействуют с подобными им комплексами, приводя к сближению, и контакту линейно отдаленных друг от друга участков генома, или возникает физический барьер, разграничивающий два участка ДНК с разными

профилями регуляции. В результате модификации этих двух активностей могут проявляться следующие функции инсуляторов:

1. Блокирование энхансеров [18, 24, 25].
2. Привлечение энхансеров к промоторам [14, 26].
3. Разделение областей генома с различным эпигенетическим состоянием и изоляция друг от друга генов, регулируемых по разным механизмам [24, 27].
4. Организация системы совместно регулируемых и экспрессируемых генов в результате их сближения [28].
5. Регуляция рекомбинации (V(D)J-рекомбинация в локусе IgH) [14, 29].
6. Поддержание состояния репрессированного хроматина с метками H3K27me3 [30].

К настоящему времени было охарактеризовано лишь небольшое количество инсуляторов. Это связано с тем, что для идентификации и функционального изучения этих элементов необходимо проводить детальные исследования с использованием прямых экспериментальных тестов для регистрации их энхансер-блокирующей и барьерной активности. Без этой информации оценить возможность применения инсуляторов в терапевтических генно-инженерных конструкциях затруднительно.

### **2.1.1. CTCF – многофункциональный фактор**

Впервые CTCF был обнаружен как транскрипционный фактор, узнающий СТ - богатую последовательность в промоторе *c-myc* гена кур.[31]. Последующие исследования показали, что CTCF может узнавать и связывать множество различных последовательностей [32], данные сайты были названы CTCF – связывающими сайтами (CTSs).

CTCF содержит фланкированный N- и C-концевым доменами, обогащенными лизином и аргинином, эволюционно консервативный центральный ДНК-связывающий домен, состоящий из 11 цинковых пальцев

[32]. Первичная последовательность CTCF является консервативной, что демонстрирует функциональную значимость данного белка [33].

Было показано, что CTCF в клетке вовлечен в такие процессы как канцерогенез [34], регуляция эмбрионального развития [35], дифференцировка клеток [29, 36-38], апоптоз [39], импринтинг [40], подавление экспрессии генов [41] и инактивация X – хромосомы [42]. Было высказано предположение, что многофункциональность CTCF основана на связывании его с различными консенсусными последовательностями в геноме и последующими процессами, связанными с привлечением разнообразных кофакторов и посттрансляционных модификаций [33].

В связи с множеством различных эффектов, которые CTCF оказывает на транскрипционную активность различных генов, можно предложить несколько моделей описывающих процессы, в которых участвует CTCF [43-45].

Первая модель основана на данных полученных при изучении регуляции локусов генов *Igf2-H19* [46] и *MHC-II* [47], в данной модели CTCF-зависимые инсуляторы опосредуют дистантные взаимодействия, которые в свою очередь могут влиять на эпигенетическое состояние ключевых регуляторных элементов и уровень транскрипции. В данной модели функция CTCF также сопряжена со связыванием в инсуляторах таких белковых факторов, как деацетилаза гистонов SIN3 [48], рецептор тиреоидного гормона [49], нуклеофосмин [50], Kaiso [51] и содержащая DEAD-бокс РНК хеликаза р68 в комплексе с некодирующими РНК (ncRNA) [52]. Исследования показали, что белковый комплекс р68 необходим для загрузки когезинового комплекса в области, контролирующей импринтинг, локуса генов *Igf2* и *H19* [52]. Также было обнаружено, что CTCF совместно занимает сайты с транскрипционным фактором FOXA1 и рецептором эстрогена (ER). Такие сайты имеют тенденцию к расположению вблизи генов контролируемых эстрогеновым рецептором, предполагается, что CTCF может опосредовать активацию этих генов [53]. Более того, CTCF в эмбриональных стволовых клетках привлекает

в межгенные сайты основной транскрипционный фактор TAF3, где TAF3-зависимое выпетливание приводит к активации транскрипции генов [54].

Вторая модель основана на предположении, что CTCF может управлять организацией топологически связанных доменов хроматина, путем образования петель или хабов (узлов), которые ограничивают доступ транскрипционного аппарата [16]. Модель активного хроматинового хаба постулирует, что дистантные регуляторные элементы (энхансеры) и промоторы целевых генов организованы в общий комплекс [55]. Идея активного хроматинового хаба была выдвинута для объяснения физических взаимодействий в ядре между активными генами  $\beta$ -глобинового локуса и регуляторными элементами [56]. Роль CTCF в данной модели заключается в регулировании дистантных взаимодействий и эпигенетического состояния хроматина [57]. Было показано, что деплеция CTCF влияет на образование петлевых доменов между сайтами связывания CTCF, а также нарушает изолированность между соседними топологически связанными доменами [58]

### **2.1.2. Картирование сайтов связывания инсуляторных белков позвоночных**

Помимо того, что CTCF имеет повсеместный паттерн экспрессии, его сайты связывания широко представлены в геноме позвоночных. Для того чтобы лучше понять его функции было проведено картирование сайтов связывания в геноме млекопитающих.

Биоинформатический анализ консервативных последовательностей генома человека выявил примерно 15000 потенциальных CTCF-связывающих последовательностей [59]. Использование иммунопреципитации хроматина с последующей гибридизацией на чипах (ChIP-chip), позволило идентифицировать 13804 CTCF-связывающих сайта в линии фибробластов человека IMR90 [60]. Использование иммунопреципитации хроматина с последующим массивным секвенированием позволило идентифицировать в CD4<sup>+</sup> Т-клетках 26814 CTCF-связывающих сайта [61].

Анализ полученных данных показал, что в геноме человека значительная часть (60%) обнаруженных сайтов CTCF, располагается на значительном расстоянии от точек начала транскрипции (TSS) (среднее расстояние до ближайшего промотора - 48000 п.н.), 30% сайтов CTCF было обнаружено внутри интронов, 10% – внутри экзонов, и лишь незначительная доля обнаруженных сайтов располагается вблизи 3'-концов генов. [59, 60, 62, 63]. Большая часть CTCF-связывающих сайтов демонстрирует выраженную тканеспецифичность, при этом для 41% сайтов была показана связь с дифференциальным метилированием [63, 64]. Данные результаты показывают, что связывание CTCF является регулируемым процессом, который вовлечен в клеточную дифференцировку.

Детальное сравнение ~326840 CTCF-связывающих сайтов в 38 клеточных линиях человека показало, что большая часть сайтов являются консервативными, ~126200 сайтов оказались тканеспецифичными [62]. Наибольшее связывание CTCF было выявлено на границах репрессированных хроматиновых доменов, имеющих характерную H3K27me3 модификацию гистоновых белков, при этом было обнаружено, что связывание CTCF с границами репрессированных доменов также является тканеспецифичным [61].

Анализ конститутивных CTCF-связывающих сайтов в 56 клеточных линиях показал, что такие сайты расположены в зонах открытого хроматина и колокализуются с белками когезинового комплекса [64]. Такое распределение сайтов связывания CTCF отличается от распределения сайтов связывания других известных факторов транскрипции и согласуется с представлениями о роли CTCF как основного инсуляторного белка позвоночных, который участвует в образовании топологически связанных доменов.

### **2.1.3. CTCF и белки когезинового комплекса**

Когезин является основным белком, который участвует в разделении сестринских хроматид в течение S, G2 и M фазы клеточного цикла.

Когезиновый комплекс состоит из белков семейства smc (structure maintenance of chromosomes). Комплекс имеет кольцеобразную структуру, которая удерживает две хроматиды вместе [65].

Было обнаружено, что сильные сайты связывания когезина перекрывается с сайтами связывания CTCF [64, 66, 67], слабые же сайты связывания расположены в непосредственной близости от активных промоторов и энхансеров [68, 69]. Связывание когезинового комплекса с этими сайтами зависит от присутствия в них CTCF. В отсутствие CTCF когезин сохраняет способность связываться с ДНК, но при этом специфическое связывание с сайтами связывания CTCF пропадает [66]. Исследование мутантных форм CTCF показало, что за специфичность взаимодействия с CTCF-связывающими сайтами может отвечать С-концевой участок CTCF, который опосредует это взаимодействие с Scc3 субъединицей когезина. Удаление этого участка приводит к потере взаимодействия с когезином и исчезновению энхансер-блокирующей функции CTCF в локусе генов *H19-Igf2* [70].

Ряд исследований различных генов подтвердил функциональную роль когезинового комплекса в регуляции экспрессии генов. Так, исследования локуса генов *H19-Igf2* показали, что когезиновый комплекс участвует в образовании CTCF опосредованных хроматиновых петель, которые обеспечивают правильную транскрипцию гена *Igf2* [71]. Также регуляторная роль когезина была показана для локуса гена  $\gamma$ -интерферона (*IFNG*), где деплеция когезина приводила к разрушению петель хроматина между регуляторными элементами и понижению экспрессии гена *IFNG* [72]. В локусе  $\alpha$ -кадгеринов CTCF и когезин опосредуют возникновение случайного взаимодействия энхансера с промоторами переменных экзонов [36], механизм данного взаимодействия аналогичен функции CTCF при альтернативном сплайсинге в клетках иммунной системы [73]. В  $\beta$ -глобиновом локусе когезин участвует в образовании устойчивого



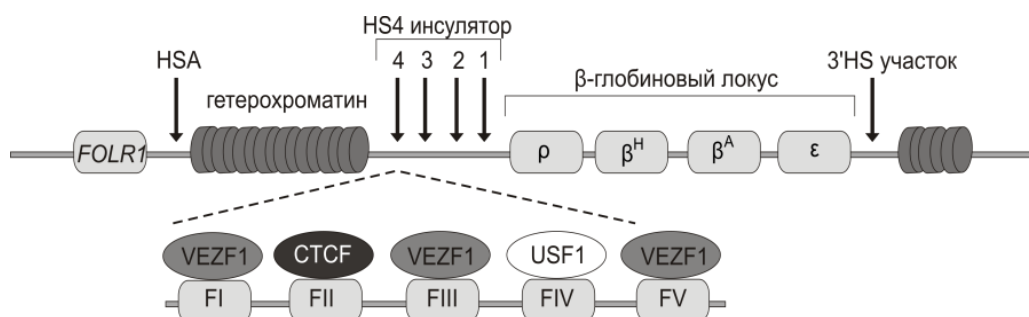
взаимодействия между энхансером локус контролирующей области (ЛКР) и генами  $\beta$ -глобинов, расположенных ниже [74].

Сравнение CTCF – независимых сайтов связывания когезина в двух типах клеток показало их клеточную специфичность. В таких сайтах когезин часто колокализуется с медиатором и РНК полимеразой II, что свидетельствует об энхансерной функции таких последовательностей. Такие сайты часто расположены вблизи активно транскрибируемых генов и связаны с тканеспецифичными транскрипционными факторами [68, 69]. Приведенные данные показывают, что CTCF и когезин выполняют как общие, так и независимые функции в регуляторных последовательностях генома.

#### **2.1.4. 5'-cHS4 инсулятор из бета-глобинового локуса кур**

Первым охарактеризованным в геноме позвоночных элементом, обладающим инсуляторными свойствами, был участок длиной 1,2 т.п.н., расположенный в 5'-области  $\beta$ -глобинового локуса кур. Локус  $\beta$ -глобиновых генов кур содержит четыре гена, которые экспрессируются на разных стадиях развития организма (рисунок 1). В локусе были обнаружены несколько специфичных для клеток эритроидного ряда участков (HS-сайтов), чувствительных к действию ДНКазы I. HS-сайт, расположенный в 5'-области локуса, обладал конститутивной чувствительностью к действию ДНКазы I [75]. Анализ структуры хроматина в непосредственной близости от этого конститутивного HS-сайта, выявил участок открытого хроматина длиной 33 т.п.н., включающий активный в эритроидных клетках  $\beta$ -глобиновый локус. Выше HS-сайта был выявлен участок гетерохроматина длиной 16 т.п.н. с геном фолатного рецептора [76]. Конститутивный HS-участок длиной 1,200 п.н., выполняющий барьерную функцию между конденсированным и активным хроматином был назван HS4-инсулятором. Данный инсулятор содержит коровый участок длиной 250 п.н., который обладает энхансер-блокирующей и барьерной активностями, и содержит пять белок-связывающих фрагментов (FI-FV). FII фрагмент представляет собой участок,

взаимодействующий с белком CTCF [77]. Удаление данного сайта приводит к потере энхансер–блокирующих свойств и не влияет на проявление барьерной активности [78]. За проявление барьерной активности инсулятора HS4 отвечают два белка – USF1, который связывается с FIV сайтом, и VEZF1, связывающийся с FI, FIII и FV сайтами (рисунок 1). Фактор USF1 привлекает модифицирующие хроматин факторы SET1 и NURF, которые специфически метилируют и ацетилируют гистоны и позиционируют нуклеосомы, что обеспечивает барьерную активность инсулятора [79, 80]. Удаление сайта связывания USF приводит к исчезновению барьерной активности инсулятора HS4 [81]. Фактор VEZF1 ограничивает распространение метилирования ДНК, защищая регуляторные последовательности и гены локуса от метилирования [82]. Мутации затрагивающие сайты связывания VEZF1 белка проявляются в изменении состояния метилирования как самого инсулятора, так и промотора одного из глобиновых генов, что отражается на уровне экспрессии этого гена [82]. Таким образом, координированное взаимодействие этих факторов, приводит к появлению барьерной активности у инсулятора HS4.



**Рисунок 1.** Куриный 5'-сHS4-инсулятор. Участок гетерохроматина длиной 16 т.п.н. расположен между геном фолатного рецептора1 (*FOLR1*) и бета–глобиновым локусом. HS4-инсулятор ограничивает гетерохроматин от бета–глобинового локуса. Инсулятор включает 5 сайтов (FI-FV) связывающих белки VEZF1, CTCF, USF1. CTCF-необходим для проявления энхансер-блокирующей активности и для привязывания инсулятора к ядрышку, USF1 и VEZF1 необходимы для проявления барьерной активности. HSA и 3'HS сайты проявляют энхансер-блокирующую активность.[83]

Хотя белку CTCF в данном инсуляторе отведена роль энхансер–блокирующего фактора, в других CTCF-связывающих инсуляторах он может

обеспечивать барьерную активность, защищая трансгены от эффекта положения [84]. Более того CTCF фланкирует домены связанные с ядерными белками ламинами и в некоторых случаях разделяет участки гетерохроматина – области с повышенным содержанием H3K27me3, от участков активного хроматина, содержащих маркер H3K5ac, что подтверждает роль CTCF в обеспечении барьерной активности инсуляторных элементов [85].

### **2.1.5. Механизмы функционирования инсуляторов**

Для объяснения функционирования инсуляторов было предложено три различные модели.

Согласно первой модели, энхансеры представляют собой, регуляторные элементы, привлекающие регуляторные комплексы, которые затем следуют вдоль хромосомы к промотору [86]. Данная модель основана на том, что активация и репрессия хроматина имеет процессивный характер [87]. Инсуляторы в данной модели выполняют функцию барьерных элементов, блокирующих продвижение регуляторных комплексов вдоль хроматиновой фибриллы.

Барьерная активность может быть, как пассивной, так и активной.

Пассивная барьерная активность осуществляется путем сборки на инсуляторном элементе большого белкового комплекса, который нарушает регулярную структуру хроматина и блокирует продвижение регуляторного комплекса [88, 89].

Активная барьерная активность осуществляется путем привлечения специализированных ферментативных комплексов, которые обеспечивают поддержание состояния активного или репрессированного хроматина. Данный механизм был продемонстрирован для 5'cHS4 инсулятора из  $\beta$  – глобинового локуса кур [80]. Модель защиты от гетерохроматина, основана на том, что специфические факторы, такие как, USF1 и TFIIIC, взаимодействуя с инсулятором, вызывают локальное освобождение участка ДНК от нуклеосом, или с помощью специфических модификаций гистоновых белков делают этот

участок хроматина открытым. Такие локальные изменения хроматина приводят к созданию барьера, блокирующего распространение гетерохроматина [90]. Барьерные инсуляторы посредством белка CTCF могут быть привязаны к специфическим компартментам ядра, а с помощью других факторов, взаимодействующих с инсулятором, обеспечивается ограничение распространения метилирования. Можно предположить, что механизмы лежащие в основе активации промоторов генов лежат также в основе активации и функционирования инсуляторов [83].

Вторая модель, объясняющая функционирование инсуляторов, основана на том, что энхансеры взаимодействуют с промотором посредством механизма выпетливания. В данной модели перемещение транскрипционных факторов связанных с энхансером в область с целевым промотором происходит путем выпетливания хроматина [91]. Данные факторы могут привлекать или взаимодействовать с такими компонентами транскрипционного аппарата как медиатор, TFIID или PolII или же изменять организацию хроматина в области промотора и фланкирующих его участков, что приводит к активации или репрессии транскрипции. Использование молекулярно-биологического метода ChIA-PET показало, что данные взаимодействия опосредованы CTCF и выявило локальную (в пределах 1 млн. п.н.) природу таких взаимодействий [92]. Использование РНК-интерференции для искусственного снижения уровня CTCF, выявило сокращение числа дистальных взаимодействий между регуляторными элементами и снижение активности связанных с ними генов [93]. Было высказано предположение, что при таком типе взаимодействия энхансера с промотором инсулятор, находящийся между ними, играет роль транскрипционной «ловушки», взаимодействуя с основными факторами транскрипции в составе преинициаторного комплекса [91, 94]. Процесс захвата в данном механизме предполагает существование белок-белковых взаимодействий между факторами, связанными с энхансером и инсулятором. Так как инсуляторы проявляют небольшую специфичность в отношении энхансеров, которые они блокируют, белки, которые взаимодействуют с

инсуляторами должны обладать широким спектром потенциальных белок-белковых взаимодействий.

Третья модель основана на том, что энхансер–блокирующие элементы взаимодействуют друг с другом, что приводит к выпетливанию промежуточных областей участков генома [57]. Данное предположение основано на том, что эукариотические хромосомы имеют петельно-доменную архитектуру [95]. В данной модели нарушение влияния энхансеров и сайленсеров на регулируемые ими гены достигается за счет образования топологически разобщенных хроматиновых петель. Это приводит к тому, что энхансеры, расположенные в одном петлевом домене, не могут из-за пространственной отдаленности взаимодействовать с промоторами расположенными в другом петлевом домене [96]. Энхансеры и сайленсеры расположенные в пределах одной петли могут контактировать с их целевыми генами путем скольжения хроматиновых фибрилл или при помощи внутри-петельных контактов [97]. Возможность независимого движения в трехмерном пространстве регуляторных элементов, расположенных в одной петле, и целевых генов в другой, также усложняет установление стабильного взаимодействия [98].

Экспериментальная проверка указанных выше моделей показала, что барьерная модель инсулятора наиболее хорошо описывает функционирование инсуляторных генов tPHK, SINE-B2 и MIR элементов [21, 99, 100]. Модель, в которой инсуляторы образуют топологически связанные домены, может в полной мере объяснить, как барьерную, так и энхансер-блокирующую функцию большинства CTCF-связывающих инсуляторов и последние данные полученные с использованием методов 3C и Hi-C подтверждают то, что инсуляторы выполняют роль структурных элементов, физически ограничивающих действие энхансеров, [19, 101]. Модель транскрипционной ловушки на данный момент является лишь теоретической и не имеет экспериментального подтверждения. Тем не менее, в последние годы произошел большой прогресс в исследовании молекулярных механизмов

работы инсуляторов, и дальнейшая работа в данном направлении может не только подтвердить уже имеющиеся, но и выявить новые механизмы функционирования инсуляторов.

### **2.1.6. Функциональные методы изучения инсуляторов**

Основным методом для функционального изучения инсуляторных последовательностей является метод репортерного гена. Несмотря на то, что CTCF проявляет энхансер-блокирующую функцию, находясь между энхансером и промотором в составе плазмиды, содержащей репортерный ген, не все CTCF-связывающие последовательности проявляют инсуляторную активность [45].

Изучение энхансер-блокирующей активности CTCF-связывающих сайтов в их естественном окружении является гораздо более сложной задачей. Наиболее ярким примером изучения CTCF-связывающих сайтов в их нативном окружении является локус *H19-Igf2*. Использование методики MS-MLPA (methylation-specific multiplex ligation-dependent probe amplification) и последующего бисульфитного секвенирования позволило выявить в данном локусе мутации затрагивающие сайты связывания CTCF, которые влияют на статус метилирования ICR (imprinting control region) [20].

Использование сайт-специфического мутагенеза также позволяет изучать функционирование CTCF-связывающих сайтов. Данный подход был использован при установлении функции CTCF-связывающих сайтов в локусе, кодирующем тяжелые цепи иммуноглобулина мыши, [38]. В последнее время использование направленного мутагенеза при помощи методики CRISPR/Cas позволило выявить основополагающую роль CTCF в образовании петельных доменов [19]. Сайт-специфический мутагенез дополняют методы «потери функциональности», которые используют генетические подходы нарушающие функционирование и экспрессию CTCF. Использование данного подхода позволило изучить роль CTCF в регуляции экспрессии генов протокадгеринов *Pcdh*, определяющих разнообразие нейронов [37].

## 2.2. ЭНХАНСЕРЫ

Энхансеры представляют собой гетерогенный класс *цис*-регуляторных элементов, усиливающих транскрипцию определенных генов. Впервые энхансер был идентифицирован в геноме онкогенного вируса SV40 благодаря способности значительно усиливать транскрипцию сопряженного гена бета-глобина кролика. [102]. В дальнейшем энхансеры были обнаружены в геномах других организмов. Первый энхансер человека был найден в локусе тяжелой цепи гена иммуноглобулина [103]. Недавние исследования также показали, что действию энхансеров могут подвергаться гены микро РНК [104, 105]. За последние 30 лет было идентифицировано большое количество энхансеров и показано, что их активность может зависеть от типа клеток и тканей, стадии развития организма, сигналов от клеточных рецепторов и др. В настоящий момент можно выделить несколько функциональных когорт энхансеров: локус контролирующие области [106], теновые энхансеры [107], супер-энхансеры [108, 109], Мега-Транс энхансеры [110], плейотропные энхансеры [111] и подготовленные (poised) энхансеры [112]. Изучение энхансеров позволило выделить ряд общих свойств характерных для этих элементов:

- 1 Энхансерные элементы имеют длину 50-1500 п.н. и содержат сайты связывания транскрипционных факторов, которые в свою очередь привлекают другие повсеместно экспрессируемые или тканеспецифичные факторы [113].

2. Энхансеры часто располагаются на большом удалении от активируемого промотора или промоторов (до 10 т.п.н. у дрозофилы и 100 т.п.н. у человека) и их функционирование не зависит от геномного окружения.

3. Действие энхансеров не зависит от их ориентации по отношению к последовательности активируемого промотора [114].

4. Энхансеры имеют собственные маркеры хроматина, такие как монометилирование четвертого остатка лизина гистона H3 (H3K4me1). Кроме этого, энхансеры связывают белок p300, транскрипционный коактиватор с ацетилтрансферазной активностью [115]. Для активных энхансеров характерно расположение в областях генома с пониженным содержанием

нуклеосом [116], что определяет их высокую чувствительность к ДНК нуклеазам, таким как ДНКаза I [117]. Также для последовательностей энхансеров характерны нуклеосомы содержащие особые варианты гистоновых белков, такие как H3.3 и H2A.Z [118]. Данные модификации гистонов в совокупности с гиперчувствительностью к ДНКазе I и связыванием определенных коактиваторов являются признаками активных энхансеров и используются при их картировании

Изучение функционирования энхансеров представляет большой интерес, что связано с признанием их значения не только в экспрессии генов, но и в процессе эволюции и развитии болезней, так называемых «энхансеропатий», связанных с однонуклеотидными полиморфизмами [119]. Несмотря на большой промежуток времени, прошедший с момента открытия энхансеров, наши знания об этих элементах остаются на удивление неполными. Для подавляющего большинства энхансеров животных их пространственно-временная активность неизвестна [120], предполагается, что экспрессия лишь небольшого количества генов связана определенными специфическими энхансерами [121].

### **2.2.1. Структура энхансеров**

Энхансеры являются гетерогенным классом *цис*-регуляторных элементов. Так, энхансеры значительно отличаются друг от друга, как по длине, так и по нуклеотидной последовательности, которая определяет то, какие факторы транскрипции связываются с энхансером. В отличие от простой и понятной природы генетического кода, правила регуляторного кода остаются неясными. Регуляторный код представляет собой набор правил, которые предопределяют энхансерную функцию последовательности и задают особенности активности энхансеров в различных типах клеток. Последовательности энхансеров содержат короткие последовательности ДНК, называемые мотивами, которые связывают транскрипционные факторы. Однако, в отличие от кодонов генетического кода, мотивы имеют различную

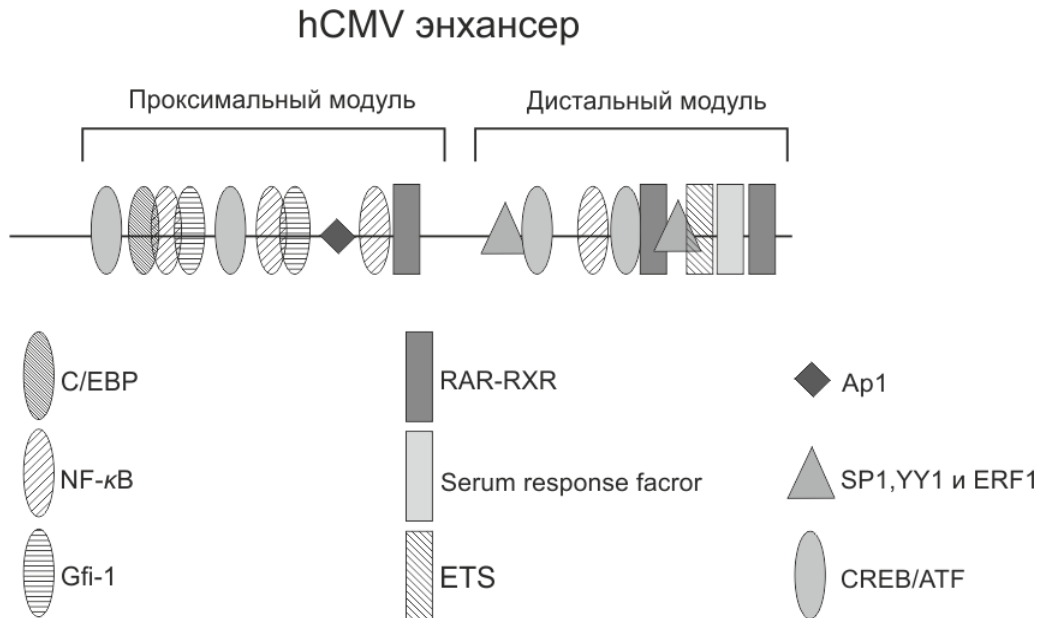


длину и могут быть разделены участками нейтральной последовательности. Показано, что порядок и расположение мотивов транскрипционных факторов в паралогичных энхансерах, выполняющих схожую функцию, сильно варьирует [122]. Исследование ортологов энхансера гена *eve* у *Drosophila* также показало наличие различий в их структуре при сохранении общего паттерна экспрессии [123]. Для регуляторных элементов в отличие от белок-кодирующих генов также характерно отсутствие четко определенных границ.

Функциональный анализ цис-регуляторных элементов показал избыточность и дискретность большинства энхансеров [124], что привело к формированию трех моделей организации энхансеров: билборда, энхансеосомы и модели коллективного связывания транскрипционных факторов [125, 126].

### **Модель билборда**

В модели билборда энхансер состоит из отдельных функциональных модулей, которые способны независимо регулировать транскрипцию. Каждый отдельный модуль состоит из небольшого количества сайтов связывания транскрипционных факторов.



**Рисунок 2.** Энхансер цитомегаловируса человека hCMV. Энхансер состоит из двух модулей: проксимального и дистального. Показаны транскрипционные факторы, связывающиеся с энхансером [127].

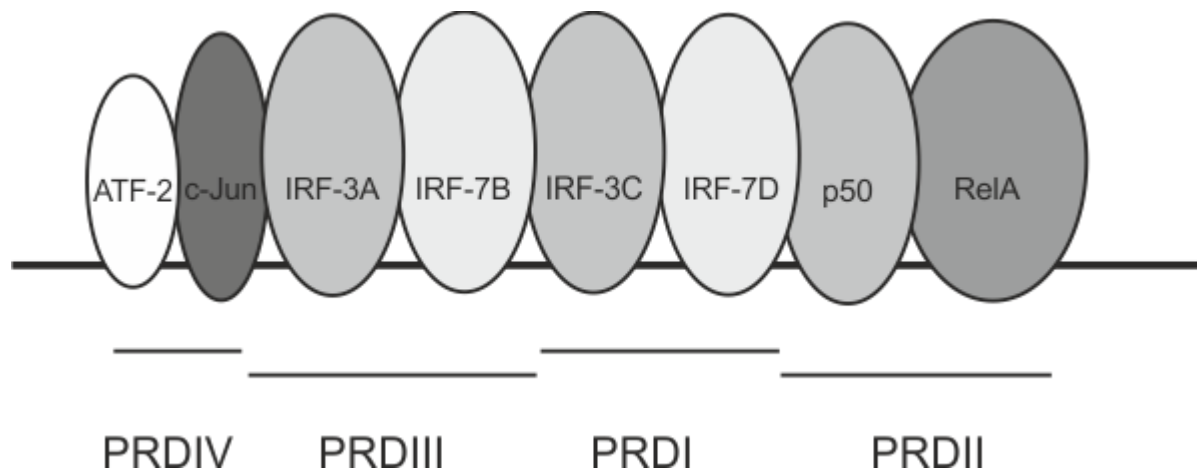
Наиболее хорошо изученным энхансером, обладающим таким типом организации, является энхансер вируса CMV человека (рисунок 2). CMV энхансер содержит большое количество сайтов связывания эукариотических транскрипционных факторов NF-κB, CREB/ATF, AP-1, RAR-RXR, SP1, YY1 и ERF, которые формируют два независимых модуля [128]. Делеционный анализ показал, что модули CMV энхансера могут функционировать, как независимо друг от друга, так и кооперативно, привлекая РНК полимеразу II и усиливая тем самым транскрипцию сопряженных с энхансером генов [128]. Использование направленного мутагенеза для исключения сайтов связывания NF-κB или CREB/ATF никак не сказывалось на репликации вируса в клетках человеческих фибробластов [129]. Данное исследование наглядно показало компенсаторную роль модулей при такой организации энхансера. Аналогичной двухмодульной организацией обладает энхансер вируса SV40 [130].

В данной модели транскрипционные факторы комбинаторным способом обеспечивают точный паттерн экспрессии генов. При этом функционирование энхансера не зависит от связывания транскрипционных факторов кооперативным способом по закону «все или ничего», что обеспечивает необходимую эволюционную гибкость энхансера без потери функциональности энхансера [131]. Наличие указанных свойств делает данную модель энхансера наиболее предпочтительной для описания структуры большинства энхансеров и недавние полногеномные исследования подтверждают, что такой тип организации энхансеров является преобладающим в геноме [132].

### ***Модель энхансеосомы***

Модель энхансеосомы описывает энхансер как матрицу для сборки сложного нуклеопротеидного комплекса называемого энхансеосомой. В отличие от модели энхансеров, обладающих структурой билборда, для сайтов связывания транскрипционных факторов в энхансеосомах характерен

синергизм действия. Потеря хотя бы одного из этих сайтов ведет к потере функциональности всего энхансера [133]. Сайты связывания транскрипционных факторов в энхансеосомах могут как пересекаться [134], так и нет [135].



**Рисунок 3.** Схематическое изображение энхансеосомы гена  $\beta$  – интерферона. ATF-2/c-Jun связывается с участком PRDIV, IRF-3 и IRF-7 связываются с участком PRDIII-I и NF $\kappa$ B связывается с участком PRDII [134].

Наиболее хорошо изученным энхансером, обладающим такой структурой, является энхансер гена  $\beta$  – интерферона. Экспрессия гена  $\beta$  – интерферона требует координированного связывания с ДНК транскрипционных факторов ATF-2/c-Jun, IRF-3, IRF-7 и NF- $\kappa$ B (p50/RelA) [136]. Энхансер  $\beta$ -интерферона разделен на четыре позитивных регуляторных домена (PRDs). Белки IRF связываются с PRDI и III, NF- $\kappa$ B с PRDII, и ATF-2/c-Jun с PRDIV, каждый из этих факторов может взаимодействовать с коактиватором CBP/p300 (Рисунок 3). При вирусной инфекции *in vivo* эти транскрипционные факторы кооперативно связываются со свободным от нуклеосом участком промотора. Важно отметить, что индивидуальные транскрипционные факторы не способны активировать транскрипцию интерферона  $\beta$ , и только при наличии всех компонентов, входящих в состав энхансеосомы, этот процесс может быть запущен.

Несмотря на то, что имеются данные полногеномного анализа, в которых показана распространенность попарного связывания

транскрипционных факторов [137], энхансеосомы и подобные им регуляторные элементы не являются распространенными, что связано с тем, что для сборки энхансеосомы необходимо коллективное связывание нескольких транскрипционных факторов. Возможным объяснением существования таких элементов является то, что они необходимы только в уникальных регуляторных условиях, например, для активации энхансеров, регулируемых транскрипционными факторами с низкой концентрацией в ядре [138] или для предотвращения нежелательного взаимодействия транскрипционных факторов и эктопической энхансерной активности [139].

#### ***Модель коллективного связывания транскрипционных факторов.***

Данный тип энхансеров схож по своей структуре с энхансерами, обладающими организацией билборда, так как положение, количество и ориентация сайтов связывания транскрипционных факторов не является столь важными для их функционирования. Однако, в отличие от энхансеров-билбордов, для обеспечения полной функциональности данного типа энхансеров необходимо связывание полного набора транскрипционных факторов, что является характерным свойством энхансеосом [126]. Для транскрипционных факторов необходимых для функционирования данного типа энхансеров также может быть характерно их привлечение в энхансер посредством белок-белковых взаимодействий с уже связавшимися факторами [126]. Данные последних исследований свидетельствуют о том, что подобная организация энхансеров может быть характерной для регуляторных элементов контролирующей морфогенез [140].

На данный момент, наиболее изученными энхансерами, обладающими таким типом организации, являются кардиальные энхансеры *Drosophila*, индуцирующие дифференциацию сердечной поперечнополосатой мышечной ткани [126].

Различные особенности того как организованы сайты связывания транскрипционных факторов в энхансеосоме, билборде или модели коллективного связывания могут помочь объяснить различную степень

консервативности между *цис*-регуляторными элементами. Секвенирование геномов различных организмов показало, что наиболее консервативными являются регуляторные элементы, связанные с критически важными в процессе эмбрионального развития генами [141]. Согласно описанным выше свойствам, модель энхансесомы накладывает жесткие ограничения на сайты связывания транскрипционных факторов, что обеспечивает высокую консервативность последовательности. Напротив, в модели билборда и коллективного связывания данные ограничения кооперативного связывания транскрипционными факторами ослаблены, что определяет их более высокую скорость эволюции последовательности. К сожалению, немногие высоко консервативные энхансеры были подвергнуты функциональной диссекции и, таким образом, нам не хватает информации, какие модели лучше всего характеризуют функцию и консервативность энхансеров.

### **2.2.2. Функциональные модели активации транскрипции энхансерами**

Энхансеры могут способствовать инициации транскрипции (и, таким образом, усиливать уровень транскрипции гена) несколькими способами. Так связывание энхансера с активатором может приводить к привлечению комплексов перестройки хроматина и позиционирования нуклеосом. Связывание энхансера с активатором также может приводить к привлечению основных факторов аппарата транскрипции к кор-промотору [142]. Передачу сигнала от активаторов, связанных с энхансером, к инициаторному комплексу, расположенному в области кор-промотора, может обеспечивать медиатор [114, 143].

Описанные выше модели и различные особенности связывания транскрипционных факторов позволяют объяснить различные регулирующие стратегии характерные для них. Функциональный анализ энхансеров привел к формированию двух моделей транскрипционной активации генов, бинарной «стохастической» и непрерывной «реостатической» [91, 144].

Некооперативные взаимодействия транскрипционных факторов с ДНК хорошо подходят для точной регуляции экспрессии генов, которая часто необходима для гомеостатических ответов, например, для генов, участвующих в воспалительной реакции организма, при которой сила воспалительной реакции зависит от микробной нагрузки [145]. Реостатическая модель предполагает, что энхансер может регулировать уровень экспрессии сопряженного гена «количественно», активность гена при этом будет меняться постепенно. Опыты, в которых контролировалась концентрация транскрипционного фактора, связывающегося с последовательностью энхансера, подтвердили эту модель [146].

Стохастический механизм регуляции активности генов хорошо применим там, где транскрипционная активность зависит от активности энхансера и может находиться в двух возможных состояниях, «включенном» и «выключенном» [147]. Данный механизм характерен для генов, которые участвуют в процессах связанных с развитием, например органогенез, и требующих четкой демаркации анатомических границ между тканями [148]. В данной модели активация транскрипции происходит благодаря кооперативному связыванию транскрипционных факторов с последовательностью энхансера [149]. Данный механизм характерен для энхансеров имеющих структуру билборда и коллективного связывания.

### **2.2.3. Механизмы действия энхансеров**

Существует две основные модели действия энхансеров: модель трекинга и модель выпетливания.

Согласно модели выпетливания взаимодействие энхансера с промотором происходит благодаря сближению их последовательностей в пространстве, при котором ДНК образует петлю. Способность энхансера связываться с последовательностями промотора посредством ДНК-белковых и белок-белковых взаимодействий, при его линейной удаленности от сайта начала транскрипции гена, обеспечивается механизмами сближения

последовательностей энхансера и промотора в пространстве, в которые могут быть вовлечены такие белки, как когезин или CTCF [93, 150]. Полный набор белков, опосредующих взаимодействие энхансера и промотора, не определен, но отдельные его представители описаны и включают GATA1, EKLF (KLF1), TFIИ-I, SATB-1 (для разных случаев показана необходимость наличия разных факторов) [151].

Модель «трекинга», предполагает привлечение факторов к промотору, которое осуществляется в результате предварительной сборки комплекса в области энхансера (сборки комплекса с участием основных факторов транскрипции и РНК-полимеразы II) и дальнейшего перемещения комплекса к промотору по ДНК [142]. Эту модель подтверждает исследование, в котором были обнаружены короткие РНК транскрибируемые на протяжении молекулы ДНК между энхансером и промотором [152]. Эта модель, также была подтверждена исследованием регуляции транскрипции гена *HNF-4α* (hepatocyte nuclear factor), в котором было показано наличие между энхансером и промотором комплексов транскрипционных факторов связанных с РНК-полимеразой II [153]. Несмотря на то, что вероятно одни энхансеры функционируют согласно модели «трекинга», а другие согласно модели выпетливания, данные получаемые в последние годы все больше подтверждают механизм функционирования энхансера через образование петли [154].

Несмотря на то, что большинство энхансеров действуют преимущественно в *цис*-положении, эксперименты показали, что энхансеры не обязаны находиться на той же молекуле ДНК, что и их целевые промоторы [155]. Например, для *Drosophila* данное явление известно как трансекция [156]. Действие энхансеров в *транс*-положении описано также и у позвоночных [157]. Обращает на себя внимание то, что активация энхансером промотора в *транс*- не менее сильна, чем в *цис*-, но такое явление имеет гораздо более спорадический характер и имеет место только в нескольких клетках в организме [156, 157].

Важно отметить, что физическая близость в результате выпетливания хроматина является необходимым условием для активации бета-глобиновых генов их локус контролирующей областью, расположенной в *цис*-положении [158]. Развитие методов получения изображений с высоким разрешением [159] и методов фиксации конформации хромосом (3С) [160], также показало, что взаимодействия энхансера с промотором является следствием их непосредственного физического контакта, в то время как интроны, как правило, при таком взаимодействии выпетливаются [161]. Несмотря на большое количество полногеномных данных о внутриядерных контактах, в том числе данных о геноме человека на основе которых были построены 3D-карты с высоким разрешением [162], количество наших знаний о механизмах формирования, поддержания и контроля взаимодействующих комплексов регуляторных элементов, в частности, наиболее изученных комплексов промотор-энхансер, все еще недостаточно для понимания их функционирования. Вероятно, одной из причин этого является то, что разрешающая способность полногеномных 3С методик (с разрешением более 1 т.п.н) все еще не является достаточной, и исследуемый фрагмент может включать в себя помимо промоторов и энхансеров другие регуляторные последовательности. Кроме того, близкое пространственное расположение элементов не обязательно означает, наличие их функциональной связи [163, 164].

Другая причина состоит в том, что в геноме осложнены механизмы взаимодействий промотора с энхансером, и активный промотор взаимодействует с несколькими энхансерами, распределенными главным образом в пределах 500 т.п.н. от промотора [17, 162, 165]. Энхансерный элемент, в свою очередь, контактирует в среднем с двумя промоторами, причем эти элементы вместе с сайленсерами и инсуляторами участвуют во множественных взаимодействиях, которые образуют сложные сети [17, 164].



Механизм повышения уровня транскрипции, работающий при сближении энхансера и промотора в петлевой модели, неизвестен. Предложены две модели, объясняющие функционирование энхансера.

В первой модели происходит увеличение локальной концентрации факторов транскрипции вблизи промотора, происходящее при сближении промотора и энхансера [114]. Во второй модели происходит релокализация комплекса энхансер-промотор в ядерный субкомпаратмент с определенными условиями (например, с высокой концентрацией или наличием специфического фактора) [166].

В обоих случаях взаимодействие промотора и энхансера приводит к изменению окружения промотора, способствующему сборке преинициаторного комплекса (PIC). Оба механизма ведут к повышению вероятности ассоциации промотора и транскрипционных факторов, необходимых для сборки PIC и инициации транскрипции. В первом случае имеет место специализированное и направленное привлечение факторов, в другом - создание более благоприятного для самостоятельной ассоциации факторов окружения. Однако, следует иметь в виду возможность перемещения гена в область заранее собранных комплексов, что ближе к первой модели.

#### **2.2.4. Регуляция активности энхансеров**

##### ***Структура хроматина и гистоновый код***

Изучение энхансеров показало, что активность энхансеров определяется различными паттернами модификаций гистоновых белков, структурой хроматина, транскрипционными факторами и некодирующими РНК.

Энхансер может находиться в трех состояниях: активном, праймированном и подготовленном. Для активных энхансеров характерны модификации гистонов H3K4me1 и H3K27ac [115, 167]. Праймированное состояние предшествует активации энхансера и характеризуется высоким уровнем H3K4me1 и отсутствием H3K27ac [167]. Энхансеры генов,

отвечающих за эмбриогенез, находятся в подготовленном состоянии и для них характерно триметилирование по 27 лизиновому остатку гистона H3 (H3K27me3), замещающее H3K27ac [112, 168]. Исследование энхансеров генов вовлеченных в дифференцировку стволовых клеток крови показало, что подготовленное состояние энхансеров также характеризуется наличием модификаций гистонов H3K9me1 и H3K27me1 [169]. Присутствие H3K27ac широко используется, для различия активных энхансеров от уравновешенных (poised). Несмотря на то, что H3K4me1 и H3K27ac являются характерными метками энхансеров, дополнительные модификации гистонов, такие как H3K9ac и H3K18ac, кротонилирование гистонов и H3K79me2 и H3K79me3 были идентифицированы в энхансерных регионах, но до сих пор полногеномно не картированы [170, 171].

Недавние исследования показали, что 5-гидроксиметилцитозин (5hmC), который является промежуточным продуктом при деметилировании ДНК, также может являться маркером активных энхансеров [172] и супер-энхансеров [173]. Изучение распределения 5hmC в геноме показало высокую степень колоколизации 5hmC с маркерами H3K4me1 и H3K27ac [174]. Также было показано что 5hmC в отличии от 5mC фланкирует сайты связывания транскрипционных факторов в активных энхансерах [172].

### ***Транскрипционные факторы***

Ключевые регуляторные последовательности ДНК, обеспечивающие энхансерную активность, могут быть представлены несколькими мотивами ДНК длиной 6-8 пар оснований [175]. Связывание транскрипционных факторов с энхансером определяется доступностью ДНК. Открытие хроматина может опосредоваться кооперативным связыванием транскрипционных факторов, привлечением «пионерских» факторов, а также непосредственным депонированием гистона H2A.Z.

Кооперативное связывание транскрипционных факторов облегчает преодоление нуклеосомного барьера и инициирует регуляторное событие на

хроматине. Изучение *in vitro* кооперативного связывания транскрипционных факторов GAL4, USF и NF-κB подтвердило данное предположение [176].

Инициация связывания транскрипционных факторов, обеспечивающих функционирование энхансера, может также происходить опосредованно с привлечением пионер-факторов [177], которые опосредуют связывание других факторов через смещение нуклеосом, привлечение ферментов модифицирующих гистоны или посредством защиты ДНК от метилирования [178]. Наиболее изученным примером такого пионер-фактора является FOXA1, который связываясь с хроматином, дестабилизирует локальные нуклеосомные взаимодействия, а также стабилизирует позиционирование нуклеосом, регулируя доступность сайтов связывания транскрипционных факторов [179, 180].

Открытие хроматина может также происходить при образовании доменов хроматина, обладающих высокоподвижными нуклеосомами, путем непосредственного включения в нуклеосомы гистона H2A.Z, который привлекает «пионерские» транскрипционные факторы. Включение гистона H2A.Z в нуклеосомы происходит в виде димера H2A.Z/H2B и опосредуется белковым комплексом TIP60/p400 [181]. Недавние исследования дифференцировки стволовых клеток мыши показали, что открытию хроматина предшествует образование доменов богатых гистоном H2A.Z, которые привлекают связывание пионер-фактора FOXA2, который в свою очередь привлекает факторы ремоделинга хроматина INO80 и SWI/SNF [182].

### ***Некодирующие РНК***

Анализ библиотек кДНК [183], геномные высокоплотные чипы (genome tiling arrays) [184, 185] и CAGE (cap analysis gene expression) [186], показали, что примерно 93 % эукариотического генома транскрибируется, большая часть РНК транскриптов представлена некодирующими РНК (ncRNA) [187, 188].

Основными центрами образования некодирующей РНК наряду с промоторами являются энхансеры. В настоящий момент для человека описано около 40000-65000 различных РНК транскрибированных с энхансеров [189].

Образование энхансерной РНК (eRNA) является характерным маркером активных энхансеров [189].

Недавние исследования показали, что внутригенные энхансеры могут выполнять функцию альтернативных тканеспецифичных промоторов, которые транскрибируются с образованием большого количества длинных РНК, называемых мульти-экзонными РНК (meRNA) [190]. meRNA проявляют низкий кодирующий потенциал и предполагается, что данные РНК могут быть побочными продуктами активации энхансеров или выполнять регуляторную роль.

Показано, что уровень энхансерной РНК (eRNA) коррелирует с активацией соответствующих энхансеров и уровнем их активности в геноме [167, 191]. Исследование дистальных взаимодействий между точками старта транскрипции (TSS) и энхансерами с использованием метода 5C подтвердило присутствие eRNA в активных энхансерах и выявило корреляцию между экспрессией генов, взаимодействием энхансеров с промоторами и присутствием eRNA. Более того, было показано, что только 7% дистальных взаимодействий в геноме осуществляется с близлежащими генами [17]. Недавние исследования показали, что присутствие связанной eRNA значительно усиливает связывание энхансерами транскрипционного фактора YY1, который может привлекать гистоновые ацетилазы, что приводит к активации транскрипции [192]. Функциональную роль eRNA также подтвердили исследования, в которых нокдаун РНК-транскриптов приводил к нарушению экспрессии гена-мишени [193].

#### **2.2.5. Идентификация и функциональный анализ энхансеров**

Идентификация энхансеров в геноме является важнейшей задачей для понимания механизмов регуляции экспрессии генов. Идентификацию предполагаемых энхансеров и их картирование в геноме осуществляют с помощью методов классической генетики, сравнительной геномики и биохимических методов.

Биохимические подходы основаны на отличительных особенностях хроматина, таких как чувствительность к ДНКазе I, связывание с транскрипционными факторами, а также паттерны модификаций гистонов, которые связаны с функциональными некодирующими областями, как было указано выше. В последние годы широкое распространение получили методы основанные на иммунопреципитации хроматина с последующей гибридизацией на ДНК-микрочипах (ChIP-CHIP) или последующим секвенированием (ChIP-seq), которые позволили провести полногеномное картирование сайтов связывания транскрипционных факторов (TFBS) и эпигенетических маркеров хроматина [194, 195]. Параллельное картирование TFBS, маркеров хроматина и сайтов чувствительности к ДНКазе I позволяет проводить поиск и картирование энхансеров в полногеномном масштабе [167, 196]. Последовательности энхансеров связывают кластеры транскрипционных факторов [197] и часто ассоциированы с компонентами медиаторного комплекса и транскрипционными коактиваторами p300/CBP [68, 198]. Активные энхансеры также колокализуются с характерными модификациями гистонов H3K4me1 и H3K27ac [167].

Хотя данные свойства хроматина могут быть полезны для выявления потенциальных энхансеров, функциональный вклад каждого из перечисленных факторов остается неизвестным. Использование полногеномных методов позволило идентифицировать примерно 1 млн. потенциальных энхансерных элементов в геноме млекопитающих [117, 199, 200]. И только часть из картированных таким образом потенциальных элементов была охарактеризована с использованием функциональных подходов.

Эволюционные подходы, основанные на консервативности функциональных последовательностей, также имели успех в определении потенциальных энхансеров и мотивов [7, 201].

Данный метод основан на предположении, что последовательности ДНК, обладающие функциональной значимостью, эволюционируют

медленнее, чем нефункциональные последовательности, которые быстрее накапливают мутации [202]. Сравнительный анализ геномов нескольких организмов позволяет идентифицировать эволюционно консервативные некодирующие элементы, которые при экспериментальном изучении часто проявляют энхансерную активность [203, 204]. Исследования показали, что высоко консервативные некодирующие элементы (CNEs – conserved non-coding elements) чаще расположены вблизи генов, участвующих в раннем эмбриональном развитии. Предположительно, это связано с тем, что мутации в энхансерах, определяющих точный паттерн экспрессии в эмбриональном развитии, приводят к возникновению патологического фенотипа [205, 206].

Секвенирование 29 геномов млекопитающих позволило с высокой точностью картировать эволюционно консервативные элементы и показало, что 4,2% последовательностей генома являются эволюционно консервативными [7]. Данный сравнительный анализ также показал, что 68% обнаруженных консервативных последовательностей расположены в интронах и межгенных областях, 30% консервативных последовательностей были локализованы с характерными маркерами активных энхансеров H3K4me1 и H3K27ac [7]. В геноме млекопитающих также были обнаружены CNE, обладающие степенью консервативности близкой к 100% [207].

Однако, несмотря на наличие или отсутствие высокой степени консервативности данное свойство не является доказательством функциональности. Это было продемонстрировано путем делетирования высоко консервативных областей, которые не оказали влияние на наблюдаемый фенотип [208] и примерами функционально охарактеризованных регуляторных элементов, практически не обнаруживающих консервативности [111, 209].

Подходы основанные на консервативности регуляторных элементов ограничены необходимостью сравнения последовательностей разных видов организмов, что вызывает проблемы связанные с идентификацией видоспецифичных элементов, в частности, из-за их быстрой скорости

эволюции [210, 211]. В то время, как скорость эволюции промоторов сопоставима со скоростью эволюции белок-кодирующих генов, анализ консервативности в сочетании с биохимическими подходами геномов 20 различных видов млекопитающих показал, что энхансеры эволюционируют в среднем в три раза быстрее и большинство энхансеров млекопитающих, по всей видимости, являются видоспецифичными [212]. Таким образом, при анализе предполагаемых энхансеров нельзя отдавать предпочтение изучению их консервативности, так как, скорее всего, это может привести к неверному толкованию результатов.

Классический генетический подход основан на направленном мутагенезе с последующим изучением фенотипического или функционального проявления внесенной мутации. Данный подход чаще всего используется в модельных организмах, таких как дрожжи и *Drosophila*, использование в качестве модельных организмов млекопитающих, таких как мыши и крысы, затруднено высокой трудоемкостью и временными издержками, поэтому лишь малое число энхансеров человека были идентифицированы с помощью классических генетических подходов. Несмотря на то, что простота и эффективность направленного мутагенеза быстро растет, этот подход по-прежнему характеризуется довольно малой пропускной способностью [213-216].

Несмотря на то, что генетические изменения и соответствующие изменения в наблюдаемом фенотипе могут достоверно идентифицировать функцию, они не могут исключить отсутствие функции у изучаемого элемента при отсутствии наблюдаемых последствий мутаций из-за высокой контекстности и избыточности, характера регуляции генов.

Идентификацию энхансеров также проводят путем клонирования изучаемой последовательности в плазмидную конструкцию перед репортерным геном, сопряженным с минимальным промотором и последующим тестированием их транскрипционной активности в клеточных линиях или трансгенных организмах. Недавно были разработаны масштабные

«многопоточные» методы репортерного гена, которые позволяют проводить функциональный анализ энхансеров путем непосредственного редактирования потенциальных энхансеров и их эпигенетического состояния. «Многопоточные» методы репортерного гена позволяют выполнять анализ десятков тысяч элементов с использованием транзientной экспрессии и количественной оценки РНК методами глубокого секвенирования [217, 218]. Совершенствование данных методов и разработка таких методик, как STARR-seq, и FIREWACH, позволили проводить одновременный независимый анализ нескольких миллионов элементов [219, 220]. Также, потенциальные энхансеры могут быть проверены с помощью методик редактирования эпигенома, которые позволяют, изменяя эпигенетическое состояние хроматина, предотвратить активацию энхансера и идентифицировать таким образом ген-мишень [221-223]. Функционально охарактеризованные энхансеры собраны в базы данных, такие как REDfly для *Drosophila* [224] и VISTA enhancer browser для мыши и человека [225].

Ограничения генетического, эволюционного и биохимического подходов не позволяют дать однозначного ответа о функциональной роли элементов, идентифицируемых с их помощью.

Кроме того, энхансеры и их активность не могут быть достоверно предсказаны по последовательности ДНК или особенностям хроматина без проведения детального исследования с использованием прямых экспериментальных тестов, также не может быть предсказана и функционально значимая часть последовательности энхансера, что исключает возможность прогнозирования функциональных последствий мутации в данных последовательностях.

В отношении функциональности регуляторных элементов, определяемых данными методами существуют две точки зрения. Согласно первой, лишь небольшая часть потенциальных регуляторных элементов является функциональной, согласно второй - все идентифицируемые элементы являются нефункциональными и лишь функциональный анализ



может доказать наличие регуляторной функции у изучаемого элемента [226-228]. В отличие от методов, используемых в эволюционном и генетическом подходе, биохимические методы позволяют получить сведения о молекулярных механизмах, лежащих в основе функционирования элементов, и данные о типе клеток, в которых эти механизмы реализуются. В совокупности с развитием высокопроизводительных методов функционального анализа повышается и скорость, с которой данные элементы проверяются с использованием генетического подхода.

### 3. МАТЕРИАЛЫ И МЕТОДЫ

#### 3.1. МАТЕРИАЛЫ

##### Реактивы

В работе были использованы следующие реактивы: этиловый спирт, изоамиловый спирт, изопропиловый спирт, ацетат натрия, фенол, хлороформ медицинский, ацетон, кислота соляная, кислота уксусная ледяная, хлорид магния, хлорид калия, хлорид натрия, цитрат натрия, гидроксид натрия, сульфат меди, перхлорат аммония - квалификация “осч”, “хч”, “чда” (Реахим, Россия).

Натрия додецилсульфат (SDS), тритон X-100 (Serva, Германия);

NP40 (Sigma-Aldrich, США);

Агароза, натрия этилендиаминтетраацетат (ЭДТА), трис-(оксиметил)аминометан (Трис), дитиотрейтол (ДТТ), глицерин, акриламид (USB, США);

Мочевина (Химмед, Россия);

Бромфеноловый синий, ксиленцианол, бромистый этидий, 2-меркаптоэтанол (Bio-Rad, США);

Дезоксинуклеозидтрифосфаты (Boehringer Mannheim, Германия);

Метиленбисакриламид, TEMED (Amersham Pharmacia Biotech, Швеция);

25 mM, 50 mM MgCl<sub>2</sub>(Promega, США);

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

NEPER, сухие смеси для приготовления культуральных сред (DMEM: F12 НАМ, RPMI-1640), эмбриональная сыворотка телят, сухие смеси для приготовления культуральных сред (DMEM: F12 НАМ, RPMI-1640), эмбриональная сыворотка телят (FBS) (Gibco BRL, США);

Ампициллина натриевая соль (Биосинтез, Россия);

Наборы для приготовления проявителя и фиксажа кислого (Випс-Мед, Россия);

Lipofectamine 2000 (Invitrogen, США).

### **Ферментные препараты**

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);

Готовая смесь для ПЦР qPCRmix-HS SYBR (Евроген, Россия);

ДНК полимераза PfuII Ultra («Stratagene»);

Обратная транскриптаза PowerScript (Clontech, США);

Эндонуклеазы рестрикции *Eco* RI, *Msp* I, *Xba* I, *Hpa* I, *Nhe* I (*Asu* NHI), *Sal* I, *Xho* I, *Pst* I, *Cfr42*I (*Sac* II), *Hind* III, *Bam* HI (MBI Fermentas, Литва);

Т4 ДНК лигаза, фрагмент Кленова ДНК-полимеразы I *E. coli*, SAP (щелочная фосфатаза креветки) (MBI Fermentas, Литва).

### **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 ЭДТА;

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

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

Буфер TBEx0,5: 45 мМ трис-борат, 1 мМ ЭДТА, pH 8,0;

Буфер В (12 мМ HEPES–KOH, pH 7,9, 12% глицерин, 60 мМ KCl, 0,3 мМ ЭДТА, 0,6 мМ дитиотрейтол);

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

Буфер PBS (Phosphate Buffered Saline): 1,7 мМ KH<sub>2</sub>PO<sub>4</sub>, 5,2 мМ Na<sub>2</sub>HPO<sub>4</sub>, 150 мМ NaCl;

Буфер TNM (5x) для отжига олигонуклеотидов: 200 мМ Трис-НСl рН 8,0, 250 мМ NaCl; 100 мМ MgCl<sub>2</sub>.

### **Питательные среды для клеток *E. coli***

Среда LB: 10г триптон, 5 г дрожжевой экстракт, 5 г NaCl на 1 литр;

Среда LB с ампициллином: к среде LB добавляли раствор ампициллина до концентрации 100 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 100 мкг/мл.

### **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);

Mint RACE primer set (Евроген, Россия).

### **Маркеры длин ДНК**

100 - 1500 п.н., 1 т.п.н. (СибЭнзим, Россия);

1000 п.н. + 10000 п.н. (СибЭнзим, Россия);

MassRuler High Range DNA Ladder, Ready-to-Use, 1500-10,000 п.н (MBI Fermentas, Литва);

MassRuler Low Range DNA Ladder, Ready-to-Use, 80-1031 п.н. (MBI Fermentas, Литва).

### **Бактериальные клетки**

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

## **Оборудование**

Центрифуги MiniSpin Plus 5415 (Eppendorf, Германия), Multifuge 3 L-R (Heraeus, Германия);

Спектрофотометр GeneQuant (Amersham Pharmacia Biotech, Швеция);

Ротатор LabQuake Shaker (Barnstead Thermoline, США);

EIA Reader Model 2550 (BioRad, США);

Приборы для горизонтального электрофореза – электрофореза SE-1 и SE-2 (Хеликон, Россия);

Источники питания Model 500 Power Supply (BioRad Laboratories, США) и Power Supply Model 100 (Bethesda Research Laboratories, США);

УФ-излучатель TC-312 A (США);

Цифровая фотокамера и УФ-излучатель White/Ultraviolet Transilluminator (UltraViolet Products, США), программы для обработки VisionWorksLS (UVP Inc., США) и GelPro Analyzer (UVP Inc, США);

ПЦР-амплификаторы OmniGene (Hybaid, MWG-BioTech, Великобритания) и MiniCycler PTC-150 (MJ Research, США);

ПЦР-Анализатор LightCycler 480 (Roche, Швейцария);

Автоматические пипетки (Gilson, Франция и Labmate, Польша);

Ламинарные шкафы (Babcock-BSH, Германия и Gelaire, Италия);

Автоматический CO<sub>2</sub> инкубатор IR 1500 (Flow Laboratories, США);

Термостатированная качалка Certomat (B.Braun Biotech International, Германия);

Термошейкер Thermomixer Comfort (Eppendorf, Германия);

Охладитель проб SC2D (Биоком, Россия);

Термостат Термит (ДНК-Технология, Россия);

Электронные весы Sartorius L610D (Германия);

Усиливающий экран ЭУВ-3А (Россия).

### **Расходные материалы**

Пленка медицинская рентгеновская «Ренекс» (ЕС), нейлоновая мембрана Hybond-N (Amersham Pharmacia Biotech, Швеция), наконечники для пипеток (Treff Lab, Швейцария и QSP, США), центрифужные пробирки на 1,5 Eppendorf (Германия); 2 мл (QSP, США), 15 мл (Greiner Bio One, Германия) и 50 мл (Corning Incorporated, США), 96-луночные планшеты (Nunc Nunclon Microwell, Дания), парафиновая пленка Microseal A Film (MJ Research, США), пробирки для ПЦР в режиме реального времени (Roche, Швейцария).

### **Ресурсы WWW**

В работе использовались следующие программы и базы данных:

Human Genome Browser: <http://genome.ucsc.edu/>;

BLAT: <http://genome.ucsc.edu/>;

GenBank: <http://www.ncbi.nlm.nih.gov/Genbank/>;

BLAST: <http://www.ncbi.nlm.nih.gov/BLAST/>;

Primer3: [www-genome.wi.mit.edu/cgi-bin/primer/primer3\\_www.cgi](http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi).

### **Олигонуклеотиды**

Подбор олигонуклеотидных праймеров осуществляли с помощью программы Primer 3 (<http://frodo.wi.mit.edu/primer3/>). Выравнивание последовательностей проводили при помощи программы ClustalW ([www.ebi.ac.uk/clustalw/](http://www.ebi.ac.uk/clustalw/)). Для поиска потенциальных сайтов связывания факторов транскрипции применяли программу TESS ([www.cbil.upenn.edu/tess/](http://www.cbil.upenn.edu/tess/)), использующую базы данных TRANSFAC, JASPAR, IMD и CBIL GibbsMat. В работе пользовались ресурсами интернет сервера NCBI (<http://www.ncbi.nlm.nih.gov/BLAST>) и UCSC Genome Browser (<http://genome.ucsc.edu/goldenPath/hgTracks.html>).

В зависимости от целей праймеры конструировали вручную или с использованием программы Primer3 ([www-genome.wi.mit.edu/cgi-bin/primer/primer3\\_www.cgi](http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi)). Праймеры были синтезированы в компании Evrogen на синтезаторе Applied Biosystems ABI 3900 (Applied Biosystems, США).

**Таблица 1.** Праймеры для ПЦР-амплификации фрагментов энхансера 12.

Праймер	Последовательность праймера
ENH12 forward	AGT TTG GGC TGT GTT CTG TG
ENH12 reverse	TTT TTC TCA TTG GAC CTT TCT G
27 primer (SalI):	ACT GAG GTC GAC TAT CCA TGA ACA
T7 primer	TAA TAC GAC TCA CTA TAG GG
M13 forward	CGC CAG GGT TTT CCC AGT CAC GAC
M13 reverse	TCA CAC AGG AAA CAG CTA TGA C
myc forward	GGG ATC GCG CTG AGT ATA AA
myc reverse	GGA TCT CCC TTC CCA GGA C
AL	CGG TAC AGG TTG AGC AGC A
AR	TCT GGT TTC TTA GGG GCA AA
BL	GGG CTG CAT GAA CAG AGA GT
BR	TCA TCT TTG CCA GTC CCT TC
CL	TAC ACC CCA GCC TAA GCA TC
CR	CGG TCT TCG TGC CGT TTT

**Таблица 2.** Праймеры для ПЦР-амплификации промотора генов PSENE1 / U2AF1L4.

Праймер	Последовательность праймера
PR1	TTT ACC CAA GCC CTC CAG

Праймер	Последовательность праймера
PR2	CTT TCA GTT ATG GAC GTT TGC
CR1mut	GCT GCT CTT ATG TAC CTT CCG TTG C GCA ACG GAA GGT ACA TAA GAG CAG C
CR2mut	ACT GTT CAG GTG CGT TCG TTG CCC T AAC GAA CGC ACC TGA ACA GTG CTC C
CR3mut	AGG GCG CTT TAG CGT TTG AGG CGC GCG CCT CAA ACG CTA AAG CGC CCT
CR4mut	TTT TAA CAG CTA CGT CGT GCA GTG T ACA CTG CAC GAC GTA GCT GTT AAA A

**Таблица 3.** Праймеры для ПЦР-амплификации потенциальных энхансеров.

Праймер	Последовательность праймера
E1 forward	GTG GAC CGT GTA GCC GAGA
E1 reverse	GAC AGC CCA GCA AGC TCA
E2 forward	GAC AGT GAG CAC AGG CTA T
E2 reverse	CAC TCT ACC CCC GGC CAG TC
E3 forward	TGC TTA CAG CTT TTC CAG C
E3 reverse	TAA AAG GAG AGC CAG CAT CAG TC
E4 forward	GCT GGC TGT CCT CAT TAT TCC
E4 reverse	GAA GGA TTT TTC TTG CCC ACT
E5 forward	ACA CTC CCA CCA GGA ATC AGC
E5 reverse	CGT GGA AAC TCA AAT GG
E6 forward	CTC CAG CAG TCC CTT GTG TC
E6 reverse	CAT CCA AGC TCC CAG AGG



**Таблица 4.** Праймеры для ПЦР-амплификации потенциальных инсуляторов.

Праймер	Последовательность праймера
C2 forward	CAC TAT GAT GGT GAA AGC AG
C2 reverse	TTT TTA AGA GAA AGG TAA GTG A
C3 forward	CAC GTC TTC CAT CAA GGA CTC A
C3 reverse	CAC TCT GCG GTT CAC AC
C4 forward	TGA ACT CTG CCC TAA GAC CT
C4 reverse	GTT TCA GAC CTC TTT CCG TA
C6 forward	GAT TCC GGG ATA AAT GAC ACA C
C6 reverse	TAG ACT CAA GAA GGC ACT
C8 forward	TTA ATA CTC CAA GGT TGT CCA
C8 reverse	CAC GAC CTA TAT TTA CAG T

**Таблица 5.** Праймеры для использованные в 5'RACE.

Праймер	Последовательность праймера
SMART II	AAG CAG TGG TAT CAA CGC AGA GTA CGC rGr GrG
3'-CDS	AAG CAG TGG TAT CAA CGC AGA GTA C(T)30N-1N
Na21cap	TGT AGC GTG AAG ACG ACA GAA AAG CAG TGG TAT CAA CGC AGA GT
Na21	TGT AGC GTG AAG ACG ACA GAA
U2out	GGG TGA AGG GTA AGG GGG CCA
U2in	CAA GGT TGT CGC ACA CAT TCA TC.
PSE	GCC TGT CCT GGG AGA ATA AGC A

**Таблица 6.** Праймеры для использованные при изучении *транс*-активации промотора энхансером.

Праймер	Последовательность праймера
SV40L SV40R	GAA GGA GCT GAC TGG GTT GA AAA GCA AGT AAA ACC TCT ACA AAT GTG
$\lambda$ left $\lambda$ right	TCC GTG AGG TGA ATG TGG TG TAG TCG GCT CAA CGT GGG TT
Ecmv 1	GTG TAT CAT ATG CCA AGT ACG C GTT TTC CCA GTC ACG ACG TTG
Ecmv 2	GGC TAT GAA CTA ATG ACC CCG TA CAC GAC AGG TTT CCC GAC T
Ecmv control	GTG TAT CAT ATG CCA AGT ACG C GTG AGT CAA ACC GCT ATC CAC
Lambda	ATA GAG CAT AAG CAG CGC AAC CAC GAC AGG TTT CCC GAC T
Lambada control	AGC TAT GAC CAT GAT TAC GCC AA ATA GAG CAT AAG CAG CGC AAC
Pcmv 1	CTC CCA CCG TAC ACG CCT A CGA TAG TAC TAA CAT ACG CTC T
Pcmv 2	TCC CGT CTT CGA GTG GGT A CGC TAG CCT CGA AAT TCG GTA
Pcmv Control	TAC ATG AGC ACG ACC CGA AAG CC CGC ACC GCT TGT GTC CGA TT

## 3.2. МЕТОДЫ

### 3.2.1. Стандартные методики.

Стандартные генно-инженерные манипуляции (обработка ДНК рестриктазами, лигирование, трансформация *E. coli*, электрофорез ДНК в агарозном и полиакриламидном геле проводили согласно руководству по молекулярному клонированию [229]. Выделение геномной ДНК проводили с использованием набора Wizard SV Genomic DNA Purification System (Promega, США) в соответствии с прилагаемой инструкцией. Выделение плазмидной ДНК проводили с использованием набора реактивов Wizard Plus Miniprep DNA Purification System («Promega», США) в соответствии с прилагаемой инструкцией.

### 3.2.2. ПЦР-амплификация фрагментов ДНК.

Для ПЦР-амплификации использовали различные пары праймеров, нуклеотидные последовательности которых приведены в таблицах 1-6.

#### *ПЦР-амплификация фрагментов геномной ДНК.*

Реакцию проводили в объеме 25 мкл. Реакционная смесь содержала 1-кратный буфер для ПЦР-реакции, 2мМ MgCl<sub>2</sub>, дезоксирибонуклеозидтрифосфаты в концентрации по 0,2 мМ каждого, 1 е.а. Hot-start *Taq*-полимеразы, 10пмоль каждого праймера (таблица 1). В качестве матрицы использовали 40 нг геномной ДНК. Реакцию проводили в следующих условиях:

- 1) активация *Taq* -полимеразы (95°C, 3мин);
- 2) 35 циклов ПЦР-амплификации (95°C, 30с/60°C, 30с/72°C, 50с);
- 3) «Достройка» (72°C, 5 мин).

Реакционную смесь смешивали с буфером для нанесения ДНК на гель и разделяли продукты ПЦР электрофорезом в 1,2%-ном агарозном геле, содержащем 0,06·мкг/мл бромистого этидия. Полосу нужной длины вырезали и элюировали в ТЕ-буфере (на каждые 10мг геля добавляли 20 мкл ТЕ-буфера).

***Повторная-ПЦР продуктов амплификации геномной ДНК (подготовка фрагментов к лигированию в вектор).***

1 мкл элюированной из агарозного геля ДНК в ТЕ-буфере использовали в качестве матрицы в ПЦР-амплификации. Реакцию проводили в объеме 25 мкл. Реакционная смесь содержала 1-кратный буфер для ПЦР-реакции, 2 мМ MgCl<sub>2</sub>, дезоксирибонуклеозидтрифосфаты в концентрации 0,2 мМ каждого, 0,5 е.а. *Taq*-полимеразы, 10 пмоль каждого праймера. Реакцию проводили в следующих условиях:

- 1) 20 циклов ПЦР-амплификации (95°C, 30с/60°C, 30с/72°C, 50с);
- 2) «Достройка» (72°C, 10 мин).

Продукты ПЦР анализировали электрофорезом в 1,2%-ном агарозном геле. 1 мкл реакционной смеси брали в реакцию лигирования амплифицированных фрагментов в вектор pGEM-T («Promega», США).

***ПЦР-скрининг колоний E. coli, содержащих плазмидную ДНК со вставками.***

Выросшие колонии *E. coli* переносили в лунки 96-луночного планшета, содержащие 200 мкл среды LB с ампициллином (100 мкг/мл), и подрачивали в течение ночи при +37°C. 5 мкл клеточной суспензии лизировали нагреванием при 95°C в течение 5 мин. К лизированным клеткам добавляли реакционную смесь для ПЦР-амплификации. Итоговый состав реакции включал: 1-кратный буфер для ПЦР-реакции, 2 мМ MgCl<sub>2</sub>, дезоксирибонуклеозидтрифосфаты в концентрации 0,2 мМ каждого, 0,5 е.а. *Taq*-полимеразы, 10 пмоль каждого праймера и 5 мкл лизированных клеток *E. coli*. Реакцию проводили в следующих условиях:

- 1) 20 циклов ПЦР-амплификации (95°C, 30с/60°C, 30с/72°C, 50с).

5 мкл реакционной смеси смешивали с буфером для нанесения ДНК на гель и разделяли продукты ПЦР электрофорезом в 1,2%-ном агарозном геле, содержащем 0,06 мкг/мл бромистого этидия.

***ПЦР-амплификация фрагментов ДНК с помощью меченых праймеров***

Реакцию проводили в объеме 20 мкл. Состав реакции включал: 1-кратный буфер для ПЦР-реакции, 2 мМ MgCl<sub>2</sub>, дезоксирибонуклеозидтрифосфаты в концентрации 0,2 мМ каждого, 1 е.а. *Taq*-полимеразы, по 4 пМоль меченых праймеров и 1 нг плазмидной ДНК. Реакцию проводили в следующих условиях:

- 1) 10-20 циклов ПЦР-амплификации (95°C, 40с/60°C, 30с/72°C, 1 мин 10с);
- 2) «Достройка» (72°C, 10 мин).

Очистку меченых продуктов ПЦР проводили в 10%-ном полиакриламидном геле.

### ***ПЦР-амплификация в реальном времени***

ПЦР осуществляли с использованием готовой смеси для ПЦР qPCRmix-NS SYBR (Евроген, Россия). Реакцию проводили в объеме 25 мкл. Состав реакции включал: 1 мкл разведенной матрицы ДНК, по 1 мкл каждого праймера (3,2 пМоль), 5 мкл смеси для ПЦР qPCRmix-NS SYBR (Евроген, Россия), 17 мкл деионизированной воды. Реакцию проводили в следующих условиях:

- 1) Активация *Taq* -полимеразы (95°C, 5 мин).
- 2) 40 циклов ПЦР-амплификации (95°C, 20с/63°C, 30с/72°C 20с);

### **3.2.3. Трансформация клеток *E. coli*.**

Компетентные клетки *E. coli* штамма DH5α были получены ранее в нашей лаборатории строго в соответствии с разработанной методикой (Inoue *et al.*, 1990).

К 50 мкл компетентных клеток *E. coli*, предварительно размороженных во льду, добавляли 10 мкл лигазной смеси, инкубировали на льду в течение 30 мин и затем выдерживали 30 секунд при 42°C, после чего клетки держали 2 мин во льду. К клеткам добавляли среду LB до конечного объема 1 мл и подращивали в течение часа при 37°C. Затем высевали 150 мкл клеток на агаризованную среду LB, содержащую ампициллин (100 мкг/мл), X-gal (20

мкг/мл) и IPTG (0,25 mM), подращивали в течение ночи при 37°C и проводили ПЦР-скрининг белых колоний *E. coli*, содержащих плазмидную ДНК со вставками. Клоны, содержащие вставки, подращивали в 5 мл LB с ампициллином (100 мкг/мл) в течение ночи на термостатированной качалке при 180 об./мин. Клетки собирали центрифугированием (10 000 об./мин), из осадков клеток выделяли плазмидную ДНК с помощью набора реактивов Wizard® Plus Minipreps DNA Purification System («Promega», США).

### 3.2.4. 5'-RACE

Для определения 5' концов генов, выделяли из клеток РНК и строили полноразмерные кДНК согласно [229], с использованием двухстадийной ПЦР, набора реактивов Encyclo PCR kits (Евроген, Россия), и праймеров SMART II и 3'-CDS (таблица 5). Амплификацию концевых фрагментов первых цепей кДНК осуществляли согласно рекомендациям к набору Mint RACE primer set (Евроген, Россия) с использованием геноспецифических праймеров U2out, U2in, PSE и праймеров Na21Cap и Na21 (таблица 5). Для гена *PSENE1* проводили амплификацию в 1 раунд, а для *U2AF1L4* в 2 раунда. ПЦР-продукты клонировали в вектор pAL-TA (Евроген, Россия) и секвенировали.

#### ***ПЦР-амплификация 5'RACE***

Реакцию проводили в объеме 25 мкл. Состав реакции включал: 2,5 мкл 10-кратного буфера для ПЦР-реакции, 0,5 мкл раствора дезоксирибонуклеозидтрифосфатов 10 mM каждого, 0,5 мкл Encyclo полимеразы 50X, 0,5 мкл геноспецифичных праймеров с концентрацией 10 мкМ, 2,5 мкл праймера Na21Cap с концентрацией 0,4 мкМ, 2,5 мкл праймера Na21 с концентрацией 2 мкМ, 1 мкл кДНК-SMART и 15 мкл деионизированной воды. Реакцию проводили в следующих условиях:

- 1) Активация Encyclo полимеразы (95°C, 3 мин).
- 2) 1 стадия ПЦР-амплификации 13 циклов (20 сек/95°C, 40с/68-62°C (с уменьшением температуры отжига на 0,5°C каждый цикл), 50с/72°C);

- 3) 2 стадия ПЦР-амплификации 20 циклов (20 сек/95°C, 40с/62°C, 50с/72°C).

### **3.2.5. Подготовка плазмид для секвенирования**

К плазмидной ДНК (200 нг) добавляли 3,2 пмоль секвенирующего праймера. Пробы упаривали при 65°C в течение 20 мин. Секвенирование проводили в Центре коллективного пользования «Геном» с использованием ABI PRISM BigDye Terminator v.3.1 kit и автоматического секвенатора ABI PRISM 3100-Avant.

### **3.2.6. Электрофорез в агарозном геле**

К предварительно прогретой до 90-95°C 1,2% суспензии агарозы в 1-кратном буфере TAE добавляли бромистый этидий до концентрации 0,06 мкг/мл, охлаждали до 50°C и заливали в форму для геля. Электрофоретическое разделение ДНК проводили в 1-кратном буфере TAE при напряжении 100-125 В. Результаты гель-электрофореза фиксировали с помощью цифровой фотокамеры White/Ultraviolet Transilluminator (UVP Inc., США) и программы GelPro Analyzer (UVP Inc., США).

### **3.2.7. Электрофорез в полиакриламидном геле (очистка меченой ДНК)**

Аmplифицированные меченные <sup>32</sup>P фрагменты ДНК очищали в 10% полиакриламидном геле с 0,17% N,N-метилден-бис-акриламидом (сшивка 30:1) в 0,5-кратном TBE-буфере. Электрофорез проводили при силе тока в 24 мА. По прохождении ксиленцианолом 1/2 длины геля электрофорез останавливали, гель закрывали лавсановой пленкой и проводили радиоавтографию в течение 15 мин. Часть геля, соответствующую искомому фрагменту ДНК, вырезали из геля, измельчали и элюировали ДНК в 300 мкл буфера для элюции ДНК из полиакриламидного геля (0,5 М ацетат аммония, 1% SDS, 1 мМ ЭДТА) при 37°C в течение ночи. Затем отфильтровывали частицы геля через стекловату, добавляли 300 мкл смеси фенол-хлороформ-

изоамиловый спирт (соотношение компонентов в смеси - 25:24:1, соответственно), центрифугировали (12000 об/мин, 4 мин), собирали водную фазу, содержащую ДНК, добавляли 2 объема 96% этанола и выдерживали 30 мин при  $-70^{\circ}\text{C}$ . После центрифугирования (14500 об/мин, 10 мин), удаляли супернатант, осадок промывали 250 мкл 75% этанола, центрифугировали (14500 об/мин, 10 мин), осадок высушивали и растворяли в 20 мкл  $\text{H}_2\text{O}$ .

### **3.2.8. Связывание меченых фрагментов ДНК с белками *in vitro***

Реакционная смесь содержала 2 мкл  $\sim 20$  нг (50000 имп/мин) меченого фрагмента ДНК, 2 мкл (2 мкг) альтернирующего двунитевого гетерополимера poly(dI-dC), 4 мкл TE-буфера, 4 мкл (2 мкг суммарного белка) ядерного экстракта HeLa, 12 мкл буфера В (12 мМ HEPES–KOH, pH 7,9, 12% глицерин, 60 мМ KCl, 0,3 мМ ЭДТА, 0,6 мМ DTT). Белковый образец добавляли в последнюю очередь. После добавления всех реагентов, смесь инкубировали в течение 20 минут при комнатной температуре.

При анализе способности последовательности ДНК конкурировать с меченым фрагментом за связывание с белками ядерного экстракта HeLa вместо TE-буфера добавляли в реакционную смесь  $\sim 60$  нг и  $\sim 120$  нг немеченых фрагментов ДНК.

### **3.2.9. Сдвиг электрофоретической подвижности в полиакриламидном геле (EMSA)**

Эксперименты по сдвигу электрофоретической подвижности проводили по ранее описанной методике [230]. По окончании инкубации (см. выше), смесь, содержащую ДНК-белковые комплексы, наносили на 7,5%-ный полиакриламидный гель (сшивка 30:1) приготовленный на 0,5-кратном буфере TBE. В свободные дорожки геля наносили красители бромфеноловый синий и ксиленцианол для контроля миграции во время проведения электрофореза. Электрофорез проводили при силе тока в 10-12 мА. По прохождении ксиленцианолом  $2/3$  длины геля электрофорез останавливали, гель закрывали



лавсановой пленкой и проводили радиоавтографию в течение 16-20 часов при 4°C.

### 3.2.10. Используемые культуры клеток

Клеточные линии CHO-K1 (CCL-61, клетки яичника китайского хомячка), HeLa (CCL-2, аденокарцинома шейки матки человека), HepG2 (HB-8065, клетки гепатоцеллюлярной карциномы человека), A549 (CCL-185, Легочная аденокарцинома человека), Jurkat (TIB-185, Т-лимфобластная лейкемия человека) культивировали в средах и при условиях, рекомендованных ATCC (American Type Culture Collection) для этих линий (таблица 7). При определении условий пассирования также использовались рекомендации ATCC. Клеточная линия NGP-127 (нейробластома человека) культивировали при условиях, описанных в [231]. Все клеточные линии культивировали при 37°C в атмосфере 5% CO<sub>2</sub>. Для открепления клеток при пассировании всех адгезивных культур использовали обработку 0,25% трипсином-версеном.

**Таблица 7.** Краткая характеристика использованных клеточных линий

<i>Название, номер по каталогу ATCC</i>	<i>Происхождение линии клеток</i>	<i>Культуральная среда</i>
CHO-K1 CCL-61	Яичник китайского хомячка	DMEM/F12+ 10% FCS
HeLa CCL-2	Аденокарцинома шейки матки человека	DMEM/F12 + 10% FCS
HepG2 HB-8065	Гепатоцеллюлярная карцинома человека	DMEM/F12 + 10% FCS
A549 CCL-185	Легочная аденокарцинома человека	DMEM/F12 + 10% FCS

<i>Название, номер по каталогу ATCC</i>	<i>Происхождение линии клеток</i>	<i>Культуральная среда</i>
Jurkat TIB-185	Т-лимфобластная лейкемия человека	RPMI 1640 + 10% FCS
NGP-127	Нейробластома человека	RPMI 1640 + 10% FCS

### **3.2.11. Трансфекция клеток с использованием реагента Lipofectamine 2000**

Трансфекцию клеток проводили в 24-луночных планшетах с использованием реагента липофектамин 2000 («Invitrogen») согласно рекомендациям производителя. В опытах по исследованию энхансерной и энхансер-блокирующих активностей для трансфекции использовали 200 нг предварительно подготовленной плазмидной ДНК и 75 нг нормировочной плазмиды рRL-ТК («Promega») на лунку. В опытах по исследованию *транс* активации для каждой трансфекции в 20 мкл ТЕ-буфера готовили смесь плазмидной ДНК, состоящей из 40 нг плазмиды с промотором и 3- или 10-кратным мольным избытком плазмиды, содержащей энхансер, добавляли 75 нг плазмиды рRL-ТК (Promega) и доводили общее количество ДНК до 500 нг необходимым количеством плазмиды, содержащей фрагмент ДНК фага лямбда.

Для трансфекции использовали адгезивные культуры клеток аденокарциномы человека HeLa, яичника хомячка CHO-K1 и клетки гепатоцеллюлярной карциномы человека HepG2. Клетки за 24 часа до трансфекции высевали в количестве  $70-100 \times 10^3$  на лунку 24-луночного планшета. Трансфекцию клеток с помощью реагента Lipofectamine 2000 осуществляли согласно рекомендациям фирмы-производителя (Invitrogen, США)

### **3.2.12. Получение клеточного лизата и измерение активности люциферазы**

После проведения трансфекции среду из лунок удаляли, клетки промывали 1 мл однократного PBS-буфера, затем добавляли 100 мкл однократного PLB-буфера (Promega, США) и инкубировали с перемешиванием в течении 10 мин. Клеточный лизат центрифугировали 10 мин при 5000 об/мин, 10 мкл лизата брали на измерение активности люциферазы.

Измерение активности люциферазы в системе двойной люциферазной детекции (Dual-Luciferase® Reporter Assay System) проводили согласно инструкции производителя с использованием автоматического люминометра Tecan GENios Pro. Обработка результатов проводилась с использованием программы Microsoft Office 2007 Excel.

### **3.2.13. 3С анализ**

#### ***Приготовление эквимольной смеси продуктов лигирования на основе смеси плазмидной ДНК***

Смесь продуктов лигирования (RLT-random ligation template) [232] получали путем гидролиза эквимольной смеси плазмид Pcmv, Ecmv и лямбда с эндонуклеазой рестрикции *Msp* I и последующим лигированием полученных фрагментов в условиях высокой концентрации ДНК. Эффективность пар праймеров в ПЦР реакции определяли путем амплификации серийных разведений матрицы и последующего вычисления с использованием программы LinRegPCR [233]. Полученный образец использовали в качестве стандарта при определении относительных количеств продуктов лигирования в экспериментах 3С- анализа.

#### ***Фиксация клеток и изоляция ядер***

Метод фиксации конформации хромосом (3С) и количественную ПЦР проводили, согласно протоколу [234]. Приблизительно  $10^6$  клеток HepG2

трансфицировали в течение 24 ч. с 10 мкг эквимольной смеси плазмид Pcmv, Ecmv и L, после чего проводили сшивку в течение 10 минут при комнатной температуре в среде, содержащей 1% формальдегида. Реакцию останавливали с помощью раствора 2 М глицина, клетки лизировали в 5 мл NP-40-буфера (10 мМ Трис-НCl, pH 8,0, 10 мМ NaCl, 0,15% NP-40, 0,15 мМ AEBSF), промывали и ресуспендировали в том же буфере без NP-40.

### ***Рестрикция, лигирование и очистка ДНК***

К выделенным ядрам добавляли SDS до конечной концентрации 0,3%, и инкубировали полученную суспензию при 37 °С в течение 1 ч. Инактивировали SDS, путем добавления детергента тритон X-100 до конечной концентрации 2,5%, и дополнительно инкубировали суспензию при 37 °С в течение 1 ч. Затем ядра обрабатывали высококонцентрированной эндонуклеазой рестрикции *Msp* I в растворе 10 мМ Трис-НCl, pH 7.9, 50 мМ NaCl, 10 мМ MgCl<sub>2</sub>, 1 мМ DTT (NEB2 буфер), с последующим лигированием, как описано в [20], и последовательно для удаления сшивок обрабатывали РНКазой А с концентрацией 30 мкг/мл в течение 30 мин при температуре 37 °С и протеиназой К с концентрацией 30 мкг/мл при 65 °С в течение ночи. ДНК экстрагировали смесью фенол-хлороформ, осаждали этанолом, и использовали для ПЦР-амплификации в режиме реального времени.

Эффективность рестрикции *Msp* I измеряли для всех плазмид, как описано [234] с использованием пар праймеров (Ecmv1, Ecmv2, Lambda, Pcmv1, Pcmv2) флануирующих рестриктные сайты и пар праймеров (Ecmv control, Lambda control, Pcmv control) к внутреннему фрагменту, не содержащему рестриктный сайт *Msp* I, в соответствующих плаزمидях. Во всех случаях эффективность была больше 85%. Последовательности пар праймеров приведены в таблице 6.

Отрицательный контроль (без сшивки) был выполнен в условиях, описанных выше, но без добавления формальдегида.

### ***Количественный анализ продуктов лигирования с использованием ПЦР в режиме реального времени***

Количественную ПЦР проводили с использованием набора реактивов qPCRmix-HS SYBR (Evrogen) и системы LightCycler 480 (Roche, США) в соответствии с руководством производителя. Фрагменты ДНК амплифицировали в течение 40 циклов при 95°C в течение 20 с, 63°C в течение 30 сек, 72°C в течение 20 с. Фрагменты амплифицированной ДНК были дополнительно проанализированы на однородность получаемого продукта при помощи электрофореза в 1,2% агарозном геле и анализом кривой плавления. Во всех реакциях был обнаружен один ПЦР-продукт ожидаемой длины (351 и 430 п.н. для продукта лигирования промотор-контроль и промотор-энхансер соответственно).

Относительное содержание продуктов лигирования промотор-энхансер и промотор- контрольный фрагмент определяли с помощью ПЦР в реальном времени следующим образом. Количество циклов, необходимых для амплификации продуктов лигирования в RLT матрице до порогового значения наносили на график против последовательных разведений RLT. Используя логарифмическую линию тренда для данного графика (коэффициент корреляции  $R^2 > 0,97$ ), были определены эффективные разведения, соответствующие продуктам лигирования промотор-контроль и промотор-энхансер, относительное содержание (Cr) каждого продукта рассчитывали, используя следующую формулу:

$$Cr = 1000 / D$$

Где:

D - эффективное разведение.

### 3.2.14. Определение количества трансфицированных плазмид в ядрах

Клетки HepG2 трансфицировали эквимольной смесью плазмид Pcmv, Ecmv и лямбда, сшивали, и выделяли ядра с использованием буфера NP-40, как описано выше (раздел 3.2.13 «3С анализ»). Затем ядерную ДНК экстрагировали смесью фенол-хлороформ, осаждали этанолом и использовали в качестве матрицы для ПЦР в реальном времени с праймерами Ecmv control и Lambda control (таблица 6), специфичными для плазмид Ecmv и L. Эквимольную смесь плазмид использовали в качестве контрольной матрицы. Нормированное количество ( $X_N$ ) каждой плазмиды вычисляли с использованием следующей формулы:

$$X_N = (1 + E)^{-dCt},$$

Где:

E - эффективность ПЦР, и dCt разница между пороговым числом циклов для амплификации ядерной ДНК и контрольной смеси.

## 4. РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

### 4.1. ФУНКЦИОНАЛЬНАЯ ДИСЕКЦИЯ ЭНХАНСЕРНОГО ЭЛЕМЕНТА, РАСПОЛОЖЕННОГО ВО ВТОРОМ ИНТРОНЕ ГЕНА *U2AF1L4* ЧЕЛОВЕКА

Ранее в нашей лаборатории был идентифицирован ряд коротких (200–500 п.н.) геномных последовательностей, проявляющих энхансерную активность в составе ретровирусной конструкции, интегрированной в геном клеток HeLa [13]. Обнаруженные последовательности были картированы в полигенном участке 19q13.12 хромосомы 19 человека длиной 1 млн п.н. Также для идентифицированных последовательностей была проверена энхансерная активность в составе вектора pGL3PV с использованием системы двойной люциферазной детекции в клетках линии HeLa [13].

Используя систему транзientной экспрессии репортерного гена, мы провели функциональный анализ одного из наиболее сильных обнаруженных энхансерных элементов (энхансер 12), расположенного во втором интроне гена *U2AF1L4* (рисунок 4).

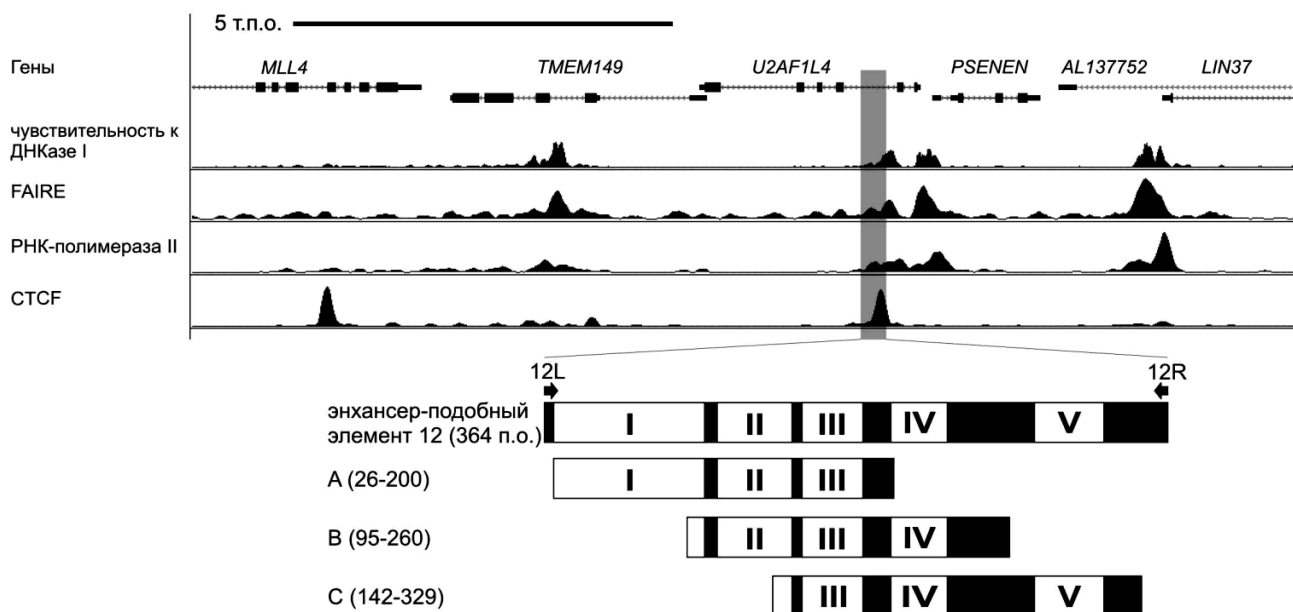
В непосредственной близости от энхансера 12 расположена промоторная область генов *PSENEN* и *U2AF1L4*. Отличительной особенностью данной промоторной области является то, что она регулирует функционально неродственные и пространственно разделенные продукты. Белок PSENEN является компонентом комплекса гамма-секретазы, который также включает в себя пресенилин и никастрин [235]. Гамма-секретаза необходима для внутримембранного протеолиза ряда мембранных белков, в том числе белка предшественника бета-амилоида (APP) и белка Notch. PSENEN является политопическим мембранным белком, который преимущественно локализуется в плазматической мембране, мембран эндоплазматического ретикулума и мембранах аппарата Гольджи. Несмотря на широкое биологическое значение фермента, данные о регуляции транскрипции компонентов комплекса гамма-секретазы отсутствуют.

Ген *U2AF1L4* кодирует белок U2AF26 связывающий РНК, который входит в состав гетеродимерного комплекса, взаимодействующего с мРНК U2 и обеспечивающего распознавание сплайсосома 3'-сплайс сайтов в пре-мРНК [236, 237]. Ген *U2AF1L4* экспрессируется тканеспецифично [236, 238], о регуляции его экспрессии на транскрипционном уровне также ничего не известно. Возможно, что обсуждаемый энхансерный элемент является одним из энхансеров, участвующих в регуляции промоторной области генов *PSENFEN* и *U2AF1L4*. Поэтому данный энхансер был подвергнут более детальному функциональному анализу.

#### **4.1.1. Анализ активности фрагментов энхансера 12**

Согласно данным геномного браузера UCSC, в участке, содержащем последовательность энхансера 12, наблюдается кластеризация сайтов связывания белков, таких как инсуляторный белок CTCF и компоненты комплекса SWI/SNF (BAF155, BAF170 и Ini1) [239]. Кроме того, полногеномный анализ сайтов чувствительности к ДНКазе I, а также метод картирования регуляторных элементов, основанный на выявлении свободных от нуклеосом фрагментов ДНК (FAIRE), выполненные в рамках проекта ENCODE, подтверждают присутствие во втором интроне гена *U2AF1L4* ДНК белковых комплексов [240-245].





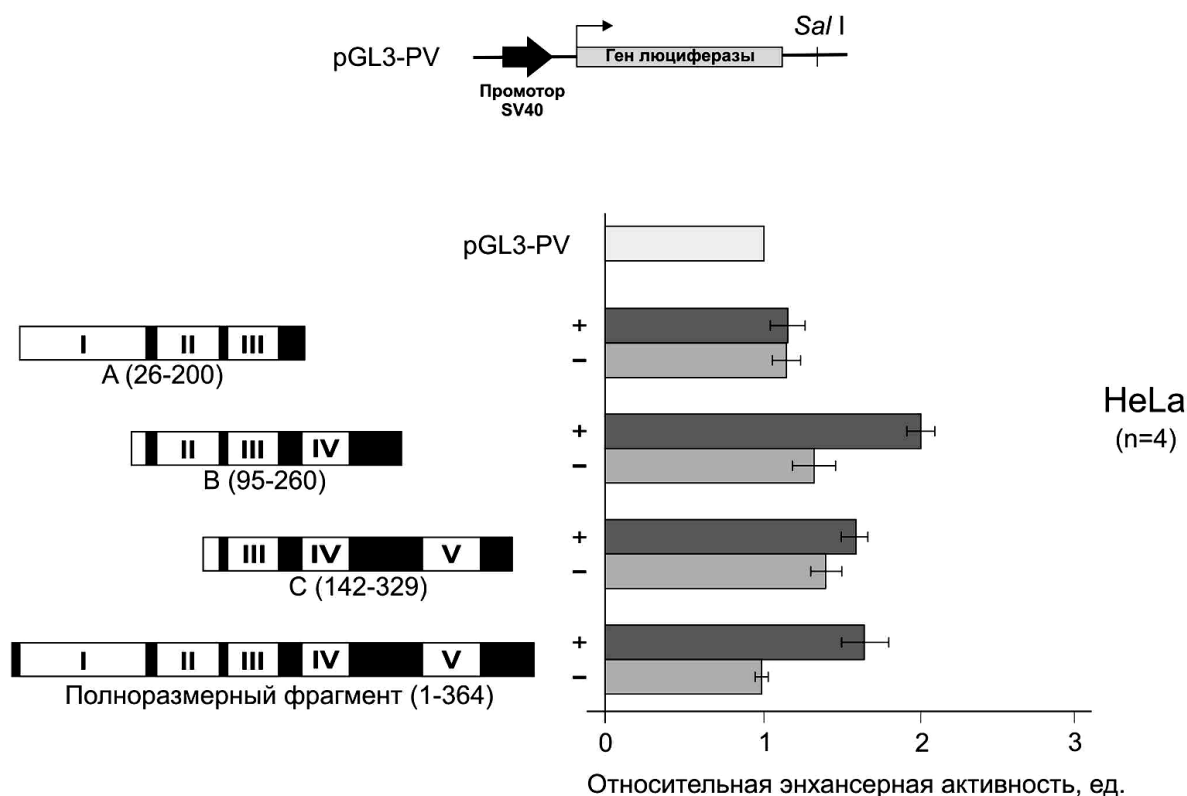
**Рисунок 4.** Расположение энхансера 12 относительно ближайших генов на хромосоме 19 человека. Представлены данные из геномного браузера UCSC (<http://genome.ucsc.edu/goldenPath/hgTracks.html>) по плотности распределения картированных участков чувствительности к ДНКазе I, регуляторных участков, обнаруженных с помощью метода выявления свободных от нуклеосом фрагментов ДНК (FAIRE – Formaldehyde Assisted Isolation of Regulatory Elements), участков, взаимодействующих с РНК полимеразой II и инсульторным белком CTCF, в геноме клеток линии HeLa. Серым цветом отмечена область, содержащая энхансер 12. Внизу показаны полноразмерный энхансер 12 и его перекрывающиеся делеционные фрагменты (А, В и С), для которых проведен функциональный анализ энхансерной активности в клеточной линии HeLa (рисунок 5). Римскими цифрами обозначены консервативные участки внутри элемента у приматов.

С помощью пары праймеров 12L и 12R была амплифицирована область, содержащая исследуемый фрагмент. Амплифицированную последовательность длиной 364 п.н. (координаты по UCSC Genome Browser, сборка 2006 г. – chr19:40927383 40927746) клонировали в вектор pGEM-T («Promega») и секвенировали.

Для того чтобы определить, какой из участков энхансера 12 отвечает за его энхансерную активность, были выбраны три перекрывающихся фрагмента А, В и С (рисунок 5). В качестве матрицы для амплификации перекрывающихся делеционных фрагментов и его полноразмерной последовательности использовали вектор pGEM-T, содержащий

последовательность энхансера 12. Продукты амплификации разделяли электрофоретически в 1% агарозном геле и элюировали целевые фрагменты ДНК с помощью набора Wizard SV Gel и PCR Clean Up System («Promega», США). Очищенные фрагменты после предварительной обработки, линейаризованной плазмиды фрагментом Кленова клонировали по сайту *Sal I* в вектор pGL3PV («Promega»), содержащий ген люциферазы светлячка под контролем промотора SV40 (рисунок 5).

Для характеристики активностей энхансера 12 и его фрагментов с помощью полученного набора плазмид, а также исходного вектора pGL3PV, используемого в качестве контрольного, была проведена серия транзientных трансфекций клеток HeLa (рисунок 5). Максимальную активность проявлял средний фрагмент В, усиливая активность промотора SV40 в два раза, что указывает на наличие в этом фрагменте участка, ответственного за энхансерную активность. Интересно, что фрагмент С, а, следовательно, и полноразмерный энхансер содержат участок связывания инсульторного белка CTCF [60, 246]. Известно, что CTCF в некоторых случаях взаимодействует с сайленсерами, а следовательно, может снижать активность пары промотор–энхансер [247], поэтому его присутствие может приводить к некоторому снижению активности полноразмерной последовательности энхансера и его фрагмента С. Фрагмент А не обладает энхансерной активностью в данной системе, поэтому с учетом перекрывания данного фрагмента с фрагментом В можно предполагать, что ядро энхансерного элемента расположено в 3' области фрагмента В. Наблюдаемый невысокий уровень энхансерной активности, возможно, связан с тем, что в клетках HeLa сильный промотор SV40 слабо подвержен модуляции со стороны клеточных энхансеров и только в присутствии сильных энхансеров, например вирусного энхансера SV40 [248], или энхансеров, расположенных в LTR эндогенных ретровирусов [249], его активность может значительно повышаться.



**Рисунок 5.** Определение энхансерной активности в клеточной линии HeLa для полного энхансера 12 и его фрагментов в прямых (+) и обратных (-) ориентациях с использованием системы двойной люциферазной детекции («Promega»). Исследуемые фрагменты клонированы 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 ( $n = 4$ ). Римскими цифрами обозначены консервативные участки у приматов

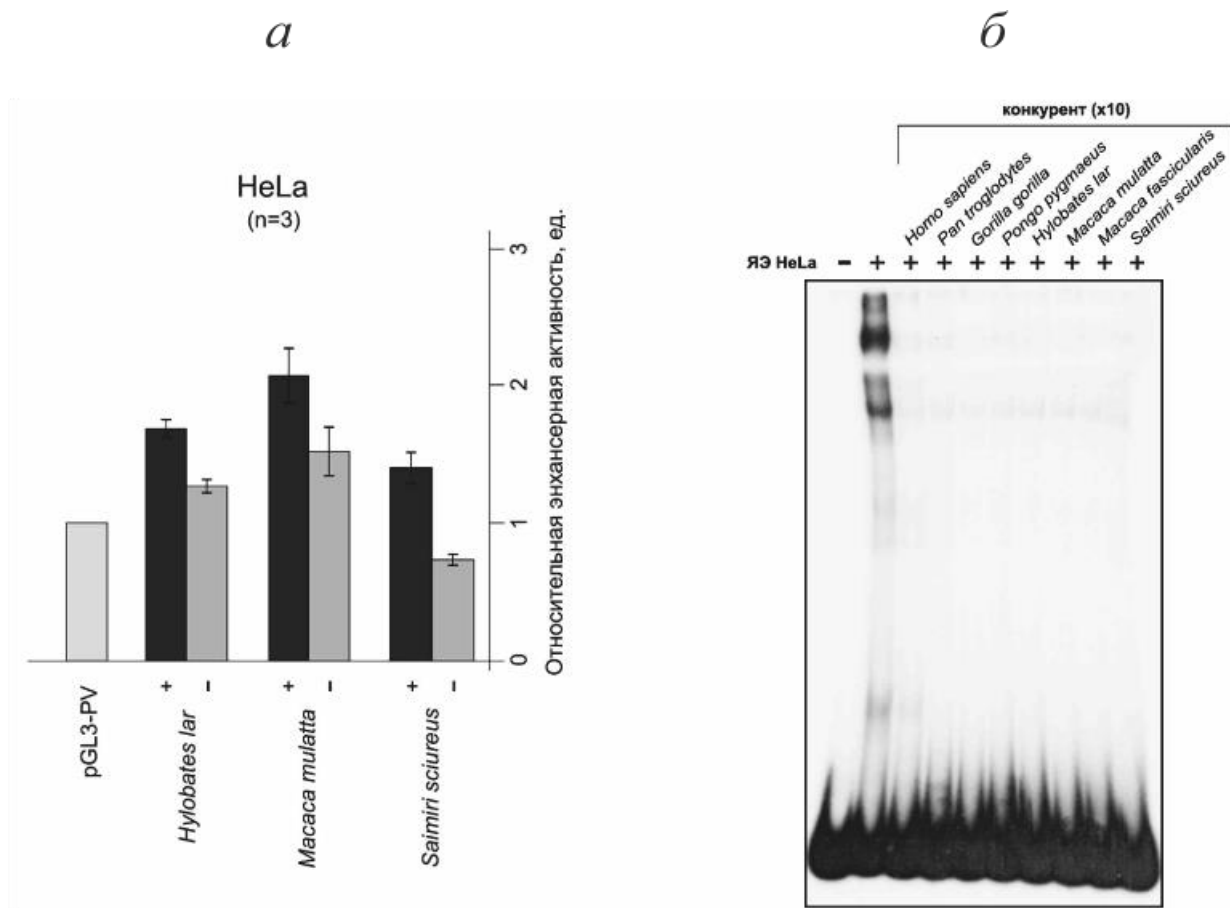
Активность исследованных последовательностей зависит от их ориентации относительно промотора, что не характерно для классических энхансеров. Полноразмерный энхансер 12, а также его фрагменты B и C проявляли наибольшую активность, находясь в прямых ориентациях (рисунок. 5).

#### 4.1.2. Анализ функциональной консервативности энхансера 12

С помощью универсальной пары праймеров (12L и 12R) были амплифицированы и клонированы последовательности, ортологичные энхансеру 12, из геномов семи видов приматов. Эти последовательности

представляют собой формы энхансера, возникшие в ходе естественного мутагенеза в процессе эволюции. Измерение энхансерной активности трех из них (*Hylobates lar*, *Macaca mulatta*, *Saimiri sciureus*) в прямой и обратной ориентациях указывает на то, что появившиеся мутации незначительно влияют на активность фрагментов (рисунок 6а), и свидетельствует о сохранении активности мутантных форм энхансера 12 в клетках человека.

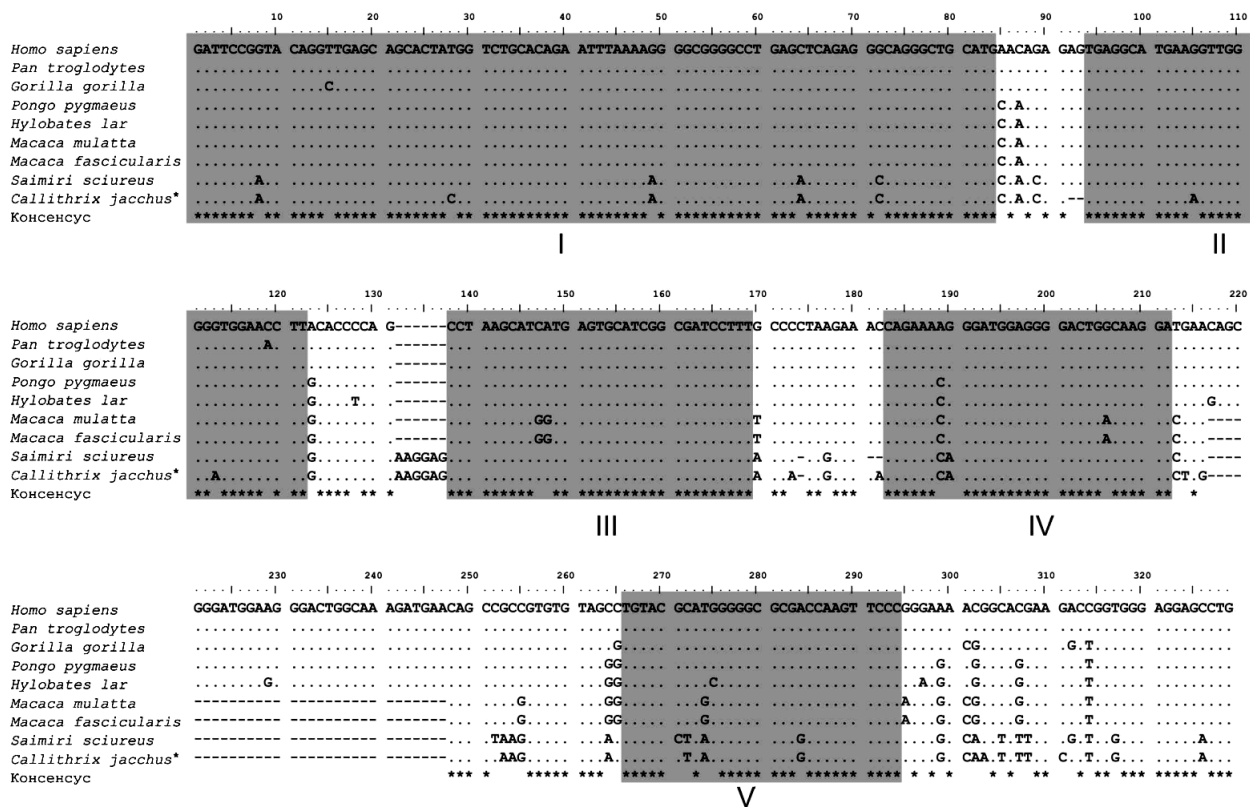
Для того чтобы выявить консервативные области в энхансере 12 нами было проведено выравнивание последовательностей энхансера 12 человека и ортологичных ему последовательностей приматов, что позволило выделить пять наиболее консервативных участков внутри энхансера 12 (I–V, рисунок 7). Известно, что мутации, затрагивающие сайты связывания транскрипционных факторов, могут приводить к нарушению ДНК-белковых взаимодействий и, следовательно, к изменению или полной утрате активности регуляторного элемента. Для проверки этой гипотезы была проверена способность последовательностей приматов взаимодействовать с белками ядерного экстракта клеток HeLa с помощью сдвига электрофоретической подвижности. Добавление немеченых последовательностей генома приматов в реакцию связывания энхансера 12 человека с белками ядерного экстракта HeLa приводит к ослаблению интенсивности зон торможения, что указывает на способность последовательностей приматов взаимодействовать с теми же белками, что и энхансер 12 (рисунок 6б). Таким образом, нуклеотидные замены, делеции и инсерции в исследованных последовательностях приматов не влияют на ДНК-белковые взаимодействия *in vitro*. Можно предположить, что консервативные области энхансера 12 и области с наименьшим числом межвидовых отличий ответственны за связывание с белками ядерного экстракта клеток HeLa и, вероятно, за проявление энхансерной активности исследуемой области генома человека.



**Рисунок 6.** Анализ функциональных свойств последовательностей приматов, ортологичных энхансеру 12. а – Энхансерная активность последовательностей гиббона (*Hylobates lar*), макаки резус (*Macaca mulatta*) и саймири (*Saimiri sciureus*) в прямой (+) и обратной (–) ориентациях в клетках HeLa с использованием системы двойной люциферазной детекции. Значения активности приведены относительно активности промотора SV40 (pGL3PV). Показана стандартная ошибка среднего для трех независимых экспериментов (n = 3); б – EMSA анализ способности последовательностей приматов, гомологичных энхансеру 12, конкурировать с энхансером 12 за связывание с белками ядерного экстракта HeLa (ЯЭ HeLa). Конкурирующие последовательности добавлялись в реакционную смесь в десятикратном избытке относительно меченого фрагмента

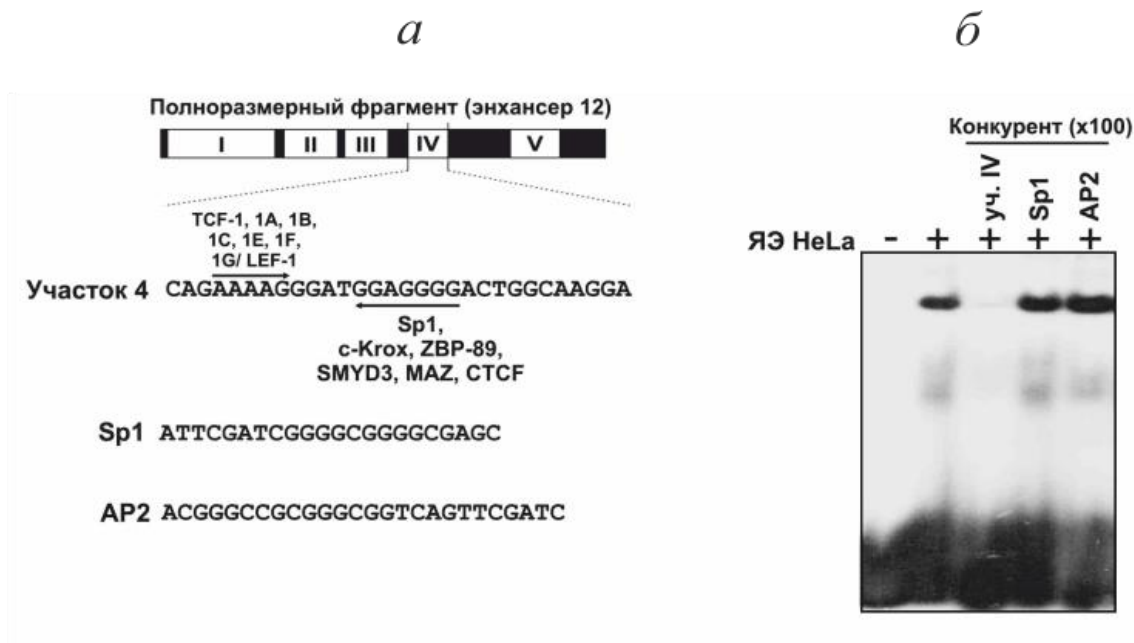
Как видно на рисунке 5, фрагмент В в отличие от фрагмента А, не проявляющего энхансерной активности, содержит консервативный участок IV длиной 30 п.н., расположенный в 3' области фрагмента. Было выдвинуто предположение, что именно этот участок отвечает за энхансерную активность. С помощью программы TESS, использующей базы данных TRANSFAC, JASPAR, IMD и CBIL GibbsMat, выявлены гипотетические сайты связывания

известных факторов транскрипции человека. Фрагмент содержит участок связывания факторов транскрипции семейства TCF-1 (T-клеточный фактор 7) и энхансерного фактора лимфоцитов LEF 1. Однако в данном сайте содержится моонуклеотидная замена у четырех из исследованных видов приматов и динуклеотидная замена у двух видов, которые могут влиять на ДНК белковые взаимодействия. В то же время сайт связывания активатора транскрипции SP1 CCCCTCC не содержит межвидовых отличий. Была проверена возможность взаимодействия консервативного участка IV с факторами SP1 и AP 2 с помощью метода сдвига электрофоретической подвижности с использованием ядерного экстракта клеток HeLa. В качестве зонда использовали двухцепочечный олигонуклеотид, содержащий последовательность участка IV, а в качестве конкурентов – олигонуклеотиды с консесусным сайтом связывания факторов SP1 (GC бокс) и активатора транскрипции AP 2 (рисунок 8а). Добавление конкурентов в реакцию не снижает интенсивности зоны торможения (рисунок 8б), что указывает на то, что консервативный участок IV не взаимодействует с этими факторами *in vitro*. В то же время добавление в реакцию немеченой консервативной области IV в 100 кратном избытке приводит к исчезновению зоны торможения (рисунок 8б).



**Рисунок 7.** Выравнивание последовательности энхансера 12 и ортологичных последовательностей из геномов восьми видов приматов с помощью программы ClustalW. Выделены серым цветом и обозначены римскими цифрами (I–V) консервативные области с наименьшим количеством межвидовых отличий. Последовательность *Callithrix jacchus* взята из базы данных UCSC Genome Browser

Стоит отметить, что с последовательностями, подобными CCCCTCC мотиву, помимо SP 1, могут взаимодействовать и другие факторы с различными активностями, например, с Krox, ZBP 89, SMYD3, MAZ, CTCF, что указывает на возможное влияние специфического геномного окружения, определяющего, какой из факторов взаимодействует с данным мотивом.



**Рисунок 8.** Взаимодействие консервативного участка IV энхансера 12 с белками ядерного экстракта HeLa *in vitro*. а – последовательность консервативного участка IV и консенсусные мотивы для связывания этим участком ряда транскрипционных факторов. Ниже приведены последовательности олигонуклеотидных конкурентов, содержащих участки связывания факторов Sp1 и AP2; б – EMSA анализ связывания консервативного участка IV с белками ядерного экстракта HeLa. Конкурирующие последовательности, содержащие сайты связывания факторов Sp1 и AP 2, и немеченый консервативный участок IV добавляли в реакционную смесь в стократном избытке относительно меченого фрагмента

Выявление короткого участка, ответственного за функциональную активность энхансера 12, позволит в дальнейшем идентифицировать связывающиеся с ним транскрипционные факторы методом ДНК аффинной хроматографии.



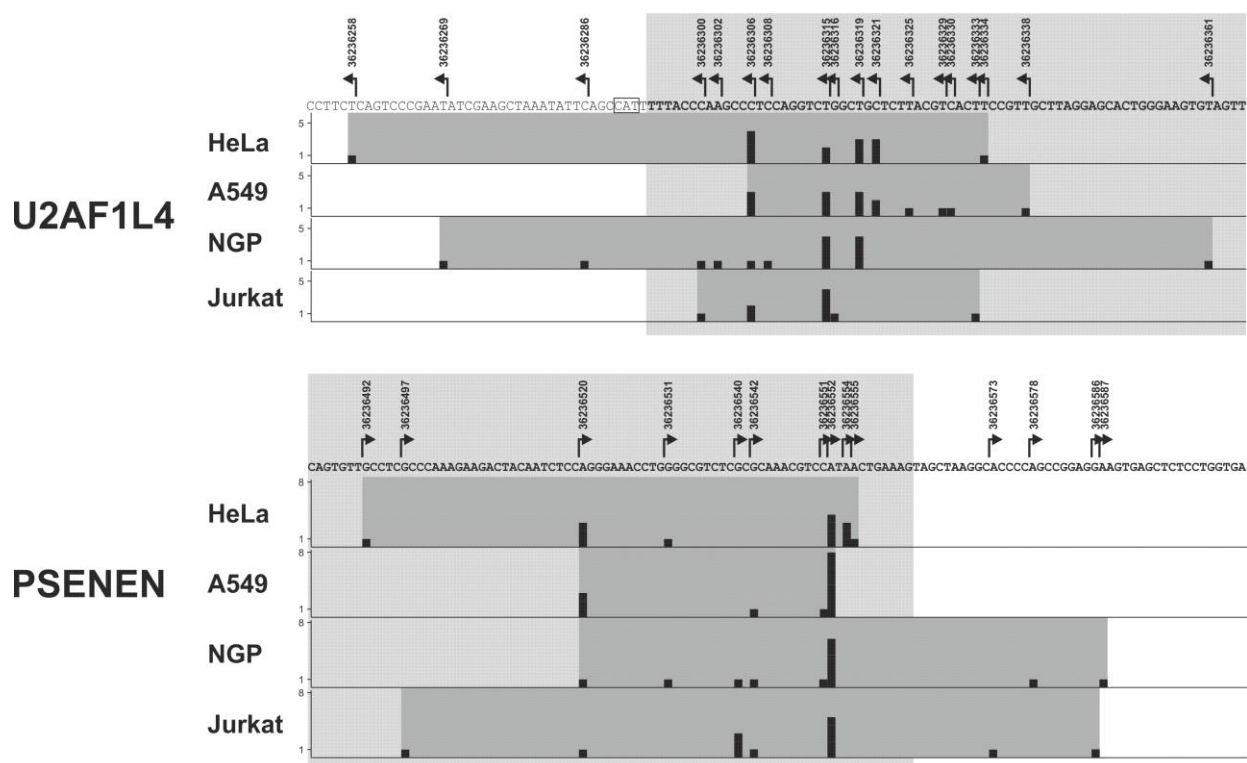
## **4.2. ИЗУЧЕНИЕ ДВУНАПРАВЛЕННОГО ПРОМОТОРА ГЕНОВ *PSENNEN* И *U2AF1L4***

В связи с тем, что проанализированный энхансер 12 расположен во втором интроне гена *U2AF1L4*, нами было сделано предположение, согласно которому промоторная область данного гена находится под контролем энхансера 12. Для этого исследовали промоторную область двух близкорасположенных генов *PSENNEN* и *U2AF1L4*.

Гены *PSENNEN* и *U2AF1L4* являются разнонаправленными генами и расположены на хромосоме 19 человека таким образом, что их первые экзоны разделены участком длиной 157 п.н.. Таким образом, общая промоторная область имеет характеристики потенциального двунаправленного промотора.

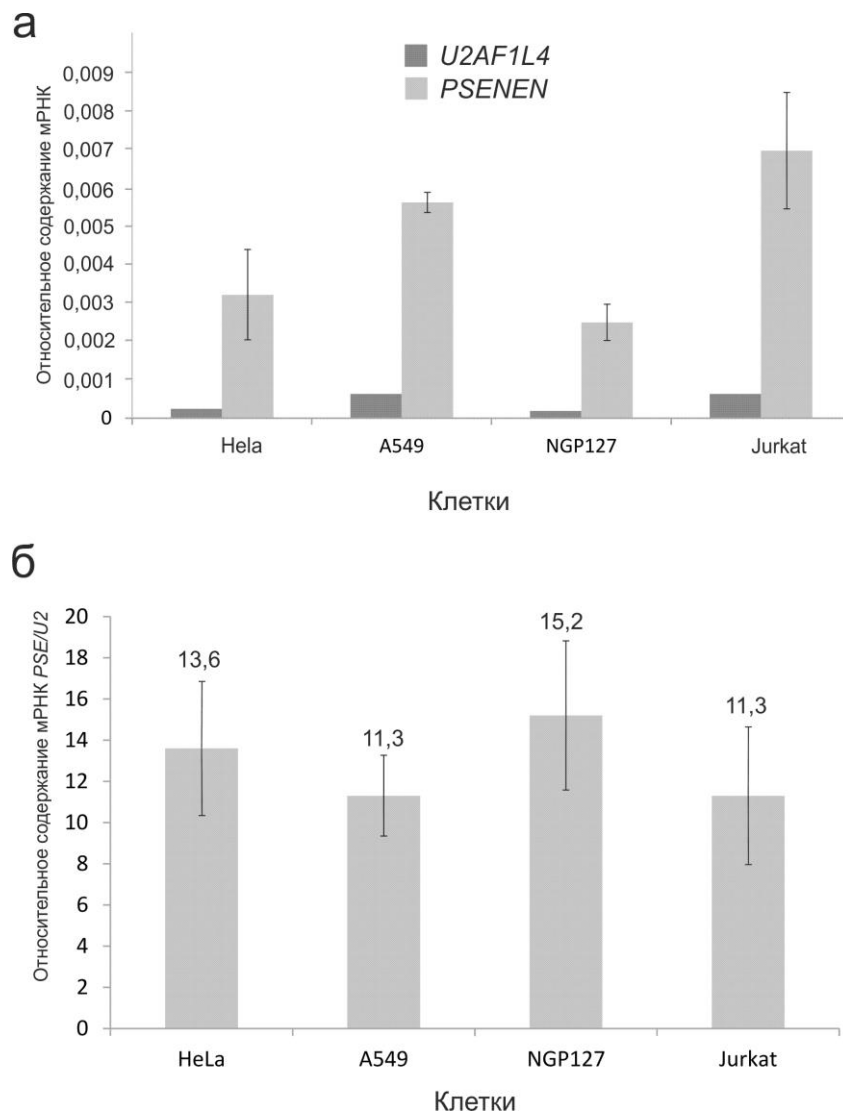
### **4.2.1. Точки инициации транскрипции и тканевая специфичность транскрипции генов *PSENNEN* и *U2AF1L4***

С помощью методики 5'-RACE SMART [250], нами были определены позиции 5'-концов мРНК для обоих генов *PSENNEN* и *U2AF1L4* в четырех клеточных линиях (HeLa, A549, NGP-127 и Jurkat). Для каждой клеточной линии выделяли и секвенировали от тринадцати до пятнадцати клонов. Рисунок 9 показывает распределение сайтов инициации транскрипции (TSS), которые были обнаружены для обоих генов. Как видно из рисунка, транскрипция в обоих направлениях характеризуется диффузным распределением TSS, однако можно выделить несколько предпочтительных участков для обоих генов. Согласно полученным данным большинство мРНК преимущественно иницируются в обоих направлениях на динуклеотидах пиримидин-пурин (-1+1), и принадлежат к классу с широким распределением TSS [251]. Тем не менее, нельзя исключить, что некоторые сайты инициации транскрипции *U2AF1L4* были потеряны из-за наличия нескольких форм сплайсинга для данного гена.



**Рисунок 9.** Распределение 5'-концов мРНК внутри промоторной области генов *PSENER* и *U2AF1L4*. Высота черных столбцов пропорциональна количеству транскриптов в данной точке инициации транскрипции для заданной клеточной линии. Светло-серая область отображает участок, перекрывающийся с клонированной последовательностью длиной 269 п.н., темно-серые области отображают всю область в которой были найдены точки инициации транскрипции.

С помощью количественной ПЦР с обратной транскрипцией, было оценено содержание мРНК генов *PSENER* и *U2AF1L4* в четырех клеточных линиях различного происхождения: HeLa, A549, NGP-127 и Jurkat. Содержание мРНК нормализовали к среднему геометрическому содержания мРНК *GAPDH*, *RPLP0* и генов 18S рРНК. Согласно полученным данным (рисунок 10а), содержание мРНК для обоих генов характеризуется умеренной тканевой специфичностью и изменяется в 2-3 раза в различных клеточных линиях.



**Рисунок 10.** Относительное содержание мРНК генов *PSENE1* и *U2AF1L4* в четырех клеточных линиях. Содержание мРНК нормировано к среднему геометрическому содержания мРНК *GAPDH*, *RPLP0* и генов 18S рРНК. б-относительное содержание мРНК *PSENE1* к *U2AF1L4* в четырех линиях клеток.

Кроме того, во всех клетках содержание мРНК *PSENE1* от 10 до 15 раз выше, чем мРНК *U2AF1L4*. Тем не менее, уровень мРНК гена *U2AF1L4* может быть несколько заниженным из-за наличия сплайс форм, не обнаруженных с помощью используемых в работе праймеров. В то же время, относительное содержание мРНК двух генов во всех клетках очень близко (рисунок 10б), что можно объяснить согласованной регуляцией транскрипции. Полученные данные согласуются с идеей того, что транскрипция обоих генов находится под контролем общего двунаправленного промотора

#### 4.2.2. Анализ активности промоторной области генов *PSENNEN* и *U2AF1L4*

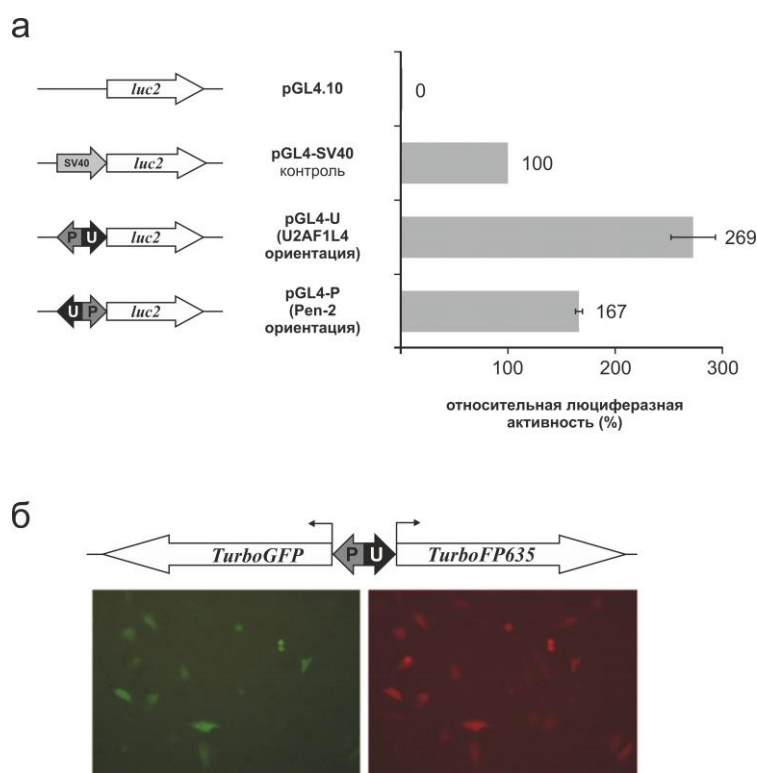
Несмотря на то, что область длиной 238 п.н., содержащая минимальный промотор гена *PSENNEN*, была идентифицирована ранее, ее активность в направлении гена *U2AF1L4* изучена не была [252]. Стоит отметить, что идентифицированный ранее минимальный промотор длиной 238 п.н. включал в себя часть кодирующей последовательности гена *U2AF1L4* вместе со стартовым кодоном АТГ и не содержал область инициации транскрипции гена *PSENNEN*, что не позволяет считать полученные ранее данные полными. Также было показано, что активность промотора в клетках линии HeLa является значительной [252].

Для того чтобы учесть в своем исследовании активность промотора в обоих направлениях, мы клонировали в прямой и обратной ориентации фрагмент длиной 269 п.н. (U2P-промотор), содержащий начальные части первых экзонов обоих генов *PSENNEN* и *U2AF1L4*, в плазмиду pGL4.10 на место промотора, выше гена люциферазы. Для этого промоторную область генов *PSENNEN* / *U2AF1L4* амплифицировали с использованием геномной ДНК-матрицы, праймеров PR1 и PR2 (таблица 2) и Pfu Ultra II полимеразы (Stratagene), в условиях, рекомендованных производителем. Целевой амплифицированный фрагмент длиной 269 п.н. очищали с помощью электрофореза в 1% агарозном геле и элюировали его с помощью набора SV Gel и PCR System Clean-Up (Promega). Плазмиду pGL4.10 (Promega), гидролизовали с помощью эндонуклеазы рестрикции *Nco* I, обрабатывали ферментом Кленова для получения тупых концов, и полученный очищенный фрагмент длиной 269 п.н. клонировали в промоторную область гена люциферазы.

Таким образом была получена плазида PGL4-P, содержащая фрагмент промотора в ориентации гена *PSENNEN* и PGL4-U, содержащая тот же фрагмент в *U2AF1L4* ориентации (рисунок 11а). Клонированная последовательность включала в себя большую часть обнаруженных TSS в

четырёх клеточных линиях. Полученными конструкциями трансфицировали клетки HeLa совместно с нормировочной плазмидой pRL-ТК, содержащей ген люциферазы *Renilla reniformis*, и определяли относительную активность люциферазы (рисунок 11а). Мутантные формы промотора были получены с помощью сайт-специфичного мутагенеза, согласно руководству [229].

Согласно полученным результатам (рисунок 11), активность фрагмента, содержащего промотор, в обеих ориентациях в клетках HeLa была выше активности минимального промотора SV40 и в 1,6 раза выше в направлении *U2AF1L4*, чем в обратной ориентации для гена *PSENNEN*.



**Рисунок 11.** а-активность промоторной области расположенной между генами *PSENNEN* и *U2AF1L4* в клетках линии HeLa трансфицированных плазмидой pGL4.10, не содержащей промотор, pGL4.10-SV40, содержащей промотор SV40, pGL4.10-U, содержащей промоторную область *PSENNEN/U2AF1L4* в ориентации гена *U2AF1L4* pGL4.10-P содержащей промоторную область *PSENNEN/U2AF1L4* в ориентации гена *PSENNEN*. б- одновременная экспрессия флуоресцентных белков TurboGFP и TurboFP635 под контролем промотора *PSENNEN/U2AF1L4* в клетках линии HeLa.

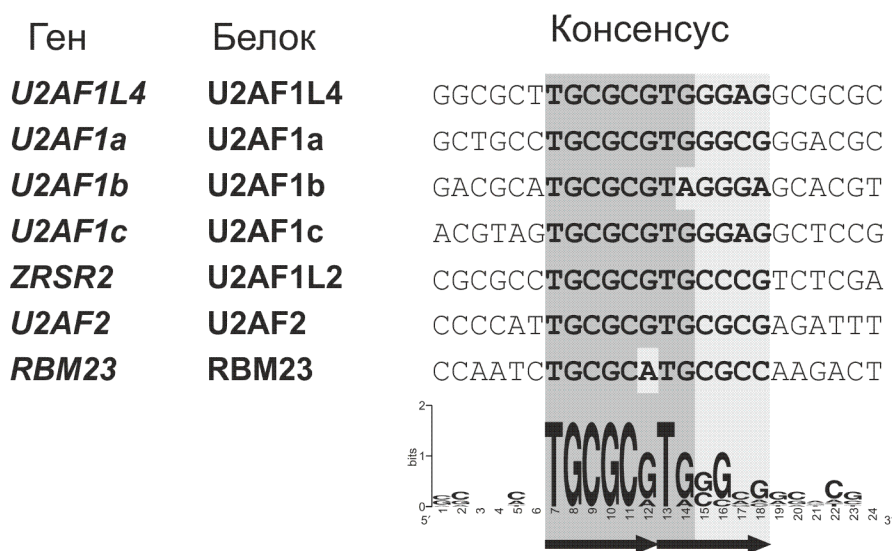
Для того чтобы продемонстрировать, что U2P-промотор может обеспечивать одновременную экспрессию обоих генов, тот же фрагмент длиной 269 п.н. клонировали между генами флуоресцирующих белков TurboGFP [253] и TurboFP635 [254] в вектор pGL4-tGFP-tFP635.

Вектор pGL4-tGFP-tFP635 был получен на основе плазмид pGL4.10 (Promega), pTurboFP635-N (Evrogen) и pTurboGFP-N (Evrogen) как описано ниже. Для получения фрагмента, содержащего ген *TurboFP635*, плазмиду pTurboFP635-N обрабатывали эндонуклеазами рестрикции *Nco* I и *Xba* I, полученный фрагмент длиной 831 п.н. лигировали в вектор pGL4.10 гидролизованной по сайту *Xba* I (концы были достроены фрагментом Кленова) и *Nco* I, в результате чего, была получена плазида pGL4-tFP635. pGL4-tGFP. Плазмиду конструировали путем введения фрагмента длиной 877 п.н., полученного путем обработки плазмиды pTurboGFP-N эндонуклеазами рестрикции *Hind* III и *Xba* I, в плазмиду pGL4.10 по сайту *Xba* I (был заполнен фрагментом Кленова) и *Hind* III. Затем обрабатывали плазмиду pGL4-tGFP эндонуклеазой рестрикции *Bam* HI и выделяли фрагмент длиной 1105 п.н. и лигировали его по сайту *Bgl* II в плазмиду pGL4-tFP635 таким образом, что гены флуоресцентных белков TurboGFP и TurboFP635 были расположены голова к голове. Плазида pGL4-tGFP-U2P-tFP635 была получена из плазмиды pGL4-tGFP-tFP635 путем клонирования 269 п.н.-фрагмента по сайту *Hind* III предварительно обработанному ферментом Кленова.

Полученную конструкцию трансфицировали в клетки HeLa, и проводили визуализацию с помощью флуоресцентной микроскопии (рисунок 11б). Согласно полученным данным, оба гена экспрессировались совместно. Согласованное регулирование подтверждается данными GeneCards (<http://www.genecards.org>). Совместное регулирование предполагает, что промотор между генами является истинным двунаправленным промотором

### 4.2.3. Функциональный анализ консервативных областей промотора генов *PSENNEN* и *U2AF1L4*

Ген *U2AF1L4* кодирует фактор сплайсинга U2AF26, который представляет собой вариант малой субъединицы вспомогательного фактора U2 (U2AF). Аминокислотная последовательность U2AF26 на 89% идентична последовательности наиболее изученного фактора U2AF35, кодируемого геном *U2AF1* [236]. Геном человека содержит также ген *ZRSR* являющийся паралогом гена *U2AF1*, кодируемый им белок также имеет высокую степень соответствия с U2AF35 [255]. Гены, кодирующие большие субъединицы U2AF представлены четырьмя паралогами: *U2AF2*, *PUF60*, *RBM39* и *RBM23*, кодирующими фактор сплайсинга U2AF65 и три U2AF65-подобных белка, соответственно [255]. Мы сравнили нуклеотидные последовательности промотора *U2AF1L4 / PSENNEN* и семи паралогов предполагаемых промоторов малых и больших субъединиц гена U2AF. Пять из восьми промоторов содержат последовательности, очень близкие к тандемному повтору (TGCGCG)<sub>2</sub> с вариабельной второй частью (рисунок 12). Промотор гена *U2AF1* содержит три такие последовательности.



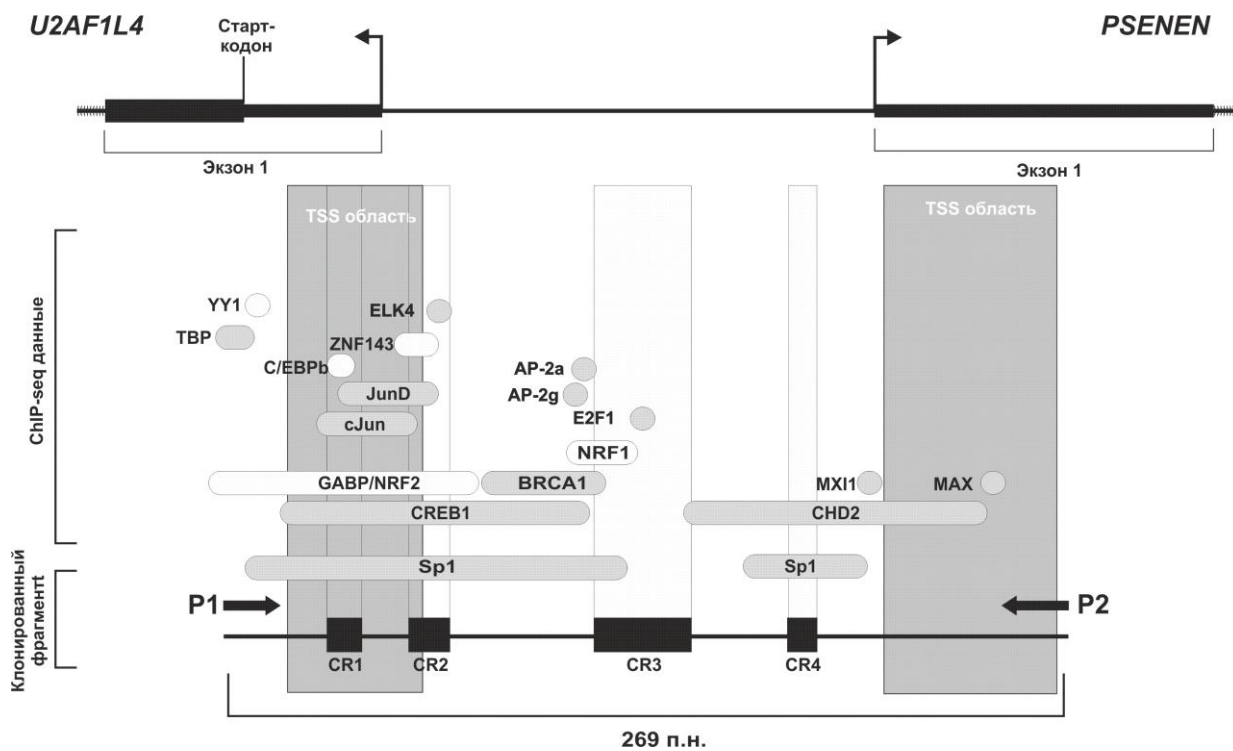
**Рисунок 12.** Связывание транскрипционного фактора NRF1 в промоторных областях паралогов гена *U2AF1L4*. Показаны последовательности, узнаваемые NRF1, внутри промоторных областей генов большой и малой субъединицы U2AF. Консенсусная последовательность была создана с использованием WebLogo [256].

Ранее [257] было показано, что консенсус (TGCGCG)<sub>2</sub>, расположенный в промоторе гена CD47 может связываться с фактором NRF1 (Nuclear Respiratory Factor 1), который активирует экспрессию некоторых ключевых метаболических генов, регулирующих клеточный рост, гены необходимые для клеточного дыхания, биосинтеза гема, а также транскрипции и репликации митохондриальной ДНК. Для NRF1 также показана связь с регулированием роста аксонов [257]. Сайты связывания NRF1 и GABP, характерны для двунаправленных промоторов [258, 259]. Еще одна особенность, которую *PSENNEN / U2AF1L4* промотор делит с большинством двунаправленных промоторов – это довольно длинная область, обогащенная динуклеотидами CpG. В полногеномных исследованиях состава последовательностей, функциональной взаимосвязи и регуляторных мотивов двунаправленных промоторов, CpG островки были обнаружены в 98% промоторов и в 61% последовательностей однонаправленных промоторов [260].

Согласно данным ChIP-seq анализа для пяти клеточных линий, в том числе HeLa [261], все промоторы генов, кодирующих субъединицы U2AF связывают NRF1 *in vivo*, по меньшей мере, в одном типе клеток.

В последовательности промотора *PSENNEN / U2AF1L4*, мы нашли четыре консервативные области (CR1-4, рисунок 13). На этом же рисунке приведены ChIP-seq данные для транскрипционных факторов, связывающихся с промоторной областью [262]. Как видно из рисунка, с промоторной областью генов *PSENNEN / U2AF1L4* связывается ряд факторов транскрипции, в том числе характерных и для двунаправленных промоторов (показаны белыми овалами). В частности, CR1 содержит сайт связывания фактора транскрипции CREB1, участвующего в регуляции транскрипции генов *PSENNEN*. Мутации сайта связывания CREB значительно уменьшают активность промоторной области длиной 238 п.н. в направлении гена *PSENNEN* [252]. Влияние этого сайта на транскрипцию в направлении *U2AF1L4* не было изучено.





**Рисунок 13.** Схематическое отображение сайтов связывания транскрипционных факторов в промоторной области длиной 269 п.н. генов *PSENE1/U2AF1L4*. Белые овалы отображают положение транскрипционных факторов характерных для двунаправленных промоторов. Черные овалы отображают положение других транскрипционных факторов, связывающихся с последовательностью. Светло-серые области показывают участки, содержащие точки инициации транскрипции. Положение консервативных областей (CR1-4) показано черными прямоугольниками. Изображение выполнено на основе ChIP-seq данных геномного браузера [261]

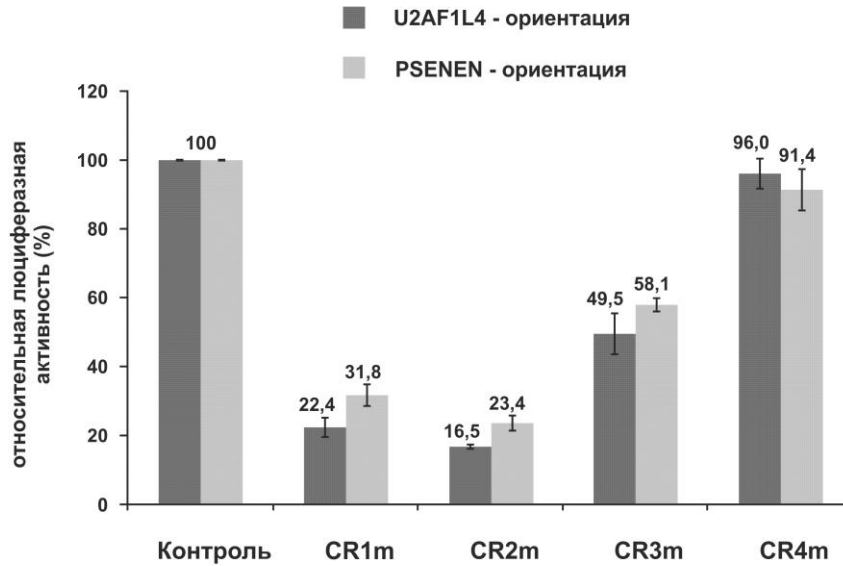
Другая консервативная область (CR2) содержит последовательность ААСТАСАСТТССС аналогичную известным консенсусам АСТАКАННТССС [258], RРАСТАКАН [263] и АСТААУРНННССС [264], которые представляют собой последовательности, широко представленные в двунаправленных промоторах человека. Следует отметить, что мотив RРАСТАКАН является сверхпредставленной последовательностью, которая связывает фактор hStaf / ZNF143 в клетках человека [263]. Данные ChIP-seq проекта ENCODE для ZNF143 и данные, представленные в работе Анно [263] показывают, что промотор *PSENE1/U2AF1L4* связывает ZNF143 в линиях клеток человека, а пики сигналов, полученных в ChIP-seq экспериментах совпадают с CR2 областью (рисунок 13). На этом же рисунке показано, что CR3 содержит предполагаемый сайт связывания NRF1. Пики сигналов для

NRF1, полученные в ChIP-seq экспериментах в рамках проекта ENCODE для нескольких линий клеток были нанесены на карту в CR3 регионе.

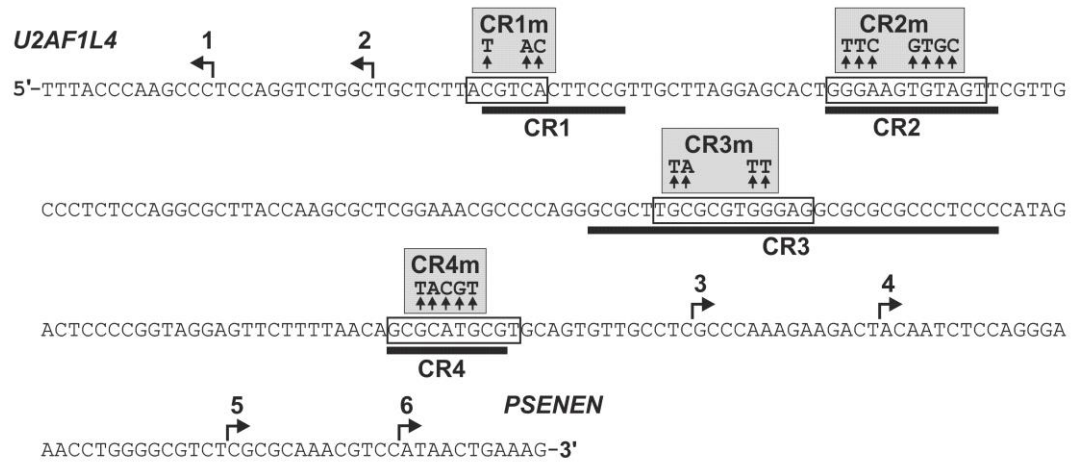
Для того чтобы выяснить функциональную роль консервативных последовательностей в активности промотора *PSENEEN / U2AF1L4*, мы ввели мутации в эти участки. Для создания нуклеотидных замен, были использованы пары праймеров: CR1mut, CR2mut, CR3mut, CR4mut. Мутантные промоторы были клонированы в вектор pGL4.10, как описано выше для фрагмента дикого типа. Для всех мутантов были подготовлены плазмиды, содержащие вставку в ориентации гена *PSENEEN* или *U2AF1L4*. Все мутации были проверены с помощью секвенирования.

Данные мутации были внесены в предполагаемые сайты связывания факторов транскрипции и сайт связывания транскрипционного фактора CREB, расположенного в области CR1 (рисунок. 13). Промоторы дикого типа и мутантные варианты клонировали в вектор pGL4.10 в обеих ориентациях на место промотора, выше гена люциферазы, затем полученные конструкции трансфицировали в клетки HeLa. Плазмиды, содержащие промотор дикого типа *PSENEEN / U2AF1L4* в *PSENEEN* ориентации (PGL4-P) и *U2AF1L4* ориентации (PGL4-U), использовали в качестве контроля (рисунок 14а). Как видно из рисунка, мутации в CR1 или CR2 приводят к снижению активности промотора *PSENEEN / U2AF1L4* в обеих ориентациях на 70%, а в CR3 на 40-50%. Мутация в области CR4 не оказывает существенного влияния на активность промотора *PSENEEN / U2AF1L4*. Полученные результаты подтверждают гипотезу о том, что транскрипция генов *PSENEEN* и *U2AF1L4* управляется двунаправленным промотором и общими транскрипционными регуляторными элементами для обоих направлений.

**а**



**б**



**Рисунок 14.** а-активность мутантных форм промотора *PSENE1/U2AF1L4* в клетках HeLa. Активность промоторной области дикого типа (контроль) принята за 100%. б-промоторная область генов *PSENE1/U2AF1L4*, в которой стрелками показаны введенные в последовательность мутации внутри консервативных областей (CR1-4)

### **4.3. ОЦЕНКА ЭНХАНСЕР-БЛОКИРУЮЩЕЙ АКТИВНОСТИ ИНСУЛЯТОРОВ КАРТИРОВАННЫХ В ЛОКУСЕ *FXVD5-COX7A1* ХРОМОСОМЫ 19 ЧЕЛОВЕКА**

Инсуляторы представляют собой последовательности ДНК, нарушающие активацию промоторов не предназначенными для данного промотора энхансерами и/или способные препятствовать распространению конденсированного состояния хроматина на его активные области [265].

С количественной стороны активность инсуляторов практически не исследована. В большинстве случаев активность инсулятора оценивают по его способности блокировать взаимодействие промотора с энхансером (энхансер блокирующие инсуляторы) в составе линейаризованной конструкции, экспрессирующей репортерный ген люциферазы. В рамках проекта по созданию функциональной карты валидированных регуляторных элементов в локусе *FXVD5-COX7A1* хромосомы 19 человека длиной 1 млн. п.н. нами была выполнена работа по поиску новых, а также проверке ранее идентифицированных *цис*-регуляторных элементов.

На первом этапе работы необходимо было получить набор векторов, предназначенных для анализа энхансер-блокирующей активности фрагментов ДНК, в том числе рабочий вектор pGL4EPV, содержащий в регуляторной области гена люциферазы светлячка энхансер и промотор вируса SV40.

#### **4.3.1. Конструирование плазмид**

Для получения плазмиды pGL4PV, содержащей в регуляторной области гена люциферазы промотор SV40, в плазмиду pGL4.10 (Promega), гидролизованную по сайтам *Xho* I и *Hind* III, лигировали фрагмент длиной 215 п.н., полученный при гидролизе плазмиды pGL3-Promoter Vector (Promega) рестриктазами *Xho* I и *Hind* III.

Плазмиду pGL4EPV, содержащую минимальный промотор и энхансер SV40, получали после гидролиза плазмиды pGL4PV *Xba* I и *Bam* HI и лигирования ее с содержащим энхансер SV40 фрагментом, полученным

обработкой pGL3-Control Vector (Promega) рестриктазами *Xba* I и *Bam* HI, и использовали для клонирования потенциальных инсуляторов по уникальному сайту *Sal* I, расположенному вне пары промотор-энхансер.

Плазмиду pGL4EPV2 получали из линейаризованной по сайту *Sal* I плазмиды pGL4PV с затупленными при помощи фрагмента Кленова концами путем лигирования с фрагментом, содержащим энхансер SV40. Энхансер длиной 353 п.н. получали после обработки плазмиды pGL3CV *Hpa* I и *Bam* HI, обработки рестриктной смеси фрагментом Кленова, и выделения фрагмента нужной длины в 1,5% агарозном геле. Эту плазмиду использовали для клонирования инсуляторов по уникальному сайту *Bam* HI, расположенному между промотором и энхансером

#### 4.3.2. Выбор фрагментов ДНК

Были отобраны четыре фрагмента ДНК, для которых имелось экспериментальное подтверждение их энхансер-блокирующей активности методом позитивно-негативной селекции [11, 12]. Свойства этих фрагментов суммированы в таблице 8.

Кроме этого, основываясь на данных иммунопреципитации хроматина с последующим массивированным секвенированием (ChIP-seq) проекта ENCODE, содержащихся в базах данных геномного браузера (UCSC Human Genome Browser, см. [261]), были выбраны шесть фрагментов ДНК, обладающих свойствами потенциальных энхансеров и пять фрагментов со свойствами потенциальных инсуляторов. Отбор основывался на способности фрагментов связывать факторы транскрипции Р300 и CTCF, их сверхчувствительности к расщеплению ДНК-азой I в составе хроматина, а также на состоянии их хроматина в соответствии с критериями работ [266, 267]. Расположение выбранных фрагментов в геноме и их свойства приведены в таблице 9. Метрическая карта локуса с выбранными фрагментами приведена на рисунке 15.

**Таблица 8.** Свойства потенциальных инсуляторов, полученных в результате негативно-позитивной селекции.

Фрагмент	Положение на хромосоме 19 человека (сборка hg19)	длина, п.н.	Состояние хроматина*, **	Связывание CTCF*	Связывание Р300*
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 в соответствующих ячейках обозначают линии клеток, в которых данный тип активности был обнаружен: H - HeLa, G - GM12878, P - HepG2, K - K562, I - IMR90, A - A549. Уровень связывания факторов транскрипции приведен в соответствии с данными ChIp-seq ENCODE [268] – от (-) – отсутствие связывания до (+++) – сильное связывание.

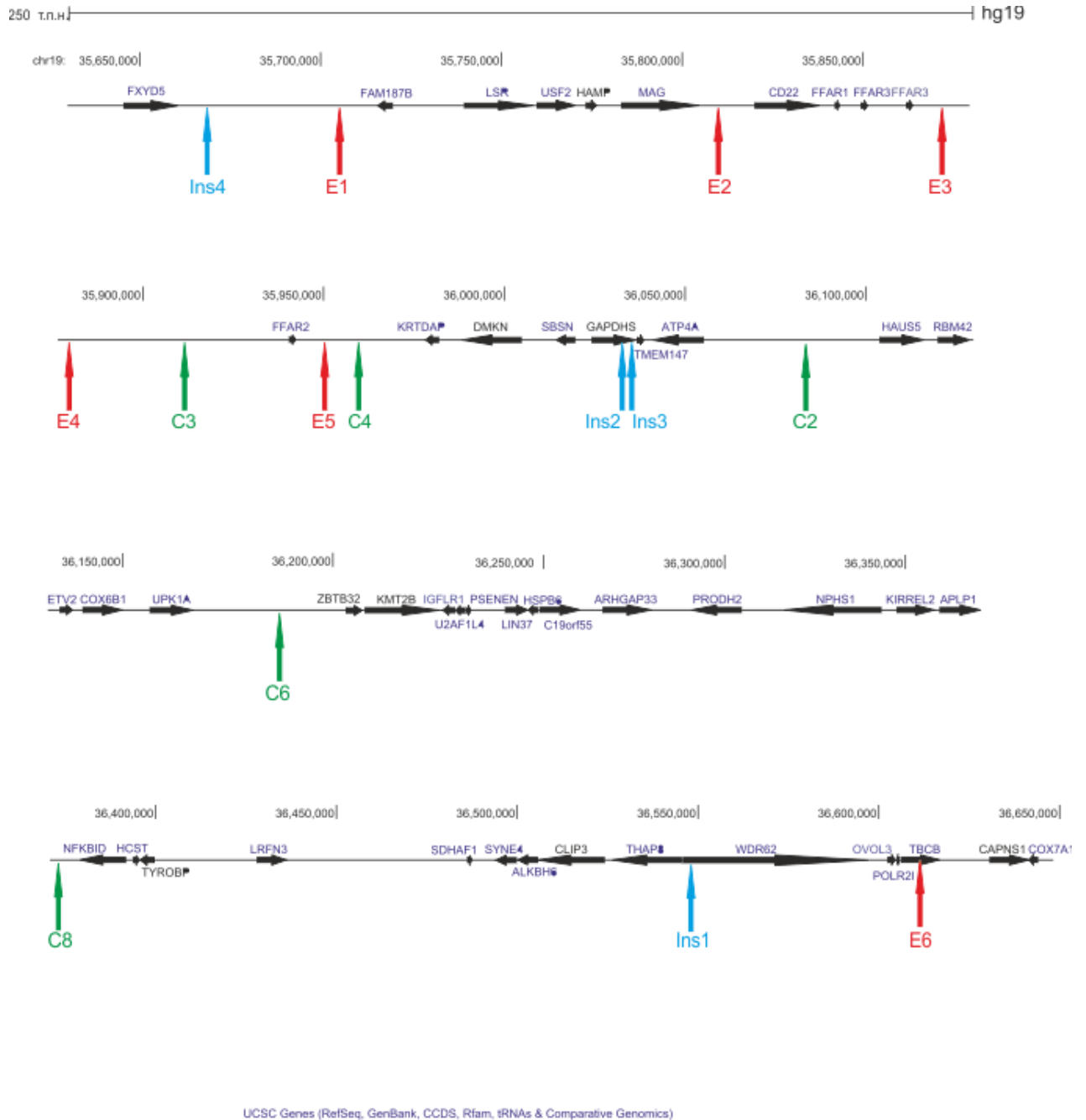
\*\* Состояние хроматина приведено в соответствии с картой, предложенной в [266, 267].

**Таблица 9.** Свойства геномных фрагментов, отобранных по функциональным характеристикам.

Фрагмент	Положение на хромосоме 19 человека (сборка hg19)	Праймеры	Длина, п.н.	Состояние хроматина*, **	Связывание CTCF*	Связывание Р300*	Сверхчувствительность к ДНК-азе I*
E1	35704659-35705401	GTGGACCGTGTAGCCGAGA GACAGCCCAGCAAGCTCA	743	Сильный энхансер <b>GK</b>	-	+++ <b>GH</b>	+++ <b>AGH</b>
E2	35809671-35810383	GACAGTGAGCACAGGCTATCA CTCTACCCCCGGCCAGTC	713	Сильный энхансер <b>KP</b>	-	-	+++ <b>АНK</b>
E3	35871105-35872127	TGCTTACAGCTTTCCAGCTAA AAGGAGAGCCAGCATCAGTC	1023	Сильный энхансер <b>KP</b>	-	+++ <b>K</b>	+++ <b>K</b>
E4	35879406-35880078	GCTGGCTGTCCTCATTATCC GAAGGATTTTCTTGCCCACT	673	Сильный энхансер <b>KP</b>	-	+++ <b>K</b>	+ <b>HKP</b>
E5	35949575-35950799	ACACTCCCACCAGGAATCA GCCGTGGAAACTCAAATGG	1225	Сильный энхансер <b>GP</b>	-	+++ <b>G</b>	++ <b>AGP</b>
E6	36610820-36611414	CTCCAGCAGTCCCTTGTC CATCCAAGCTCCCAGAGG	595	Сильный энхансер/промотор <b>P</b>	-	+++ <b>P</b>	++ <b>P</b>
C2	36083109-36083439	CACATATGATGGTGAAAAGCAG TTTGTTAAGAGAAAAGGTAAGTGA	331	Инсулятор <b>GKP</b>	+++ <b>IKG</b>	-	++ <b>AGH</b>
C3	35911268-35911580	CACGTCTTCCATCAAGGACT CACACTCTGCGGTTCCACAC	313	Сильный энхансер <b>G</b>	+++ <b>GK</b>	-	++ <b>AGHP</b>
C4	35959620-35960321	TGAACTCTGCCSTAAGACC TGTTTCAGACSTCTTCCGTA	702	Инсулятор <b>KP</b>	+++ <b>IKG</b>	-	++ <b>AGHK</b>
C6	36184653-36185486	GATTCGGGATAAATGACACA CTAGACTCAAGAAGGCACT	834	Инсулятор <b>GKP</b>	+++ <b>GK</b>	++ <b>K</b>	+++ <b>AGHK</b>
C8	36372733-36373330	TТААТАCTCCAAGGTTGTCC ACACGACSTATATTACAGT	598	Инсулятор <b>GKP</b>	+++ <b>GKI</b>	-	+++ <b>АН</b>

\*Буквы в соответствующих ячейках обозначают линии клеток, в которых данный тип активности был обнаружен: Н - HeLa, G - GM12878, P - HepG2, K - K562, I - IMR90, A - A549. Уровень связывания факторов транскрипции приведен в соответствии с данными ChIp-seq ENCODE [268] – от (-) – отсутствие связывания до (+++) – сильное связывание.

\*\*Состояние хроматина приведено в соответствии с картой, предложенной в [266, 267].

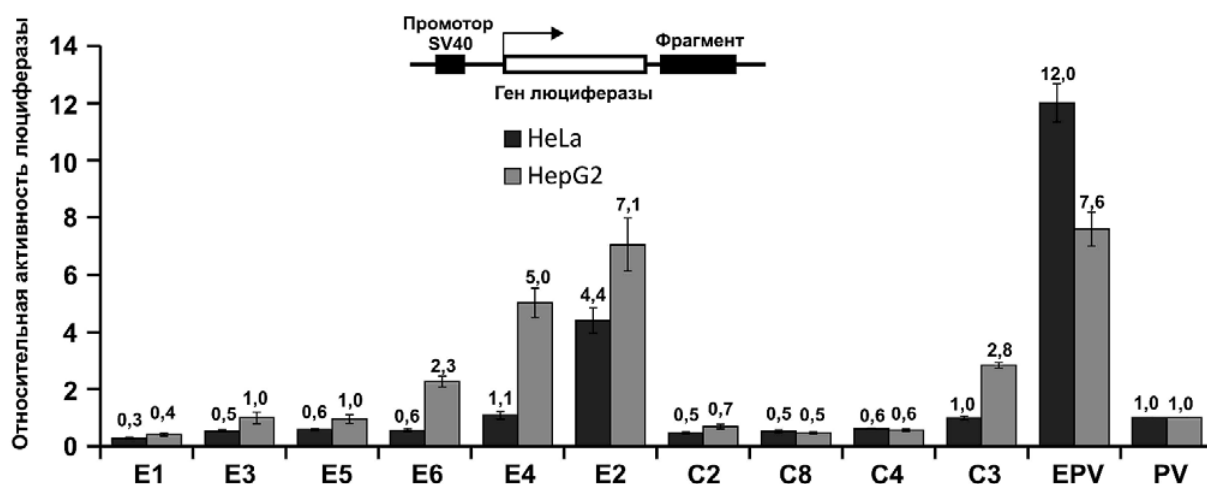


**Рисунок 15.** Метрическая карта области *FXYD5-COX7A1* хромосомы 19 человека длиной  $10^6$  п.н. Красными вертикальными линиями показано положение изученных в работе потенциальных энхансерных элементов. Зеленые линии отображают положение CTCF связывающих элементов, отобранных для изучения. Голубыми вертикальными линиями показано положение потенциальных инсуляторных последовательностей, идентифицированных ранее (№ 1-4) [11, 12]. Карта составлена с помощью UCSC Human Genome Browser (сборка hg19.).



### 4.3.3. Энхансерная активность геномных фрагментов

Для проверки энхансерной и сайленсерной активностей, выбранные фрагменты ДНК были амплифицированы с геномной ДНК-матрицы (последовательность праймеров см. в таблице 9) и клонированы в плазмиду pGL4PV на место энхансера по сайту *Sal* I с 3'-стороны от репортерного гена люциферазы (рисунок 16). Полученные конструкции трансфицировали в клетки HeLa и HepG2. В приведенном эксперименте использовались кольцевые плазмиды, во всех остальных случаях – линейаризованные.



**Рисунок 16.** Энхансерная активность фрагментов ДНК из генома человека в клетках HeLa и HepG2. Фрагменты были отобраны в качестве потенциальных энхансеров (E1–E6) или потенциальных инсуляторов (C2–C8), их свойства суммированы в таблице 9. Критерии отбора описаны в 4.3.1. Фрагменты ДНК были клонированы в плазмиду pGL4PV как показано в верхней части рисунка. Для оценки энхансерной активности использовали кольцевые плазмиды. Контрольные плазмиды: PV – pGL4PV (содержит SV40 промотор и не содержит энхансера); EPV – pGL4EPV (содержит промотор и энхансер SV40). Активность люциферазы после трансфекции клеток плазмидой pGL4PV принята за единицу

Из шести последовательностей со свойствами потенциальных энхансеров (E1–E6, см. рисунок 16 и таблица 9) способность усиливать активность промотора SV40 в данной системе в клетках HepG2 проявили три, причем только один из них (E2) обладал активностью и в клетках HeLa. Из четырех последовательностей, отобранных по связыванию фактора транскрипции CTCF (потенциальных инсуляторов) один (C3) проявил относительно невысокую энхансерную

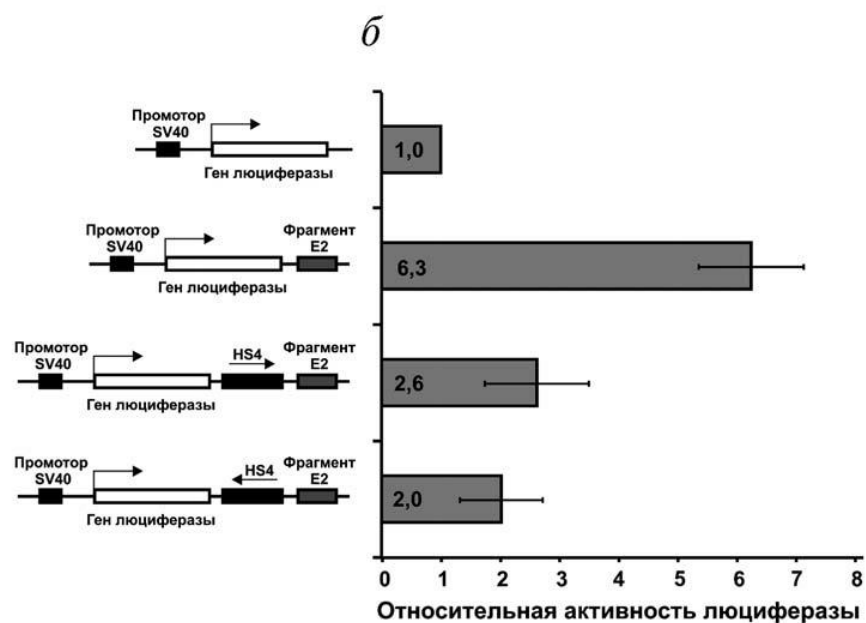
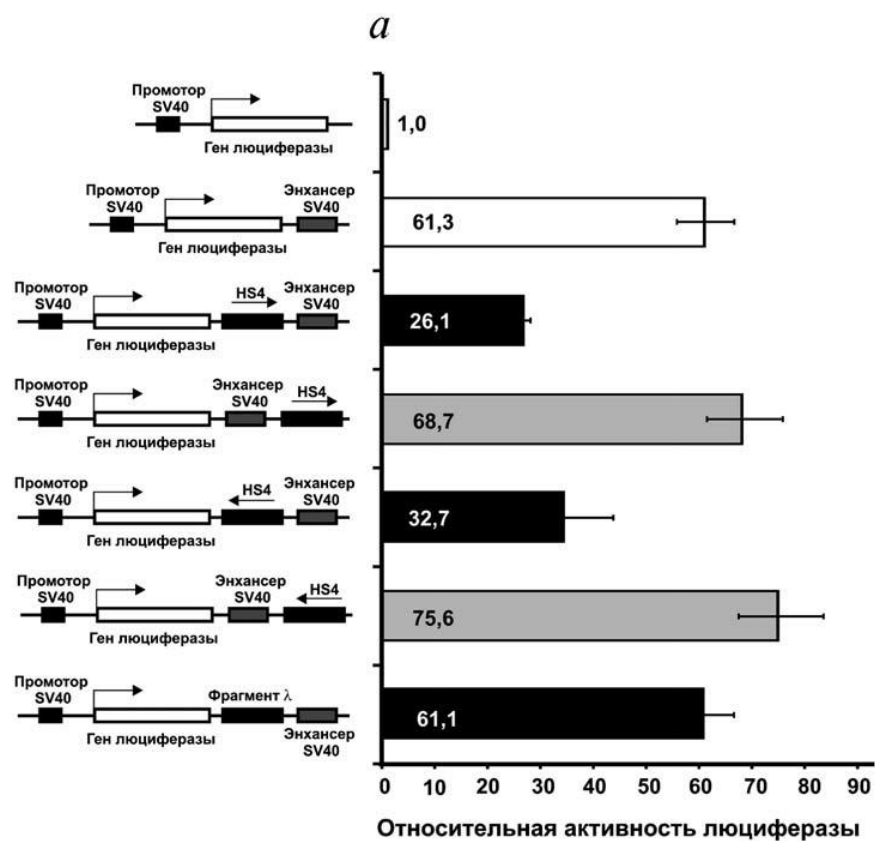
активность в клетках HepG2, остальные фрагменты, как и ожидалось, не проявили энхансерной активности. Примечательно, что фрагмент С3, кроме связывания CTCF, обладал также свойствами потенциального энхансера (см. таблицу 9).

Таким образом, из семи потенциальных энхансеров в использованной системе ожидаемую активность проявили четыре (~60%), что свидетельствует о существенной, хотя и не полной корреляции между их свойствами, использованными для отбора, и их энхансерной активностью в использованной системе.

#### **4.3.4. Энхансер-блокирующая активность инсулятора cHS4 из бета-глобинового локуса кур**

Для того чтобы подтвердить пригодность системы для анализа энхансер-блокирующей активности инсуляторов, мы проверили активность хорошо охарактеризованного инсулятора cHS4 из  $\beta$ -глобинового локуса кур. Данный инсулятор обладает доказанной энхансер-блокирующей активностью в ряде систем, включая систему с использованием транзientной трансфекции конструкций, содержащих промотор и энхансер SV40 [269].

Мы клонировали полноразмерный (1,2 т.п.н.) инсулятор cHS4 в обеих ориентациях между энхансером и промотором SV40, а также между промотором SV40 и энхансером E2. Параллельно были приготовлены контрольные конструкции, в которых cHS4 был клонирован вне пары промотор-энхансер и конструкции, не содержащие инсулятора. Конструкции были трансфицированы в клетки HeLa и определена активность люциферазы для этих конструкций (рисунок 17).

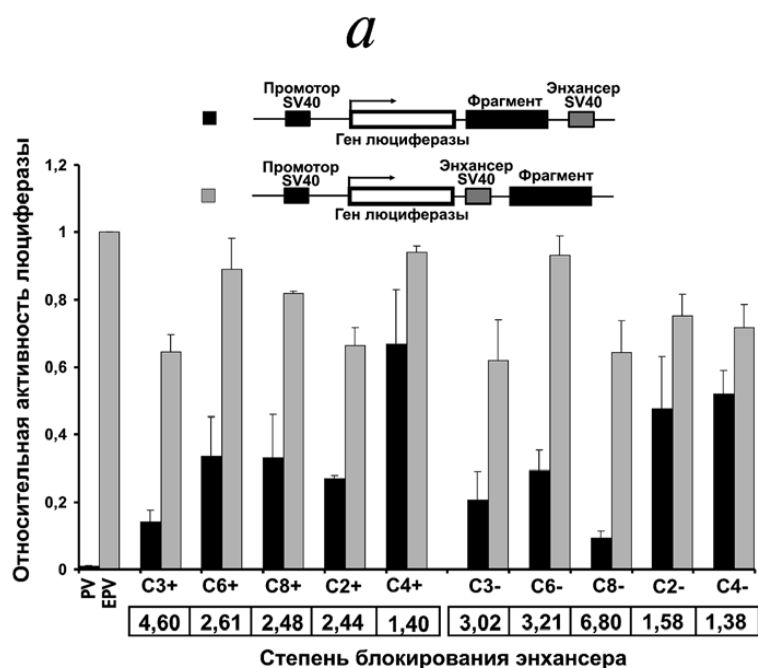


**Рисунок 17.** Эхансер блокирующая активность инсулятора HS4 из бета глобинового локуса кур в клетках HeLa по отношению пары промотор SV40 – эхансер SV40 (а) и пары промотор SV40 – эхансер E2 (б). За единицу принята активность люциферазы после трансфекции клеток плазмидами, содержащими промотор SV40 и соответствующий эхансер.

Согласно полученным данным (рисунок 17а), инсулятор cHS4 способен понижать активность пары промотор-энхансер SV40 в 2-3 раза, причем, только в том случае, когда он расположен между энхансером и промотором. Эффект подавления активности отсутствует при клонировании между энхансером и промотором фрагмента ДНК фага лямбда. Аналогичный результат получен для пары промотор SV40 – энхансер E2 (рисунок 17б). В обоих случаях энхансер-блокирующая активность слабо зависела от ориентации инсулятора относительно промотора. Таким образом было показано, что полученный набор конструкций пригоден для проверки и анализа энхансер-блокирующих инсуляторных элементов.

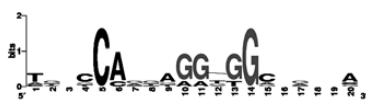
#### **4.3.5. Энхансер-блокирующая активность геномных фрагментов**

Для проверки способности фрагментов ДНК, отобранных по свойствам инсуляторов, блокировать активацию промотора близлежащим энхансером, данные фрагменты были клонированы между промотором и энхансером в вектор pGL4EPV2 по сайту *BamH I*. Контрольные конструкции, содержащие те же фрагменты на месте энхансера либо вне пары энхансер-промотор, позволяли оценить энхансерную или сайленсерную активности фрагментов, наличие которых могло бы помешать определению энхансер-блокирующей активности (см. схему на рисунке 18). Для того чтобы избежать взаимодействий между энхансером и промотором, полученные конструкции линеаризовали и трансфицировали в клетки млекопитающих, после чего измеряли активность репортерного гена люциферазы с использованием системы двойной люциферазной детекции.



**б**

	Сайт связывания CTCF	Блокирование энхансера, раз. (+/- ориентация)
C2	TAAG <u>CA</u> CTAGGTGGACAACA	2,44/1,58
C3	AGCC <u>CA</u> GGTGGGGCAGTGC	4,60/3,02
C4	TGCC <u>CA</u> GGAGGTGGCAGGCA	1,40/1,38
C6a	TTGT <u>CA</u> ATAGAGGGCGCTGG	2,61/3,21
C66	TGAC <u>CA</u> GCAGGTGGCTGCTA	2,61/3,21
C8	CTTCCAGGAGATGGCAGTA	2,48/6,80



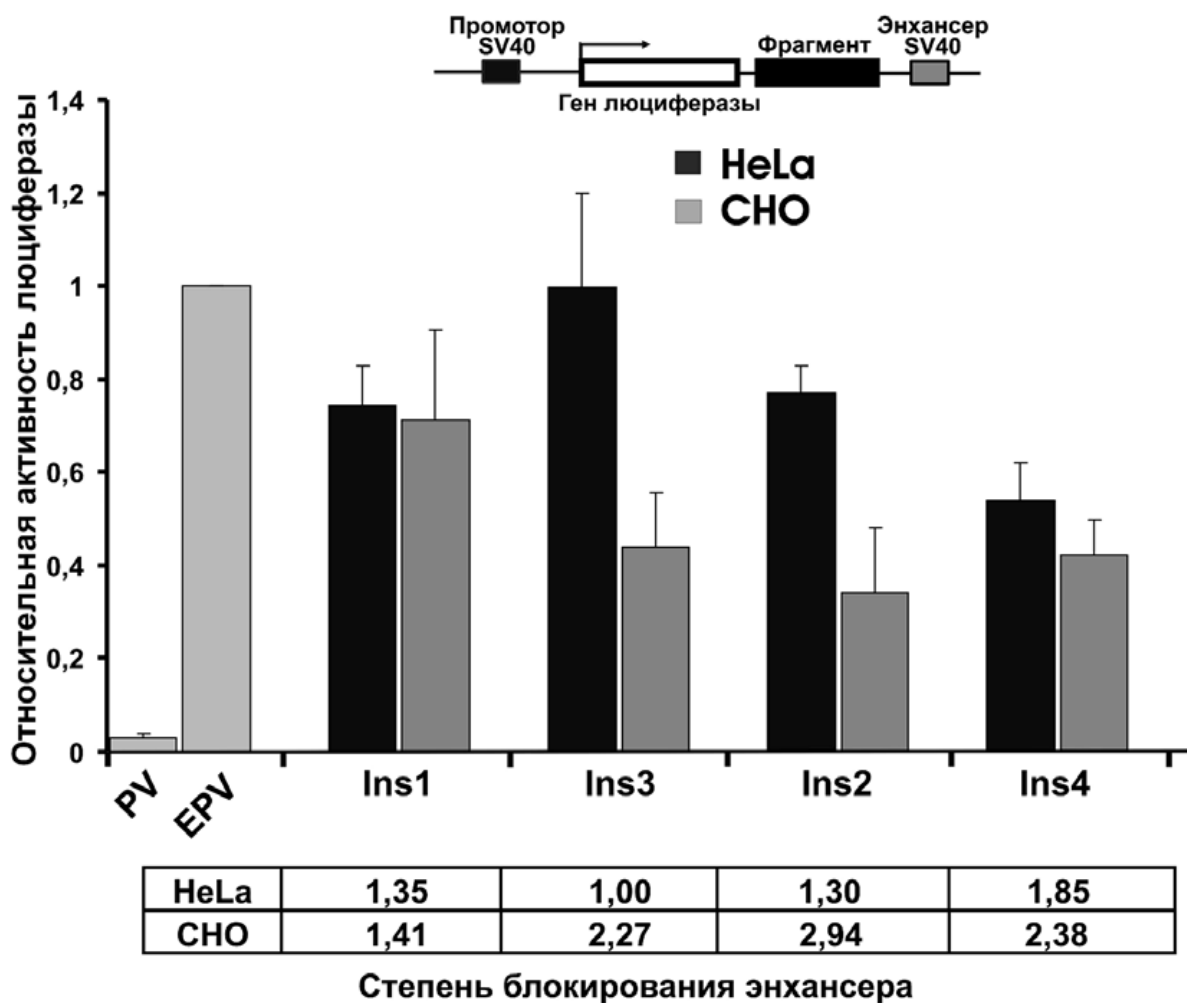
**Рисунок 18.** Энхансер блокирующая активность фрагментов ДНК из генома человека в клетках HeLa. а – Фрагменты ДНК были клонированы в плазмиду рGL4EPV2 между энхансером и промотором (темные столбцы), либо в плазмиду рGL4EPV2 вне пары энхансер промотор (светлые столбцы) в обеих ориентациях. «+» – ориентация фрагмента совпадает с ориентацией в геноме, «-» – ориентация противоположна ориентации в геноме. Для трансфекции использовали линейаризованные плазмиды PV и EPV (см. подпись к рисунку 16). Активность люциферазы после трансфекции клеток плазмидой рGL4EPV, содержащей энхансер и промотор SV40, принята за единицу; б – выравнивание участков связывания белка CTCF из фрагментов геномной ДНК человека. Во втором столбце приведена степень блокирования энхансера соответствующих фрагментов в обеих ориентациях относительно промотора. Подчеркнуты нуклеотиды, совпадающие с консенсусной последовательностью (приведена в виде WebLogo [256]).

Из пяти геномных фрагментов, отобранных по свойствам, соответствующим инсуляторам (таблица 9), четыре проявили явную способность ослаблять действие энхансера SV40 на промотор SV40 в клетках HeLa только в том случае, когда фрагменты помещены между этими элементами (рисунок 18). Возможным исключением является фрагмент С4, инсуляторные свойства которого почти не регистрируются, несмотря на сильное связывание CTCF и состояние хроматина, соответствующее инсулятору (таблица 9). Степень ослабления действия энхансера на промотор с учетом поправки на сайленсерную активность колеблется от ~2 (фрагмент С2) до более чем 7 (фрагмент С8) раз. При этом среди фрагментов присутствуют как те, активность которых сильно зависит от ориентации относительно промотора (С8), так и те, активность которых практически не зависит от ориентации (С3, С6).

Интересно, что фрагмент С3, не относящийся к инсуляторам по структурным особенностям хроматина [266, 267], но способный связывать CTCF с высокой эффективностью (см. таблицу 9), проявил высокую энхансер-блокирующую активность, тогда как фрагмент С4 с характеристиками инсулятора и также высоким связыванием CTCF, такой активности не проявил.

#### **4.3.6. Энхансер-блокирующая активность фрагментов, отобранных при помощи негативно-позитивной селекции**

Четыре фрагмента отобранных с использованием системы позитивно-негативной селекции (таблица 8) были клонированы в плазмиду pGL4EPV2 между энхансером и промотором SV40. Поскольку отбор потенциальных энхансер-блокирующих фрагментов исходно проводился в клетках СНО, полученные конструкции были линейаризованы и трансфицированы в клетки СНО, и, для сравнения, в клетки HeLa. Результаты трансфекций приведены на рисунке 19.



**Рисунок 19.** Энхансер блокирующая активность по отношению пары промотор SV40 – энхансер SV40 потенциальных инсуляторов, отобранных из генома 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.

Из рисунка 19 видно, 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. С другой стороны, энхансер-блокирующая активность двух элементов (Ins2 и Ins3) наблюдалась только в клетках CHO и практически отсутствовала в клетках HeLa, а активность Ins4 была одинаковой в этих клетках. Интересно также, что энхансер-блокирующая активность элементов Ins2 и Ins3, связывающих фактор транскрипции CTCF (таблица 8), зависит от типа клеток, тогда как активность Ins4,

для которой связывания СТСФ не наблюдалось, не зависит. Для подтверждения такой корреляции, однако, требуются дальнейшие исследования.

#### **4.4. АКТИВАЦИЯ ПРОМОТОРА ЭНХАНСЕРОМ В *ТРАНС* ПОЛОЖЕНИИ**

В конце 1980-х годов было показано, что при транзientной совместной трансфекции энхансер может активировать промотор, расположенный на другой плазмиде, когда эти элементы удерживаются вместе с помощью нековалентных связей [270, 271]. В дальнейшем было показано, что эффект активации в *транс*-положении может наблюдаться при совместной трансфекции плазмиды, содержащей энхансер и плазмиды, содержащей промотор и репортерный ген, то есть при проникновении в клетку нескольких копий плазмид, содержащих энхансер и промотор. В таком эксперименте сильный энхансер цитомегаловируса активировал промотор цитомегаловируса в *транс*-положении (то есть, энхансер расположен на отдельной плазмиде котрансфецируемой с генной конструкцией, содержащей промотор и репортерный ген) [155].

В данной работе перед нами была поставлена задача дальнейшего изучения промоторной и клеточной специфичности активации промотора энхансером в *транс*-положении.

##### **4.4.1. Конструирование плазмид**

pGL4-U плазида была получена путем вставки двунаправленного промотора U2P в вектор pGL4.10 (Promega) выше гена люциферазы, как описано в главе 4.2.2.

Для получения Pst<sup>+</sup> плазмиды, содержащей ген люциферазы светлячка (*Luc2*) под контролем минимального CMV промотора длиной 61 п.н., фрагмент длиной 318 п.н., содержащий минимальный промотор CMV и промотор гена теломеразы человека hTERT вырезали из pGL3-hTERT-CMV плазмиды [272] с помощью *Xho* I и *Bgl* II и клонировали в pGL4.10 выше гена *Luc2*. Промотор hTERT затем был вырезан из конструкции с помощью *Xho* I и *Eco* RI, образовавшиеся липкие концы в плазмиде были достроены фрагментом Кленова и лигированы.



Промотор гена тимидинкиназы вируса простого герпеса Ptk, был проверен в составе плазмиды pRL-ТК (Promega), содержащей ген люциферазы *Renilla reniformis* под контролем промотора гена HSV-*tk*. Данную плазмиду как правило используют в качестве контроля для нормирования различий эффективности трансфекции между параллельными образцами в системе двойной люциферазной детекции Promega. Однако, как было показано в работах [273, 274], некоторые последовательности данной плазмиды могут влиять на регуляторные элементы референсной плазмиды, что может привести к экспериментальным артефактам. Поскольку предлагаемая нами система была предназначена, для того чтобы показать действие энхансера на промотор в *транс*-положении, мы отказались от использования подхода двойной люциферазы. В то же время, мы использовали pRL-ТК плазмиду, для того чтобы показать эффект энхансеров на промотор гена HSV-*tk*.

Проверенные ранее геномные энхансеры E2, E4 и E6 (см. раздел 4.3.3), были клонированы в вектор pAL-ТА (Evrogen), в результате чего были получены плазмиды E2, E4 и E6, соответственно. Для получения плазмиды Esv40, энхансер SV40, амплифицировали с помощью праймеров SV40left и SV40right (последовательности указаны в таблице 6 «материалы и методы») на матрице плазмиды pGL4.13 (Promega) и клонировали в вектор pAL-ТА (Evrogen). Фрагмент ДНК, содержащий минимальный промотор CMV, был вырезан из pPNT/EmP плазмиды [11] с помощью *Xho* I и *Hind* III, после чего 5'-концы были достроены с использованием Т4 ДНК полимеразы и добавлены поли-Т-концы с использованием терминальной трансферазы. Полученный фрагмент был очищен в геле и лигирован в вектор pAL-ТА, в результате чего была получена Ecmv плазида.

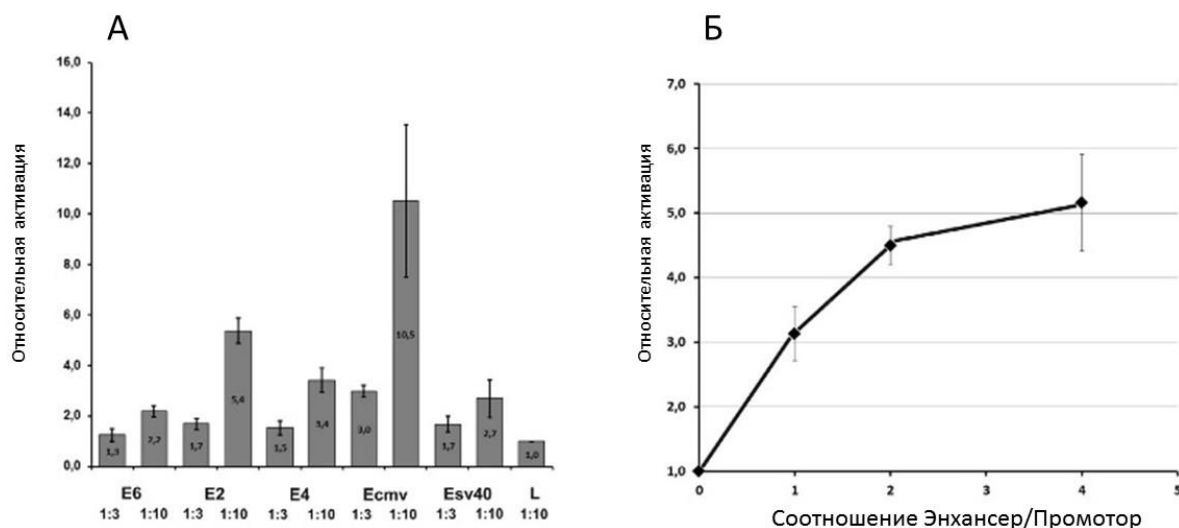
Контрольную плазмиду (L), получали с помощью ПЦР-амплификации фрагмента длиной 162 п.н. на матрице ДНК фага лямбда с использованием праймеров  $\lambda$ left и  $\lambda$ right (таблица 6) и лигированием полученного фрагмента в вектор pAL-ТА (Evrogen).

#### 4.4.2. Активация промотора энхансером в *транс*

Для того чтобы нивелировать эффект маскирования *транс*-активации, в наших опытах по трансфекции репортерного гена мы не использовали нормировку по референсной плазмиде pRL-TK, вместо этого мы предположили, что эффективность трансфекции в одном эксперименте одинакова для всех трансфицированных плазмид. Это предположение подтверждается относительно небольшой стандартной ошибкой среднего значения, полученной при измерениях относительной люциферазной активности (рисунок 20 и 21).

Результаты экспериментов по совместной трансфекции клеток HepG2 с использованием HSV-*tk* промотора (Ptk) и нескольких энхансеров различного происхождения, показаны на рисунке 20 а. Как видно из рисунка, наиболее выраженная активация (более 10 раз), наблюдается в случае совместной трансфекции конструкциями, содержащими энхансер CMV и промотор CMV при соотношении энхансер-промотор равном 10:1. Тем не менее, другие энхансеры, в том числе вирусный энхансер SV40 и три испытуемые энхансера из генома человека, также способны усиливать транскрипцию Ptk промотора в 2-5 раза в клетках HepG2. Кроме того, степень *транс*-активации зависит от соотношения энхансер-промотор.

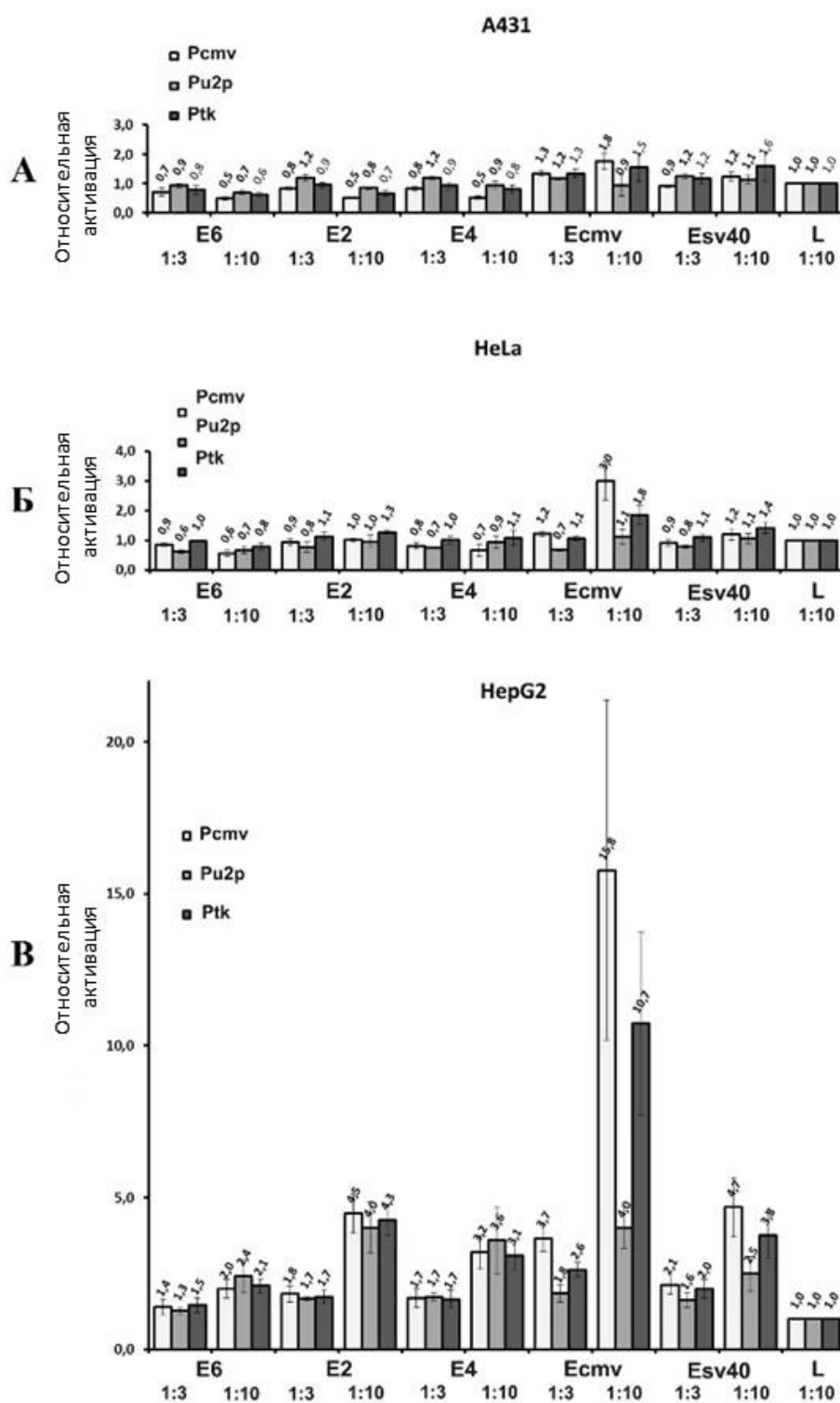
Данный эффект более детально представлен на рисунке 20 б. Как было показано, чем выше отношение энхансер-промотор, тем выше степень активации, в клетках HeLa уровень относительной активации составляет около пяти.



**Рисунок 20.** Активация промотора энхансером *in trans*. Активность энхансера была нормирована, на активность отрицательного контроля (L), содержащего фрагмент ДНК фага лямбда. (А) Активация промотора гена тимидинкиназы (Ptk) pRL-ТК плазмиды в клетках HepG2 различными котрансфицируемым энхансерами (описание в тексте). L - отрицательный контроль, содержащий фрагмент ДНК фага лямбда. 1:3, 1:10 - мольное соотношение промотор: энхансер. (Б) Влияние соотношения энхансер: промотор на активацию промотора Pcmv энхансером Ecmv при совместной трансфекции в клетки HeLa.

#### 4.4.3. Промоторная и клеточная специфичность

Эффект *trans*-активации промотора энхансером при транзientной трансфекции продемонстрировал значительную степень специфичности по отношению к промотору в клетках HepG2 (рисунок 21). Несмотря на то, что SV40 и цитомегаловирус являются филогенетически далекими объектами, наибольшую способность активировать в *trans* оба вирусных промотора (Pcmv и Ptk) проявили вирусные энхансеры Ecmv и Esv40, (рисунок 21в). В то же время, клеточные энхансеры E2 и E4, имеют значительно более низкую промоторную специфичность. Эта особенность энхансеров также была продемонстрирована в клетках HeLa (рисунок 21б).



**Рисунок. 21.** Специфичность активации энхансером промотора *in trans* в зависимости от типа клеток и промотора. L - отрицательный контроль, содержащий фрагмент ДНК фага лямбда. 1:3, 1:10 - мольное соотношение промотор: энхансер.

Согласно полученным данным активность проанализированных энхансеров в *транс*, зависит от типа клеток (рисунок 21). Так максимальный эффект усиления наблюдался в клетках гепатоцеллюлярной карциномы (HepG2), менее значительный эффект был обнаружен в клетках аденокарциномы шейки матки (HeLa) и практически не обнаруживался в клетках эпидермальной карциномы (A431). Ранее активность CMV энхансера в *транс* была показана в клетках HEK-293 и CHO-K1 [155], и, следовательно, данный эффект не ограничивается одним или несколькими типами клеток.

Специфичность эффекта *транс*-активации в различных типах клеток можно объяснить тем что в стандартных условиях липофекции, количество экзогенных энхансеров и промоторов, которые попадают в клетку [275], значительно превышает количество клеточных промоторов или энхансеров отдельно взятого типа, возможно, только за исключением промоторов и энхансеров, которые входят в состав повторяющихся элементов. Такое большое количество экзогенных регуляторных элементов может связывать транскрипционные факторы, необходимые для активации энхансера или промотора и, таким образом, нарушать клеточные регуляторные сети, что необходимо принимать во внимание при интерпретации получаемых результатов. Если предположить, что необходимые для активации факторы транскрипции, менее представлены в клетках HeLa и A431 по сравнению с клетками HepG2, как было показано в наших экспериментах.

Представленные здесь данные, показывают, что действие энхансеров в *транс* в составе генетических конструкций (и, возможно, промоторов и других регуляторных элементов) на транскрипцию генов клетки-хозяина также должно быть учтено при интерпретации экспериментальных данных, полученных с использованием транзистентных трансфекций.

#### **4.4.4. Взаимодействие промотора с энхансером**

В настоящее время активация промотора энхансером в *транс* при транзистентной трансфекции изучена слабо. Это можно отчасти объяснить тем, что

при стандартном измерении активности репортерного гена с использованием референсной плазмиды (например, в методе двойной люциферазной детекции), активация промотора в *транс*-положении будет замаскирована одновременным воздействием энхансера на тестируемый и контрольные промоторы. Поэтому наиболее детальная на сегодняшний день работа по изучению активации энхансером промотора в *транс* при совместной трансфекции была выполнена с использованием проточной цитометрии, а не анализа активности репортерного гена [155].

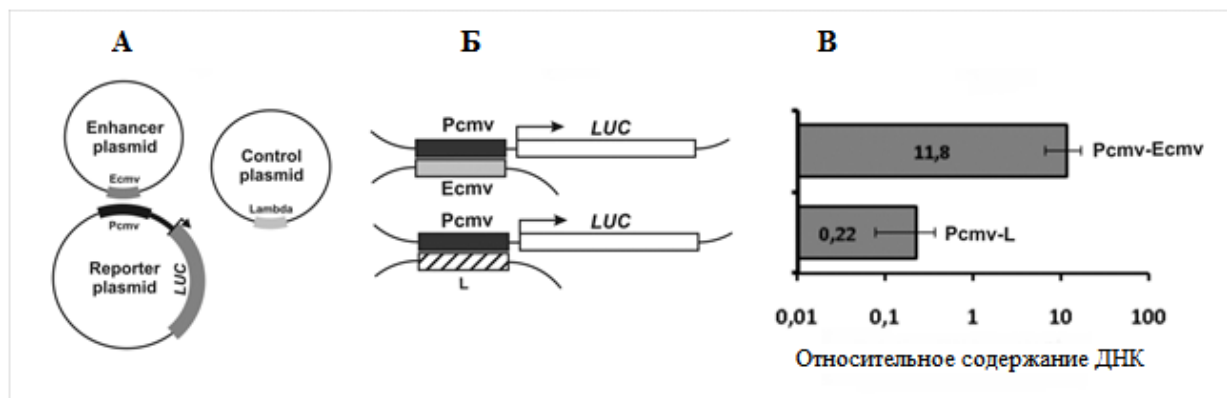
Модель дистальных (геномных) взаимодействий энхансера с промотором при котором происходит физическое сближение и контакт между энхансером и промотором, является на данный момент наиболее изученной [2].

В данной работе для того чтобы установить, что при активации энхансером промотора в *транс* регуляторные элементы, расположенные в различных плаزمидях вступают в контакт, мы применили оригинальный 3С метод, совмещенный с ПЦР в реальном времени [234].

Для этого нами была выполнена совместная трансфекция клеток НерG2 смесью, содержащей эквимольные количества трех плазмид: плазмиды, содержащей ген люциферазы светлячка под контролем минимального промотора CMV (Pcmv), плазмиды, содержащей энхансер CMV (Ecmv) и плазмиды (L), содержащей аналогичный по длине нейтральный фрагмент ДНК фага лямбда (рисунок 22а). После сшивки, очистки и лигирования ДНК (раздел 3.2.13.), частоту событий лигирования последовательностей промотор-энхансер и промотор-контрольный фрагмент определяли с помощью количественной ПЦР с использованием специфичных праймеров (таблица 6 глава 3 «Материалы и методы»), подобранных для детектирования соответствующих продуктов лигирования. Результаты представлены на рисунке 22. На рисунке 22б показано схематическое представление предполагаемых комплексов, обнаруженных с помощью соответствующих пар праймеров. Рисунок 22 в представляет относительное содержание (раздел 3.2.14) продуктов лигирования Pcmv-Ecmv и

Pcmv-L, показанных как среднее из двух независимых экспериментов трансфекций и последующего ЗС.

В отрицательном контроле, когда стадия сшивки формальдегидом была пропущена, продукты лигирования не были обнаружены даже после 40 циклов ПЦР в реальном времени.



**Рисунок. 22.** Сближение регуляторных элементов при активации промотора энхансером *in trans* как было показано с использованием метода фиксации конформации хромосом (ЗС) в клетках НерG2. (А) - схематическое представление котрансфицируемых плазмид; (Б) - предполагаемые комплексы, обнаруженные с использованием соответствующих пар праймеров; (В) - относительное содержание продуктов лигирования. Pcmv, Ecmv - CMV промотор и энхансер, соответственно. L - отрицательный контроль, содержащий фрагмент ДНК фага лямбда. LUC - репортерный ген люциферазы светлячка. Относительное содержание ДНК отражает частоту событий лигирования соответствующих регуляторных элементов.

Согласно полученным данным (рисунок 22) содержание ДНК (и соответствующая частота лигирования) Pcmv и Ecmv была в 50 раз выше, чем у Pcmv и контрольной последовательности фага лямбда. Эти данные показывают, что CMV промотор и энхансер специфично сблизилась, при котрансфекции в клетки НерG2. Продукты лигирования Pcmv-Ecmv или Pcmv-L в контроле без сшивки хроматина обнаружены не были.

Для того, чтобы исключить возможность получения ложных результатов из-за разной эффективности проникновения в ядро клетки плазмид, содержащих энхансерные и контрольные последовательности, мы сравнили относительные

количества Ecmv и L ДНК в ядрах клеток НерG2 после трансфекции. Для этого клетки НерG2 трансфицировали эквимольной смесью плазмид Pcmv, Ecmv и L, затем выделяли ДНК так же, как и для ЗС, но без стадии гидролиза *Msp* I и последующего лигирования. Полученную ДНК использовали в качестве матрицы для ПЦР в реальном времени с парами праймеров, специфичных для Ecmv и L плазмид.

Рассчитанное нормированное количество  $X_N$  (см. раздел 3.2.14.), составило 0,53 и 0,50 для Ecmv и L плазмид, соответственно, что свидетельствует о том, что обе конструкции проникают в ядро с практически одинаковой эффективностью.

Таким образом нами было показано, что способность активировать промотор *в транс* при транзистентных трансфекциях не ограничена лишь вирусными энхансерами, как было показано ранее [155], данное свойство также характерно и для геномных регуляторных элементов. Активация промотора энхансером *в транс* зависит от промотора и типа клеток и, как было показано с помощью ЗС анализа, включает в себя физический контакт промотора с энхансером.

Мы полагаем, что метод изучения энхансерной активности *в транс* с использованием совместной трансфекции и репортерного гена может быть удобной моделью для изучения многих важных свойств дистальных взаимодействий энхансера с промотором, включающих клеточную специфичность, специфичность промотора / энхансера, роль различных транскрипционных факторов и влияние структуры хроматина и мутаций в регуляторных элементах на эффект активации.



## 5. ЗАКЛЮЧЕНИЕ И ВЫВОДЫ

В данной работе проведен поиск и анализ различных типов регуляторных последовательностей, представленных в геноме человека, показана взаимосвязь между ними. Полученные данные позволяют сделать следующие выводы:

1. Проведен функциональный анализ энхансера, расположенного в локусе *FXVD5-COX7A1* хромосомы 19 человека и гомологичных ему последовательностей приматов. Определена консервативная область, отвечающая за энхансерную активность. При помощи метода EMSA показана функциональная консервативность гомологичных энхансеру последовательностей приматов при связывании с белками.

2. Определены точки начала транскрипции мРНК генов *PSENE1* и *U2AF1L4*. Продемонстрирована согласованная экспрессия данных генов в нескольких клеточных линиях. Определены консервативные области и показана их функциональная роль в работе промотора. Таким образом показано, что область между генами *PSENE1* и *U2AF1L4* на хромосоме 19 человека является истинным двунаправленным промотором.

3. Проведен функциональный анализ энхансерной, сайленсерной и энхансер-блокирующей активности последовательностей, расположенных в локусе *FXVD5-COX7A1* хромосомы 19 человека.

4. Показано, что активация промотора энхансером в *транс*-положении возможна не только для вирусных энхансеров, но и для геномных регуляторных элементов. Активность энхансера в *транс*-положении зависит от промотора и типа клеток и, как было показано с помощью 3С анализа, включает в себя физическую близость промотора с энхансером.

## 6. СПИСОК ЦИТИРОВАННОЙ ЛИТЕРАТУРЫ

1. Cavalli G, Misteli T: **Functional implications of genome topology.** *Nat Struct Mol Biol* 2013, **20**(3):290-299.
2. Levine M, Cattoglio C, Tjian R: **Looping back to leap forward: transcription enters a new era.** *Cell* 2014, **157**(1):13-25.
3. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W *et al*: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**(6822):860-921.
4. Grossman SR, Andersen KG, Shlyakhter I, Tabrizi S, Winnicki S, Yen A, Park DJ, Griesemer D, Karlsson EK, Wong SH *et al*: **Identifying recent adaptations in large-scale genomic data.** *Cell* 2013, **152**(4):703-713.
5. Fraser HB: **Gene expression drives local adaptation in humans.** *Genome Res* 2013, **23**(7):1089-1096.
6. Mouse Genome Sequencing C, Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M *et al*: **Initial sequencing and comparative analysis of the mouse genome.** *Nature* 2002, **420**(6915):520-562.
7. Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, Kheradpour P, Ernst J, Jordan G, Mauceli E *et al*: **A high-resolution map of human evolutionary constraint using 29 mammals.** *Nature* 2011, **478**(7370):476-482.
8. Ponting CP, Hardison RC: **What fraction of the human genome is functional?** *Genome Res* 2011, **21**(11):1769-1776.
9. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J *et al*: **Systematic localization of common disease-associated variation in regulatory DNA.** *Science* 2012, **337**(6099):1190-1195.
10. Ward LD, Kellis M: **HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants.** *Nucleic acids research* 2012, **40**(Database issue):D930-934.
11. Akopov SB, Ruda VM, Batrak VV, Vetchinova AS, Chernov IP, Nikolaev LG, Bode J, Sverdlov ED: **Identification, genome mapping, and CTCF binding of potential insulators within the FXYD5-COX7A1 locus of human chromosome 19q13.12.** *Mammalian genome : official journal of the International Mammalian Genome Society* 2006, **17**(10):1042-1049.
12. Дидыч ДА, Акопов СБ, Снежков ЕВ, Скапцова НВ, Николаев ЛГ, Сverdlov ЕД: **Идентификация и картирование 10 новых потенциальных инсуляторов в области FXYD5-COX7A1 хромосомы 19q13.12 человека.** *Биохимия* 2009, **74**(7):899-905.
13. Chernov I, Stukacheva E, Akopov S, Didych D, Nikolaev L, Sverdlov E: **A new technique for selective identification and mapping of enhancers within long genomic sequences.** *BioTechniques* 2008, **44**(6):775-784.
14. Yang J, Corces VG: **Insulators, long-range interactions, and genome function.** *Current opinion in genetics & development* 2012, **22**:86-92.
15. Van Bortle K, Nichols MH, Li L, Ong CT, Takenaka N, Qin ZS, Corces VG: **Insulator function and topological domain border strength scale with architectural protein occupancy.** *Genome biology* 2014, **15**(6):R82.
16. Razin SV, Gavrillov AA, Vassetzky YS, Ulianov SV: **Topologically-associating domains: gene warehouses adapted to serve transcriptional regulation.** *Transcription* 2016, **7**(3):84-90.
17. Sanyal A, Lajoie BR, Jain G, Dekker J: **The long-range interaction landscape of gene promoters.** *Nature* 2012, **489**(7414):109-113.
18. Didych DA, Kotova ES, Akopov SB, Nikolaev LG, Sverdlov ED: **DNA fragments binding CTCF in vitro and in vivo are capable of blocking enhancer activity.** *BMC research notes* 2012, **5**:178.
19. Sanborn AL, Rao SS, Huang SC, Durand NC, Huntley MH, Jewett AI, Bochkov ID, Chinnappan D, Cutkosky A, Li J *et al*: **Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes.** *Proc Natl Acad Sci U S A* 2015, **112**(47):E6456-6465.
20. Beygo J, Citro V, Sparago A, De Crescenzo A, Cerrato F, Heitmann M, Rademacher K, Guala A, Enklaar T, Anichini C *et al*: **The molecular function and clinical phenotype of partial deletions of the IGF2/H19**

- imprinting control region depends on the spatial arrangement of the remaining CTCF-binding sites.** *Human molecular genetics* 2013, **22**(3):544-557.
21. Raab JR, Chiu J, Zhu J, Katzman S, Kurukuti S, Wade PA, Haussler D, Kamakaka RT: **Human tRNA genes function as chromatin insulators.** *The EMBO journal* 2012, **31**(2):330-350.
  22. Carriere L, Graziani S, Alibert O, Ghavi-Helm Y, Boussouar F, Humbertclaude H, Jounier S, Aude JC, Keime C, Murvai J *et al*: **Genomic binding of Pol III transcription machinery and relationship with TFIIIS transcription factor distribution in mouse embryonic stem cells.** *Nucleic acids research* 2012, **40**(1):270-283.
  23. Moqtaderi Z, Wang J, Raha D, White RJ, Snyder M, Weng Z, Struhl K: **Genomic binding profiles of functionally distinct RNA polymerase III transcription complexes in human cells.** *Nat Struct Mol Biol* 2010, **17**(5):635-640.
  24. Nègre N, Brown CD, Shah PK, Kheradpour P, Morrison Ca, Henikoff JG, Feng X, Ahmad K, Russell S, White RaH *et al*: **A comprehensive map of insulator elements for the Drosophila genome.** *PLoS genetics* 2010, **6**:e1000814.
  25. Baiamonte E, Spinelli G, Maggio A, Acuto S, Cavalieri V: **The Sea Urchin sns5 Chromatin Insulator Shapes the Chromatin Architecture of a Lentivirus Vector Integrated in the Mammalian Genome.** *Nucleic acid therapeutics* 2016.
  26. Kurukuti S, Tiwari VK, Tavoosidana G, Pugacheva E, Murrell A, Zhao Z, Lobanenkov V, Reik W, Ohlsson R: **CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2.** *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**:10684-10689.
  27. Cubenas-Potts C, Corces VG: **Topologically Associating Domains: An invariant framework or a dynamic scaffold?** *Nucleus* 2015, **6**(6):430-434.
  28. Xu Z, Wei G, Chepelev I, Zhao K, Felsenfeld G: **Mapping of INS promoter interactions reveals its role in long-range regulation of SYT8 transcription.** *Nature structural & molecular biology* 2011, **18**:372-378.
  29. Lin SG, Guo C, Su A, Zhang Y, Alt FW: **CTCF-binding elements 1 and 2 in the Igh intergenic control region cooperatively regulate V(D)J recombination.** *Proc Natl Acad Sci U S A* 2015, **112**(6):1815-1820.
  30. Van Bortle K, Ramos E, Takenaka N, Yang J, Wahi JE, Corces VG: **Drosophila CTCF tandemly aligns with other insulator proteins at the borders of H3K27me3 domains.** *Genome Res* 2012, **22**(11):2176-2187.
  31. Lobanenkov VV, Nicolas RH, Adler VV, Paterson H, Klenova EM, Polotskaja AV, Goodwin GH: **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* 1990, **5**(12):1743-1753.
  32. Ohlsson R, Renkawitz R, Lobanenkov V: **CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease.** *Trends in genetics : TIG* 2001, **17**(9):520-527.
  33. Ohlsson R, Lobanenkov V, Klenova E: **Does CTCF mediate between nuclear organization and gene expression?** *BioEssays : news and reviews in molecular, cellular and developmental biology* 2010, **32**(1):37-50.
  34. Tiffen JC, Bailey CG, Marshall AD, Metierre C, Feng Y, Wang Q, Watson SL, Holst J, Rasko JE: **The cancer-testis antigen BORIS phenocopies the tumor suppressor CTCF in normal and neoplastic cells.** *International journal of cancer* 2013, **133**(7):1603-1613.
  35. Moore JM, Rabaia NA, Smith LE, Fagerlie S, Gurley K, Loukinov D, Disteche CM, Collins SJ, Kemp CJ, Lobanenkov VV *et al*: **Loss of maternal CTCF is associated with peri-implantation lethality of Ctfc null embryos.** *PloS one* 2012, **7**(4):e34915.
  36. Dekker J: **CTCF and cohesin help neurons raise their self-awareness.** *Proc Natl Acad Sci U S A* 2012, **109**(23):8799-8800.
  37. Hirayama T, Tarusawa E, Yoshimura Y, Galjart N, Yagi T: **CTCF is required for neural development and stochastic expression of clustered Pcdh genes in neurons.** *Cell reports* 2012, **2**(2):345-357.
  38. Guo C, Yoon HS, Franklin A, Jain S, Ebert A, Cheng HL, Hansen E, Despo O, Bossen C, Vettermann C *et al*: **CTCF-binding elements mediate control of V(D)J recombination.** *Nature* 2011, **477**(7365):424-430.
  39. Mustafa M, Lee JY, Kim MH: **CTCF negatively regulates HOXA10 expression in breast cancer cells.** *Biochemical and biophysical research communications* 2015, **467**(4):828-834.
  40. Bell AC, Felsenfeld G: **Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene.** *Nature* 2000, **405**(6785):482-485.

41. Qi H, Liu M, Emery DW, Stamatoyannopoulos G: **Functional validation of a constitutive autonomous silencer element.** *PLoS one* 2015, **10**(4):e0124588.
42. Ding Z, Ni Y, Timmer SW, Lee BK, Battenhouse A, Louzada S, Yang F, Dunham I, Crawford GE, Lieb JD *et al*: **Correction: Quantitative genetics of CTCF binding reveal local sequence effects and different modes of X-chromosome association.** *PLoS Genet* 2015, **11**(4):e1005177.
43. Nikolaev LG, Akopov SB, Didych DA, Sverdlov ED: **Vertebrate Protein CTCF and its Multiple Roles in a Large-Scale Regulation of Genome Activity.** *Current genomics* 2009, **10**(5):294-302.
44. Herold M, Bartkuhn M, Renkawitz R: **CTCF: insights into insulator function during development.** *Development* 2012, **139**(6):1045-1057.
45. Phillips JE, Corces VG: **CTCF: master weaver of the genome.** *Cell* 2009, **137**(7):1194-1211.
46. Kurukuti S, Tiwari VK, Tavoosidana G, Pugacheva E, Murrell A, Zhao Z, Lobanenkov V, Reik W, Ohlsson R: **CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2.** *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**(28):10684-10689.
47. Majumder P, Gomez JA, Chadwick BP, Boss JM: **The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions.** *The Journal of experimental medicine* 2008, **205**(4):785-798.
48. Lutz M, Burke LJ, Barreto G, Goeman F, Greb H, Arnold R, Schultheiss H, Brehm A, Kouzarides T, Lobanenkov V *et al*: **Transcriptional repression by the insulator protein CTCF involves histone deacetylases.** *Nucleic acids research* 2000, **28**(8):1707-1713.
49. Lutz M, Burke LJ, LeFevre P, Myers FA, Thorne AW, Crane-Robinson C, Bonifer C, Filippova GN, Lobanenkov V, Renkawitz R: **Thyroid hormone-regulated enhancer blocking: cooperation of CTCF and thyroid hormone receptor.** *The EMBO journal* 2003, **22**(7):1579-1587.
50. Yusufzai TM, Tagami H, Nakatani Y, Felsenfeld G: **CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species.** *Molecular cell* 2004, **13**(2):291-298.
51. Defossez PA, Kelly KF, Filion GJ, Perez-Torrado R, Magdinier F, Menoni H, Nordgaard CL, Daniel JM, Gilson E: **The human enhancer blocker CTC-binding factor interacts with the transcription factor Kaiso.** *The Journal of biological chemistry* 2005, **280**(52):43017-43023.
52. Yao H, Brick K, Evrard Y, Xiao T, Camerini-Otero RD, Felsenfeld G: **Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA.** *Genes & development* 2010, **24**(22):2543-2555.
53. Ross-Innes CS, Brown GD, Carroll JS: **A co-ordinated interaction between CTCF and ER in breast cancer cells.** *BMC genomics* 2011, **12**:593.
54. Liu Z, Scannell DR, Eisen MB, Tjian R: **Control of embryonic stem cell lineage commitment by core promoter factor, TAF3.** *Cell* 2011, **146**(5):720-731.
55. de Laat W, Grosveld F: **Spatial organization of gene expression: the active chromatin hub.** *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* 2003, **11**(5):447-459.
56. Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W: **Looping and interaction between hypersensitive sites in the active beta-globin locus.** *Molecular cell* 2002, **10**(6):1453-1465.
57. Splinter E, Heath H, Kooren J, Palstra RJ, Klous P, Grosveld F, Galjart N, de Laat W: **CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus.** *Genes & development* 2006, **20**(17):2349-2354.
58. Zuin J, Dixon JR, van der Reijden MI, Ye Z, Kolovos P, Brouwer RW, van de Corput MP, van de Werken HJ, Knoch TA, van IWF *et al*: **Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells.** *Proc Natl Acad Sci U S A* 2014, **111**(3):996-1001.
59. Xie X, Mikkelsen TS, Gnirke A, Lindblad-Toh K, Kellis M, Lander ES: **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* 2007, **104**(17):7145-7150.
60. Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, Zhang MQ, Lobanenkov VV, Ren B: **Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome.** *Cell* 2007, **128**(6):1231-1245.

61. Cuddapah S, Jothi R, Schones DE, Roh TY, Cui K, Zhao K: **Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains.** *Genome Res* 2009, **19**(1):24-32.
62. Chen H, Tian Y, Shu W, Bo X, Wang S: **Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome.** *PLoS one* 2012, **7**(7):e41374.
63. Wang H, Maurano MT, Qu H, Varley KE, Gertz J, Pauli F, Lee K, Canfield T, Weaver M, Sandstrom R *et al*: **Widespread plasticity in CTCF occupancy linked to DNA methylation.** *Genome Res* 2012, **22**(9):1680-1688.
64. Li Y, Huang W, Niu L, Umbach DM, Covo S, Li L: **Characterization of constitutive CTCF/cohesin loci: a possible role in establishing topological domains in mammalian genomes.** *BMC genomics* 2013, **14**:553.
65. Nasmyth K, Haering CH: **Cohesin: its roles and mechanisms.** *Annual review of genetics* 2009, **43**:525-558.
66. Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, Tsutsumi S, Nagae G, Ishihara K, Mishiro T *et al*: **Cohesin mediates transcriptional insulation by CCCTC-binding factor.** *Nature* 2008, **451**(7180):796-801.
67. Rubio ED, Reiss DJ, Welcsh PL, Disteché CM, Filippova GN, Baliga NS, Aebersold R, Ranish JA, Krumm A: **CTCF physically links cohesin to chromatin.** *Proc Natl Acad Sci U S A* 2008, **105**(24):8309-8314.
68. Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS *et al*: **Mediator and cohesin connect gene expression and chromatin architecture.** *Nature* 2010, **467**(7314):430-435.
69. Yan J, Enge M, Whittington T, Dave K, Liu J, Sur I, Schmierer B, Jolma A, Kivioja T, Taipale M *et al*: **Transcription factor binding in human cells occurs in dense clusters formed around cohesin anchor sites.** *Cell* 2013, **154**(4):801-813.
70. Xiao T, Wallace J, Felsenfeld G: **Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity.** *Molecular and cellular biology* 2011, **31**(11):2174-2183.
71. Nativio R, Wendt KS, Ito Y, Huddleston JE, Uribe-Lewis S, Woodfine K, Krueger C, Reik W, Peters JM, Murrell A: **Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus.** *PLoS Genet* 2009, **5**(11):e1000739.
72. Hadjur S, Williams LM, Ryan NK, Cobb BS, Sexton T, Fraser P, Fisher AG, Merckenschlager M: **Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus.** *Nature* 2009, **460**(7253):410-413.
73. Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, Oberdoerffer P, Sandberg R, Oberdoerffer S: **CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing.** *Nature* 2011, **479**(7371):74-79.
74. Chien R, Zeng W, Kawauchi S, Bender MA, Santos R, Gregson HC, Schmiesing JA, Newkirk DA, Kong X, Ball AR, Jr. *et al*: **Cohesin mediates chromatin interactions that regulate mammalian beta-globin expression.** *The Journal of biological chemistry* 2011, **286**(20):17870-17878.
75. Hebbes TR, Clayton AL, Thorne AW, Crane-Robinson C: **Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain.** *The EMBO journal* 1994, **13**(8):1823-1830.
76. Prioleau MN, Nony P, Simpson M, Felsenfeld G: **An insulator element and condensed chromatin region separate the chicken beta-globin locus from an independently regulated erythroid-specific folate receptor gene.** *The EMBO journal* 1999, **18**(14):4035-4048.
77. Recillas-Targa F, Pikaart MJ, Burgess-Beusse B, Bell AC, Litt MD, West AG, Gaszner M, Felsenfeld G: **Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities.** *Proc Natl Acad Sci U S A* 2002, **99**(10):6883-6888.
78. Yusufzai TM, Felsenfeld G: **The 5'-HS4 chicken beta-globin insulator is a CTCF-dependent nuclear matrix-associated element.** *Proc Natl Acad Sci U S A* 2004, **101**(23):8620-8624.
79. West AG, Huang S, Gaszner M, Litt MD, Felsenfeld G: **Recruitment of histone modifications by USF proteins at a vertebrate barrier element.** *Molecular cell* 2004, **16**(3):453-463.

80. Li X, Wang S, Li Y, Deng C, Steiner LA, Xiao H, Wu C, Bungert J, Gallagher PG, Felsenfeld G *et al*: **Chromatin boundaries require functional collaboration between the hSET1 and NURF complexes.** *Blood* 2011, **118**(5):1386-1394.
81. Huang S, Li X, Yusufzai TM, Qiu Y, Felsenfeld G: **USF1 recruits histone modification complexes and is critical for maintenance of a chromatin barrier.** *Molecular and cellular biology* 2007, **27**(22):7991-8002.
82. Dickson J, Gowher H, Strogantsev R, Gaszner M, Hair A, Felsenfeld G, West AG: **VEZF1 elements mediate protection from DNA methylation.** *PLoS Genet* 2010, **6**(1):e1000804.
83. Raab JR, Kamakaka RT: **Insulators and promoters: closer than we think.** *Nature reviews Genetics* 2010, **11**(6):439-446.
84. Ottaviani A, Rival-Gervier S, Boussouar A, Foerster AM, Rondier D, Sacconi S, Desnuelle C, Gilson E, Magdinier F: **The D4Z4 macrosatellite repeat acts as a CTCF and A-type lamins-dependent insulator in facio-scapulo-humeral dystrophy.** *PLoS Genet* 2009, **5**(2):e1000394.
85. Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W *et al*: **Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions.** *Nature* 2008, **453**(7197):948-951.
86. Calo E, Wysocka J: **Modification of enhancer chromatin: what, how, and why?** *Molecular cell* 2013, **49**(5):825-837.
87. Berger SL: **The complex language of chromatin regulation during transcription.** *Nature* 2007, **447**(7143):407-412.
88. Bi X, Broach JR: **UASrpg can function as a heterochromatin boundary element in yeast.** *Genes & development* 1999, **13**(9):1089-1101.
89. Core LJ, Lis JT: **Paused Pol II captures enhancer activity and acts as a potent insulator.** *Genes & development* 2009, **23**(14):1606-1612.
90. Valenzuela L, Kamakaka RT: **Chromatin insulators.** *Annual review of genetics* 2006, **40**:107-138.
91. Blackwood EM, Kadonaga JT: **Going the distance: a current view of enhancer action.** *Science* 1998, **281**(5373):60-63.
92. Tang Z, Luo OJ, Li X, Zheng M, Zhu JJ, Szalaj P, Trzaskoma P, Magalska A, Wlodarczyk J, Rusczycki B *et al*: **CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription.** *Cell* 2015, **163**(7):1611-1627.
93. Handoko L, Xu H, Li G, Ngan CY, Chew E, Schnapp M, Lee CWH, Ye C, Ping JLH, Mulawadi F *et al*: **CTCF-mediated functional chromatin interactome in pluripotent cells.** *Nature genetics* 2011, **43**:630-638.
94. Gohl D, Aoki T, Blanton J, Shanower G, Kappes G, Schedl P: **Mechanism of chromosomal boundary action: roadblock, sink, or loop?** *Genetics* 2011, **187**(3):731-748.
95. Paulson JR, Laemmli UK: **The structure of histone-depleted metaphase chromosomes.** *Cell* 1977, **12**(3):817-828.
96. Geyer PK: **The role of insulator elements in defining domains of gene expression.** *Curr Opin Genet Dev* 1997, **7**(2):242-248.
97. Kulaeva OI, Nizovtseva EV, Polikanov YS, Ulianov SV, Studitsky VM: **Distant activation of transcription: mechanisms of enhancer action.** *Molecular and cellular biology* 2012, **32**(24):4892-4897.
98. West AG, Fraser P: **Remote control of gene transcription.** *Human molecular genetics* 2005, **14 Spec No 1**:R101-111.
99. Roman AC, Gonzalez-Rico FJ, Molto E, Hernando H, Neto A, Vicente-Garcia C, Ballestar E, Gomez-Skarmeta JL, Vavrova-Anderson J, White RJ *et al*: **Dioxin receptor and SLUG transcription factors regulate the insulator activity of B1 SINE retrotransposons via an RNA polymerase switch.** *Genome Res* 2011, **21**(3):422-432.
100. Wang J, Vicente-Garcia C, Seruggia D, Molto E, Fernandez-Minan A, Neto A, Lee E, Gomez-Skarmeta JL, Montoliu L, Lunyak VV *et al*: **MIR retrotransposon sequences provide insulators to the human genome.** *Proc Natl Acad Sci U S A* 2015, **112**(32):E4428-4437.
101. Engel N, Raval AK, Thorvaldsen JL, Bartolomei SM: **Three-dimensional conformation at the H19/Igf2 locus supports a model of enhancer tracking.** *Human molecular genetics* 2008, **17**(19):3021-3029.
102. Banerji J, Rusconi S, Schaffner W: **Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences.** *Cell* 1981, **27**(2 Pt 1):299-308.

103. Banerji J, Olson L, Schaffner W: **A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes.** *Cell* 1983, **33**(3):729-740.
104. Sheng Y, Previti C: **Genomic features and computational identification of human microRNAs under long-range developmental regulation.** *BMC genomics* 2011, **12**:270.
105. Cohen ML, Kim S, Morita K, Kim SH, Han M: **The GATA factor elt-1 regulates C. elegans developmental timing by promoting expression of the let-7 family microRNAs.** *PLoS Genet* 2015, **11**(3):e1005099.
106. Kim A, Dean A: **Chromatin loop formation in the beta-globin locus and its role in globin gene transcription.** *Molecules and cells* 2012, **34**(1):1-5.
107. El-Sherif E, Levine M: **Shadow Enhancers Mediate Dynamic Shifts of Gap Gene Expression in the Drosophila Embryo.** *Current biology : CB* 2016, **26**(9):1164-1169.
108. Farley EK, Olson KM, Zhang W, Brandt AJ, Rokhsar DS, Levine MS: **Suboptimization of developmental enhancers.** *Science* 2015, **350**(6258):325-328.
109. Didych DA, Tyulkina DV, Pleshkan VV, Alekseenko IV, Sverdlov ED: **[ Super-enhancers. Are they regulators of regulatory genes of development and cancer?].** *Molekuliarnaia biologii* 2015, **49**(6):915-922.
110. Liu Z, Merkurjev D, Yang F, Li W, Oh S, Friedman MJ, Song X, Zhang F, Ma Q, Ohgi KA *et al*: **Enhancer activation requires trans-recruitment of a mega transcription factor complex.** *Cell* 2014, **159**(2):358-373.
111. Blow MJ, McCulley DJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F *et al*: **ChIP-Seq identification of weakly conserved heart enhancers.** *Nat Genet* 2010, **42**(9):806-810.
112. Creighton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA *et al*: **Histone H3K27ac separates active from poised enhancers and predicts developmental state.** *Proc Natl Acad Sci U S A* 2010, **107**(50):21931-21936.
113. Davidson EH: **Genomic Regulatory Systems: Development and Evolution.** *Academic, New York, 2001* 2001.
114. Plank JL, Dean A: **Enhancer function: mechanistic and genome-wide insights come together.** *Molecular cell* 2014, **55**(1):5-14.
115. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching Ka *et al*: **Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome.** *Nature genetics* 2007, **39**:311-318.
116. He HH, Meyer CA, Shin H, Bailey ST, Wei G, Wang Q, Zhang Y, Xu K, Ni M, Lupien M *et al*: **Nucleosome dynamics define transcriptional enhancers.** *Nat Genet* 2010, **42**(4):343-347.
117. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B *et al*: **The accessible chromatin landscape of the human genome.** *Nature* 2012, **489**(7414):75-82.
118. Goldberg AD, Banaszynski LA, Noh KM, Lewis PW, Elsaesser SJ, Stadler S, Dewell S, Law M, Guo X, Li X *et al*: **Distinct factors control histone variant H3.3 localization at specific genomic regions.** *Cell* 2010, **140**(5):678-691.
119. Buchanan CC, Torstenson ES, Bush WS, Ritchie MD: **A comparison of cataloged variation between International HapMap Consortium and 1000 Genomes Project data.** *Journal of the American Medical Informatics Association : JAMIA* 2012, **19**(2):289-294.
120. Yanez-Cuna JO, Kvon EZ, Stark A: **Deciphering the transcriptional cis-regulatory code.** *Trends in genetics : TIG* 2013, **29**(1):11-22.
121. Richardson L, Venkataraman S, Stevenson P, Yang Y, Moss J, Graham L, Burton N, Hill B, Rao J, Baldock RA *et al*: **EMAGE mouse embryo spatial gene expression database: 2014 update.** *Nucleic acids research* 2014, **42**(Database issue):D835-844.
122. Brown CD, Johnson DS, Sidow A: **Functional architecture and evolution of transcriptional elements that drive gene coexpression.** *Science* 2007, **317**(5844):1557-1560.
123. Ludwig MZ, Palsson A, Alekseeva E, Bergman CM, Nathan J, Kreitman M: **Functional evolution of a cis-regulatory module.** *PLoS biology* 2005, **3**(4):e93.
124. Atchison ML: **Enhancers: Mechanisms of Action and Cell Specificity.** *Annual Review of Cell Biology* 1988, **4**(1):127-153.
125. Kulkarni MM, Arnosti DN: **Information display by transcriptional enhancers.** *Development* 2003, **130**(26):6569-6575.

126. Junion G, Spivakov M, Girardot C, Braun M, Gustafson EH, Birney E, Furlong EE: **A transcription factor collective defines cardiac cell fate and reflects lineage history.** *Cell* 2012, **148**(3):473-486.
127. Stinski MF, Isomura H: **Role of the cytomegalovirus major immediate early enhancer in acute infection and reactivation from latency.** *Medical microbiology and immunology* 2008, **197**(2):223-231.
128. Meier JL, Keller MJ, McCoy JJ: **Requirement of multiple cis-acting elements in the human cytomegalovirus major immediate-early distal enhancer for viral gene expression and replication.** *Journal of virology* 2002, **76**(1):313-326.
129. Benedict CA, Angulo A, Patterson G, Ha S, Huang H, Messerle M, Ware CF, Ghazal P: **Neutrality of the canonical NF-kappaB-dependent pathway for human and murine cytomegalovirus transcription and replication in vitro.** *Journal of virology* 2004, **78**(2):741-750.
130. Herr W, Clarke J: **The SV40 enhancer is composed of multiple functional elements that can compensate for one another.** *Cell* 1986, **45**(3):461-470.
131. Crocker J, Tamori Y, Erives A: **Evolution acts on enhancer organization to fine-tune gradient threshold readouts.** *PLoS biology* 2008, **6**(11):e263.
132. Smith RP, Taher L, Patwardhan RP, Kim MJ, Inoue F, Shendure J, Ovcharenko I, Ahituv N: **Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model.** *Nat Genet* 2013, **45**(9):1021-1028.
133. Merika M, Thanos D: **Enhanceosomes.** *Curr Opin Genet Dev* 2001, **11**(2):205-208.
134. Panne D: **The enhanceosome.** *Curr Opin Struct Biol* 2008, **18**(2):236-242.
135. He J, Ye J, Cai Y, Riquelme C, Liu JO, Liu X, Han A, Chen L: **Structure of p300 bound to MEF2 on DNA reveals a mechanism of enhanceosome assembly.** *Nucleic acids research* 2011, **39**(10):4464-4474.
136. Maniatis T, Falvo JV, Kim TH, Kim TK, Lin CH, Parekh BS, Wathélet MG: **Structure and function of the interferon-beta enhanceosome.** *Cold Spring Harb Symp Quant Biol* 1998, **63**:609-620.
137. Kazemian M, Pham H, Wolfe SA, Brodsky MH, Sinha S: **Widespread evidence of cooperative DNA binding by transcription factors in Drosophila development.** *Nucleic acids research* 2013, **41**(17):8237-8252.
138. Papatsenko D, Levine M: **A rationale for the enhanceosome and other evolutionarily constrained enhancers.** *Current biology : CB* 2007, **17**(22):R955-957.
139. Liu F, Posakony JW: **Role of architecture in the function and specificity of two Notch-regulated transcriptional enhancer modules.** *PLoS Genet* 2012, **8**(7):e1002796.
140. Uhl JD, Zandvakili A, Gebelein B: **A Hox Transcription Factor Collective Binds a Highly Conserved Distal-less cis-Regulatory Module to Generate Robust Transcriptional Outcomes.** *PLoS Genet* 2016, **12**(4):e1005981.
141. Weirauch MT, Hughes TR: **Conserved expression without conserved regulatory sequence: the more things change, the more they stay the same.** *Trends in genetics : TIG* 2010, **26**(2):66-74.
142. Szutorisz H, Dillon N, Tora L: **The role of enhancers as centres for general transcription factor recruitment.** *Trends in biochemical sciences* 2005, **30**:593-599.
143. Conaway JW, Florens L, Sato S, Tomomori-Sato C, Parmely TJ, Yao T, Swanson SK, Banks CA, Washburn MP, Conaway RC: **The mammalian Mediator complex.** *FEBS Lett* 2005, **579**(4):904-908.
144. Lorberbaum DS, Barolo S: **Gene regulation: when analog beats digital.** *Current biology : CB* 2013, **23**(23):R1054-1056.
145. Giorgetti L, Siggers T, Tiana G, Caprara G, Notarbartolo S, Corona T, Pasparakis M, Milani P, Bulyk ML, Natoli G: **Noncooperative interactions between transcription factors and clustered DNA binding sites enable graded transcriptional responses to environmental inputs.** *Molecular cell* 2010, **37**(3):418-428.
146. Rossi FM, Kringstein AM, Spicher A, Guicherit OM, Blau HM: **Transcriptional control: rheostat converted to on/off switch.** *Molecular cell* 2000, **6**(3):723-728.
147. Walters MC, Fiering S, Eidemiller J, Magis W, Groudine M, Martin DI: **Enhancers increase the probability but not the level of gene expression.** *Proc Natl Acad Sci U S A* 1995, **92**(15):7125-7129.
148. Lopes FJ, Spirov AV, Bisch PM: **The role of Bicoid cooperative binding in the patterning of sharp borders in Drosophila melanogaster.** *Developmental biology* 2012, **370**(2):165-172.
149. Segal E, Raveh-Sadka T, Schroeder M, Unnerstall U, Gaul U: **Predicting expression patterns from regulatory sequence in Drosophila segmentation.** *Nature* 2008, **451**(7178):535-540.



150. Ong CT, Corces VG: **CTCF: an architectural protein bridging genome topology and function.** *Nature reviews Genetics* 2014, **15**(4):234-246.
151. Krivega I, Dean A: **Enhancer and promoter interactions-long distance calls.** *Current opinion in genetics & development* 2012, **22**:79-85.
152. Zhu X, Ling J, Zhang L, Pi W, Wu M, Tuan D: **A facilitated tracking and transcription mechanism of long-range enhancer function.** *Nucleic acids research* 2007, **35**(16):5532-5544.
153. Hatzis P, Talianidis I: **Dynamics of enhancer-promoter communication during differentiation-induced gene activation.** *Molecular cell* 2002, **10**:1467-1477.
154. Krivega I, Dale RK, Dean A: **Role of LDB1 in the transition from chromatin looping to transcription activation.** *Genes & development* 2014, **28**(12):1278-1290.
155. D'Aiuto L, De Marco R, Edward N, Rizzo A, Chaillet JR, Montecalvo A, Lotze MT, Gambotto A: **Evidence of the capability of the CMV enhancer to activate in trans gene expression in mammalian cells.** *DNA and cell biology* 2006, **25**(3):171-180.
156. Mellert DJ, Truman JW: **Transvection is common throughout the Drosophila genome.** *Genetics* 2012, **191**(4):1129-1141.
157. Noordermeer D, de Wit E, Klous P, van de Werken H, Simonis M, Lopez-Jones M, Eussen B, de Klein A, Singer RH, de Laat W: **Variiegated gene expression caused by cell-specific long-range DNA interactions.** *Nature cell biology* 2011, **13**(8):944-951.
158. Deng W, Lee J, Wang H, Miller J, Reik A, Gregory PD, Dean A, Blobel GA: **Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor.** *Cell* 2012, **149**(6):1233-1244.
159. Bickmore WA, van Steensel B: **Genome architecture: domain organization of interphase chromosomes.** *Cell* 2013, **152**(6):1270-1284.
160. de Laat W, Dekker J: **3C-based technologies to study the shape of the genome.** *Methods* 2012, **58**(3):189-191.
161. Ong CT, Corces VG: **Enhancer function: new insights into the regulation of tissue-specific gene expression.** *Nature reviews Genetics* 2011, **12**(4):283-293.
162. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES *et al*: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell* 2014, **159**(7):1665-1680.
163. Edelman LB, Fraser P: **Transcription factories: genetic programming in three dimensions.** *Curr Opin Genet Dev* 2012, **22**(2):110-114.
164. van Arensbergen J, van Steensel B, Bussemaker HJ: **In search of the determinants of enhancer-promoter interaction specificity.** *Trends in cell biology* 2014, **24**(11):695-702.
165. Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, Yen CA, Schmitt AD, Espinoza CA, Ren B: **A high-resolution map of the three-dimensional chromatin interactome in human cells.** *Nature* 2013, **503**(7475):290-294.
166. Schoenfelder S, Sexton T, Chakalova L, Cope NF, Horton A, Andrews S, Kurukuti S, Mitchell Ja, Umlauf D, Dimitrova DS *et al*: **Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells.** *Nature genetics* 2010, **42**:53-61.
167. Zentner GE, Tesar PJ, Scacheri PC: **Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions.** *Genome Res* 2011, **21**(8):1273-1283.
168. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J: **A unique chromatin signature uncovers early developmental enhancers in humans.** *Nature* 2011, **470**(7333):279-283.
169. Cui K, Zang C, Roh TY, Schones DE, Childs RW, Peng W, Zhao K: **Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation.** *Cell stem cell* 2009, **4**(1):80-93.
170. Rousseaux S, Khochbin S: **Histone Acylation beyond Acetylation: Terra Incognita in Chromatin Biology.** *Cell journal* 2015, **17**(1):1-6.
171. Bonn S, Zinzen RP, Girardot C, Gustafson EH, Perez-Gonzalez A, Delhomme N, Ghavi-Helm Y, Wilczynski B, Riddell A, Furlong EE: **Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development.** *Nat Genet* 2012, **44**(2):148-156.
172. Yu M, Hon GC, Szulwach KE, Song CX, Zhang L, Kim A, Li X, Dai Q, Shen Y, Park B *et al*: **Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome.** *Cell* 2012, **149**(6):1368-1380.

173. Ding J, Huang X, Shao N, Zhou H, Lee DF, Faiola F, Fidalgo M, Guallar D, Saunders A, Shliaha PV *et al*: **Tex10 Coordinates Epigenetic Control of Super-Enhancer Activity in Pluripotency and Reprogramming**. *Cell stem cell* 2015, **16**(6):653-668.
174. Stroud H, Feng S, Morey Kinney S, Pradhan S, Jacobsen SE: **5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells**. *Genome biology* 2011, **12**(6):R54.
175. Ghandi M, Lee D, Mohammad-Noori M, Beer MA: **Enhanced regulatory sequence prediction using gapped k-mer features**. *PLoS computational biology* 2014, **10**(7):e1003711.
176. Adams CC, Workman JL: **Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative**. *Molecular and cellular biology* 1995, **15**(3):1405-1421.
177. Iwafuchi-Doi M, Zaret KS: **Cell fate control by pioneer transcription factors**. *Development* 2016, **143**(11):1833-1837.
178. Sherwood RI, Hashimoto T, O'Donnell CW, Lewis S, Barkal AA, van Hoff JP, Karun V, Jaakkola T, Gifford DK: **Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape**. *Nature biotechnology* 2014, **32**(2):171-178.
179. Zaret KS, Carroll JS: **Pioneer transcription factors: establishing competence for gene expression**. *Genes & development* 2011, **25**(21):2227-2241.
180. Iwafuchi-Doi M, Donahue G, Kakumanu A, Watts JA, Mahony S, Pugh BF, Lee D, Kaestner KH, Zaret KS: **The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation**. *Molecular cell* 2016, **62**(1):79-91.
181. Svtelisl A, Gevry N, Gaudreau L: **Regulation of gene expression and cellular proliferation by histone H2A.Z**. *Biochemistry and cell biology = Biochimie et biologie cellulaire* 2009, **87**(1):179-188.
182. Li Z, Gadue P, Chen K, Jiao Y, Tuteja G, Schug J, Li W, Kaestner KH: **Foxa2 and H2A.Z mediate nucleosome depletion during embryonic stem cell differentiation**. *Cell* 2012, **151**(7):1608-1616.
183. Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, Kondo S, Nikaido I, Osato N, Saito R, Suzuki H *et al*: **Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs**. *Nature* 2002, **420**(6915):563-573.
184. Cheng J, Kapranov P, Drenkow J, Dike S, Brubaker S, Patel S, Long J, Stern D, Tammana H, Helt G *et al*: **Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution**. *Science* 2005, **308**(5725):1149-1154.
185. Kapranov P, Cheng J, Dike S, Nix DA, Dutttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermuller J, Hofacker IL *et al*: **RNA maps reveal new RNA classes and a possible function for pervasive transcription**. *Science* 2007, **316**(5830):1484-1488.
186. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C *et al*: **The transcriptional landscape of the mammalian genome**. *Science* 2005, **309**(5740):1559-1563.
187. Dinger ME, Amaral PP, Mercer TR, Mattick JS: **Pervasive transcription of the eukaryotic genome: functional indices and conceptual implications**. *Briefings in functional genomics & proteomics* 2009, **8**(6):407-423.
188. Hangauer MJ, Vaughn IW, McManus MT: **Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs**. *PLoS Genet* 2013, **9**(6):e1003569.
189. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T *et al*: **An atlas of active enhancers across human cell types and tissues**. *Nature* 2014, **507**(7493):455-461.
190. Kowalczyk MS, Hughes JR, Garrick D, Lynch MD, Sharpe JA, Sloane-Stanley JA, McGowan SJ, De Gobbi M, Hosseini M, Vernimmen D *et al*: **Intragenic enhancers act as alternative promoters**. *Molecular cell* 2012, **45**(4):447-458.
191. Wang D, Garcia-Bassets I, Benner C, Li W, Su X, Zhou Y, Qiu J, Liu W, Kaikkonen MU, Ohgi KA *et al*: **Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA**. *Nature* 2011, **474**(7351):390-394.
192. Sigova AA, Abraham BJ, Ji X, Molinie B, Hannett NM, Guo YE, Jangi M, Giallourakis CC, Sharp PA, Young RA: **Transcription factor trapping by RNA in gene regulatory elements**. *Science* 2015, **350**(6263):978-981.

193. Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, Merkurjev D, Zhang J, Ohgi K, Song X *et al*: **Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation.** *Nature* 2013, **498**(7455):516-520.
194. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW *et al*: **Histone modifications at human enhancers reflect global cell-type-specific gene expression.** *Nature* 2009, **459**(7243):108-112.
195. Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F *et al*: **ChIP-seq accurately predicts tissue-specific activity of enhancers.** *Nature* 2009, **457**(7231):854-858.
196. Harmston N, Lenhard B: **Chromatin and epigenetic features of long-range gene regulation.** *Nucleic acids research* 2013, **41**(15):7185-7199.
197. Gotea V, Visel A, Westlund JM, Nobrega MA, Pennacchio LA, Ovcharenko I: **Homotypic clusters of transcription factor binding sites are a key component of human promoters and enhancers.** *Genome Res* 2010, **20**(5):565-577.
198. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA: **Master transcription factors and mediator establish super-enhancers at key cell identity genes.** *Cell* 2013, **153**(2):307-319.
199. Arner E, Daub CO, Vitting-Seerup K, Andersson R, Lilje B, Drablos F, Lennartsson A, Ronnerblad M, Hrydziuszko O, Vitezic M *et al*: **Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells.** *Science* 2015, **347**(6225):1010-1014.
200. Consortium EP: **An integrated encyclopedia of DNA elements in the human genome.** *Nature* 2012, **489**(7414):57-74.
201. Alfoldi J, Lindblad-Toh K: **Comparative genomics as a tool to understand evolution and disease.** *Genome Res* 2013, **23**(7):1063-1068.
202. Tagle DA, Koop BF, Goodman M, Slightom JL, Hess DL, Jones RT: **Embryonic epsilon and gamma globin genes of a prosimian primate (*Galago crassicaudatus*). Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints.** *Journal of molecular biology* 1988, **203**(2):439-455.
203. Nobrega MA, Ovcharenko I, Afzal V, Rubin EM: **Scanning human gene deserts for long-range enhancers.** *Science* 2003, **302**(5644):413.
204. Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD *et al*: **In vivo enhancer analysis of human conserved non-coding sequences.** *Nature* 2006, **444**(7118):499-502.
205. Plessy C, Dickmeis T, Chalmel F, Strahle U: **Enhancer sequence conservation between vertebrates is favoured in developmental regulator genes.** *Trends in genetics : TIG* 2005, **21**(4):207-210.
206. Woolfe A, Elgar G: **Organization of conserved elements near key developmental regulators in vertebrate genomes.** *Advances in genetics* 2008, **61**:307-338.
207. Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, Haussler D: **Ultraconserved elements in the human genome.** *Science* 2004, **304**(5675):1321-1325.
208. Ahituv N, Zhu Y, Visel A, Holt A, Afzal V, Pennacchio LA, Rubin EM: **Deletion of ultraconserved elements yields viable mice.** *PLoS biology* 2007, **5**(9):e234.
209. Friedli M, Barde I, Arcangeli M, Verp S, Quazzola A, Zakany J, Lin-Marq N, Robyr D, Attanasio C, Spitz F *et al*: **A systematic enhancer screen using lentivector transgenesis identifies conserved and non-conserved functional elements at the *Olig1* and *Olig2* locus.** *PLoS one* 2010, **5**(12):e15741.
210. Cotney J, Leng J, Yin J, Reilly SK, DeMare LE, Emera D, Ayoub AE, Rakic P, Noonan JP: **The evolution of lineage-specific regulatory activities in the human embryonic limb.** *Cell* 2013, **154**(1):185-196.
211. Tamura K, Battistuzzi FU, Billing-Ross P, Murillo O, Filipinski A, Kumar S: **Estimating divergence times in large molecular phylogenies.** *Proc Natl Acad Sci U S A* 2012, **109**(47):19333-19338.
212. Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, Park TJ, Deaville R, Erichsen JT, Jasinska AJ *et al*: **Enhancer evolution across 20 mammalian species.** *Cell* 2015, **160**(3):554-566.
213. Li Y, Rivera CM, Ishii H, Jin F, Selvaraj S, Lee AY, Dixon JR, Ren B: **CRISPR reveals a distal super-enhancer required for *Sox2* expression in mouse embryonic stem cells.** *PLoS one* 2014, **9**(12):e114485.

214. Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, Jaenisch R: **One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering.** *Cell* 2013, **153**(4):910-918.
215. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM: **RNA-guided human genome engineering via Cas9.** *Science* 2013, **339**(6121):823-826.
216. Findlay GM, Boyle EA, Hause RJ, Klein JC, Shendure J: **Saturation editing of genomic regions by multiplex homology-directed repair.** *Nature* 2014, **513**(7516):120-123.
217. Ulirsch JC, Nandakumar SK, Wang L, Giani FC, Zhang X, Rogov P, Melnikov A, McDonel P, Do R, Mikkelsen TS *et al*: **Systematic Functional Dissection of Common Genetic Variation Affecting Red Blood Cell Traits.** *Cell* 2016, **165**(6):1530-1545.
218. Patwardhan RP, Hiatt JB, Witten DM, Kim MJ, Smith RP, May D, Lee C, Andrie JM, Lee SI, Cooper GM *et al*: **Massively parallel functional dissection of mammalian enhancers in vivo.** *Nature biotechnology* 2012, **30**(3):265-270.
219. Murtha M, Tokcaer-Keskin Z, Tang Z, Strino F, Chen X, Wang Y, Xi X, Basilico C, Brown S, Bonneau R *et al*: **FIREWACH: high-throughput functional detection of transcriptional regulatory modules in mammalian cells.** *Nature methods* 2014, **11**(5):559-565.
220. Arnold CD, Gerlach D, Stelzer C, Boryn LM, Rath M, Stark A: **Genome-wide quantitative enhancer activity maps identified by STARR-seq.** *Science* 2013, **339**(6123):1074-1077.
221. Mendenhall EM, Williamson KE, Reyon D, Zou JY, Ram O, Joung JK, Bernstein BE: **Locus-specific editing of histone modifications at endogenous enhancers.** *Nature biotechnology* 2013, **31**(12):1133-1136.
222. Kearns NA, Pham H, Tabak B, Genga RM, Silverstein NJ, Garber M, Maehr R: **Functional annotation of native enhancers with a Cas9-histone demethylase fusion.** *Nature methods* 2015, **12**(5):401-403.
223. Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE, Gersbach CA: **Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers.** *Nature biotechnology* 2015, **33**(5):510-517.
224. Gallo SM, Gerrard DT, Miner D, Simich M, Des Soye B, Bergman CM, Halfon MS: **REDfly v3.0: toward a comprehensive database of transcriptional regulatory elements in Drosophila.** *Nucleic acids research* 2011, **39**(Database issue):D118-123.
225. Visel A, Minovitsky S, Dubchak I, Pennacchio LA: **VISTA Enhancer Browser--a database of tissue-specific human enhancers.** *Nucleic acids research* 2007, **35**(Database issue):D88-92.
226. Brunet TD, Doolittle WF: **Getting "function" right.** *Proc Natl Acad Sci U S A* 2014, **111**(33):E3365.
227. Kellis M, Wold B, Snyder MP, Bernstein BE, Kundaje A, Marinov GK, Ward LD, Birney E, Crawford GE, Dekker J *et al*: **Reply to Brunet and Doolittle: Both selected effect and causal role elements can influence human biology and disease.** *Proc Natl Acad Sci U S A* 2014, **111**(33):E3366.
228. Kellis M, Wold B, Snyder MP, Bernstein BE, Kundaje A, Marinov GK, Ward LD, Birney E, Crawford GE, Dekker J *et al*: **Defining functional DNA elements in the human genome.** *Proc Natl Acad Sci U S A* 2014, **111**(17):6131-6138.
229. Sambrook J, Russell DW: **Molecular cloning : a laboratory manual**, 3rd edn. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press; 2001.
230. Fried M, Crothers DM: **Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis.** *Nucleic Acids Res* 1981, **9**(23):6505-6525.
231. Elkahoul AG, Bittner M, Hoskins K, Gemmill R, Meltzer PS: **Molecular cytogenetic characterization and physical mapping of 12q13-15 amplification in human cancers.** *Genes, chromosomes & cancer* 1996, **17**(4):205-214.
232. Splinter E, Grosveld F, de Laat W: **3C technology: analyzing the spatial organization of genomic loci in vivo.** *Methods in enzymology* 2004, **375**:493-507.
233. Ruijter JM, Ramakers C, Hoogaars WM, Karlen Y, Bakker O, van den Hoff MJ, Moorman AF: **Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data.** *Nucleic acids research* 2009, **37**(6):e45.
234. Hagege H, Klous P, Braem C, Splinter E, Dekker J, Cathala G, de Laat W, Forne T: **Quantitative analysis of chromosome conformation capture assays (3C-qPCR).** *Nature protocols* 2007, **2**(7):1722-1733.
235. Wolfe MS: **Structure, mechanism and inhibition of gamma-secretase and presenilin-like proteases.** *Biological chemistry* 2010, **391**(8):839-847.

236. Shepard J, Reick M, Olson S, Graveley BR: **Characterization of U2AF(6), a splicing factor related to U2AF(35)**. *Molecular and cellular biology* 2002, **22**(1):221-230.
237. Heyd F, Carmo-Fonseca M, Moroy T: **Differential isoform expression and interaction with the P32 regulatory protein controls the subcellular localization of the splicing factor U2AF26**. *The Journal of biological chemistry* 2008, **283**(28):19636-19645.
238. Chen F, Ji C, Dou T, Zheng N, Qiu R, Peng J, Fang W, Feng C, Xie Y, Mao Y: **Cloning and characterization of a novel splice variant of human U2AF1L3 gene**. *DNA sequence : the journal of DNA sequencing and mapping* 2006, **17**(4):282-286.
239. Celniker SE, Dillon LA, Gerstein MB, Gunsalus KC, Henikoff S, Karpen GH, Kellis M, Lai EC, Lieb JD, MacAlpine DM *et al*: **Unlocking the secrets of the genome**. *Nature* 2009, **459**(7249):927-930.
240. Sabo PJ, Kuehn MS, Thurman R, Johnson BE, Johnson EM, Cao H, Yu M, Rosenzweig E, Goldy J, Haydock A *et al*: **Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays**. *Nature methods* 2006, **3**(7):511-518.
241. Crawford GE, Davis S, Scacheri PC, Renaud G, Halawi MJ, Erdos MR, Green R, Meltzer PS, Wolfsberg TG, Collins FS: **DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays**. *Nature methods* 2006, **3**(7):503-509.
242. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE: **High-resolution mapping and characterization of open chromatin across the genome**. *Cell* 2008, **132**(2):311-322.
243. Crawford GE, Holt IE, Whittle J, Webb BD, Tai D, Davis S, Margulies EH, Chen Y, Bernat JA, Ginsburg D *et al*: **Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS)**. *Genome research* 2006, **16**(1):123-131.
244. Giresi PG, Kim J, McDaniel RM, Iyer VR, Lieb JD: **FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin**. *Genome research* 2007, **17**(6):877-885.
245. Giresi PG, Lieb JD: **Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements)**. *Methods* 2009, **48**(3):233-239.
246. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K: **High-resolution profiling of histone methylations in the human genome**. *Cell* 2007, **129**(4):823-837.
247. Klochkov D, Rincon-Arano H, Ioudinkova ES, Valadez-Graham V, Gavrilov A, Recillas-Targa F, Razin SV: **A CTCF-dependent silencer located in the differentially methylated area may regulate expression of a housekeeping gene overlapping a tissue-specific gene domain**. *Molecular and cellular biology* 2006, **26**(5):1589-1597.
248. Ruda VM, Akopov SB, Trubetskiy DO, Manuylov NL, Vetchinova AS, Zavalova LL, Nikolaev LG, Sverdlov ED: **Tissue specificity of enhancer and promoter activities of a HERV-K(HML-2) LTR**. *Virus research* 2004, **104**(1):11-16.
249. Abrarova ND, Stukacheva EA, Pleshkan VV, Vinogradova TV, Sverdlov ED: **[Functional analysis of the HERV-K LTR residing in the KIAA1245/NBPF subfamily]**. *Molekuliarnaia biologii* 2010, **44**(4):627-634.
250. Matz M, Shagin D, Bogdanova E, Britanova O, Lukyanov S, Diatchenko L, Chenchik A: **Amplification of cDNA ends based on template-switching effect and step-out PCR**. *Nucleic acids research* 1999, **27**(6):1558-1560.
251. Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CA, Taylor MS, Engstrom PG, Frith MC *et al*: **Genome-wide analysis of mammalian promoter architecture and evolution**. *Nat Genet* 2006, **38**(6):626-635.
252. Wang R, Zhang YW, Sun P, Liu R, Zhang X, Zhang X, Xia K, Xia J, Xu H, Zhang Z: **Transcriptional regulation of PEN-2, a key component of the gamma-secretase complex, by CREB**. *Molecular and cellular biology* 2006, **26**(4):1347-1354.
253. Shagin DA, Barsova EV, Yanushevich YG, Fradkov AF, Lukyanov KA, Labas YA, Semenova TN, Ugalde JA, Meyers A, Nunez JM *et al*: **GFP-like proteins as ubiquitous metazoan superfamily: evolution of functional features and structural complexity**. *Molecular biology and evolution* 2004, **21**(5):841-850.
254. Shcherbo D, Merzlyak EM, Chepurnykh TV, Fradkov AF, Ermakova GV, Solovieva EA, Lukyanov KA, Bogdanova EA, Zarskiy AG, Lukyanov S *et al*: **Bright far-red fluorescent protein for whole-body imaging**. *Nature methods* 2007, **4**(9):741-746.
255. Mollet I, Barbosa-Morais NL, Andrade J, Carmo-Fonseca M: **Diversity of human U2AF splicing factors**. *The FEBS journal* 2006, **273**(21):4807-4816.

256. Crooks GE, Hon G, Chandonia JM, Brenner SE: **WebLogo: a sequence logo generator**. *Genome Res* 2004, **14**(6):1188-1190.
257. Chang WT, Huang AM: **Alpha-Pal/NRF-1 regulates the promoter of the human integrin-associated protein/CD47 gene**. *The Journal of biological chemistry* 2004, **279**(15):14542-14550.
258. Lin JM, Collins PJ, Trinklein ND, Fu Y, Xi H, Myers RM, Weng Z: **Transcription factor binding and modified histones in human bidirectional promoters**. *Genome Res* 2007, **17**(6):818-827.
259. Collins PJ, Kobayashi Y, Nguyen L, Trinklein ND, Myers RM: **The ets-related transcription factor GABP directs bidirectional transcription**. *PLoS Genet* 2007, **3**(11):e208.
260. Liu B, Chen J, Shen B: **Genome-wide analysis of the transcription factor binding preference of human bi-directional promoters and functional annotation of related gene pairs**. *BMC systems biology* 2011, **5 Suppl 1**:S2.
261. Rosenbloom KR, Dreszer TR, Pheasant M, Barber GP, Meyer LR, Pohl A, Raney BJ, Wang T, Hinrichs AS, Zweig AS *et al*: **ENCODE whole-genome data in the UCSC Genome Browser**. *Nucleic acids research* 2010, **38**(Database issue):D620-625.
262. Euskirchen GM, Rozowsky JS, Wei CL, Lee WH, Zhang ZD, Hartman S, Emanuelsson O, Stolc V, Weissman S, Gerstein MB *et al*: **Mapping of transcription factor binding regions in mammalian cells by ChIP: comparison of array- and sequencing-based technologies**. *Genome Res* 2007, **17**(6):898-909.
263. Anno YN, Myslinski E, Ngondo-Mbongo RP, Krol A, Poch O, Lecompte O, Carbon P: **Genome-wide evidence for an essential role of the human Staf/ZNF143 transcription factor in bidirectional transcription**. *Nucleic acids research* 2011, **39**(8):3116-3127.
264. Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, Lander ES, Kellis M: **Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals**. *Nature* 2005, **434**(7031):338-345.
265. Gurudatta BV, Corces VG: **Chromatin insulators: lessons from the fly**. *Brief Funct Genomic Proteomic* 2009, **8**(4):276-282.
266. Ernst J, Kellis M: **Discovery and characterization of chromatin states for systematic annotation of the human genome**. In: *Nat Biotechnol*. vol. 28. United States; 2010: 817-825.
267. Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M *et al*: **Mapping and analysis of chromatin state dynamics in nine human cell types**. In: *Nature*. vol. 473. England; 2011: 43-49.
268. Consortium EP, Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Harrow J *et al*: **An integrated encyclopedia of DNA elements in the human genome**. *Nature*, **489**(7414):57-74.
269. Recillas-Targa F, Bell AC, Felsenfeld G: **Positional enhancer-blocking activity of the chicken beta-globin insulator in transiently transfected cells**. *Proc Natl Acad Sci U S A* 1999, **96**(25):14354-14359.
270. Mueller-Sturm HP, Sogo JM, Schaffner W: **An enhancer stimulates transcription in trans when attached to the promoter via a protein bridge**. *Cell* 1989, **58**(4):767-777.
271. Dunaway M, Droge P: **Transactivation of the Xenopus rRNA gene promoter by its enhancer**. *Nature* 1989, **341**(6243):657-659.
272. Kuz'mich AI, Kopantsev EP, Vinogradova TV, Sverdlov ED: **[Comparative analysis of activity of different promoters for NIS gene expression in melanoma cells]**. *Molekuliarnaia biologiiia* 2014, **48**(1):142-152.
273. Farr A, Roman A: **A pitfall of using a second plasmid to determine transfection efficiency**. *Nucleic acids research* 1992, **20**(4):920.
274. Adam GI, Miller SJ, Ulleras E, Franklin GC: **Cell-type-specific modulation of PDGF-B regulatory elements via viral enhancer competition: a caveat for the use of reference plasmids in transient transfection assays**. *Gene* 1996, **178**(1-2):25-29.
275. Cohen RN, van der Aa MA, Macaraeg N, Lee AP, Szoka FC, Jr.: **Quantification of plasmid DNA copies in the nucleus after lipoplex and polyplex transfection**. *Journal of controlled release : official journal of the Controlled Release Society* 2009, **135**(2):166-174.

## **БЛАГОДАРНОСТИ**

Выражаю искреннюю признательность всем сотрудникам лаборатории структуры и функций генов человека ИБХ РАН за помощь в работе и создание благоприятного трудового климата.

Особую благодарность за помощь и наставничество хочу выразить Акопову Сергею Борисовичу, Николаеву Льву Григорьевичу, Дидычу Дмитрию Александровичу.