

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

Котова Елена Сергеевна

**Идентификация и анализ активности СТСФ-зависимых  
регуляторных элементов**

Специальность 03.01.03 – молекулярная биология

Диссертация на соискание ученой степени  
кандидата биологических наук

Научный руководитель:  
Акопов Сергей Борисович  
доктор биологических наук

Москва - 2016

ОГЛАВЛЕНИЕ	
ВВЕДЕНИЕ.....	4
ОСНОВНАЯ ЧАСТЬ.....	9
1. Обзор литературы .....	9
1.1. CTCF-многофункциональный высококонсервативный транскрипционный фактор.....	9
1.2. Функциональные свойства CTCF .....	10
1.2.1. Сайт связывания CTCF.....	10
1.2.2. Взаимодействие CTCF с ДНК.....	11
1.2.3. Влияние метилирования CpG ДНК на связывание CTCF .....	13
1.2.4. Взаимодействие CTCF с белками.....	14
1.3. Роль транскрипционного фактора CTCF в регуляции ДНК-зависимых процессов .....	15
1.3.1. Гипотеза функциональных доменов хроматина. Инсуляторы... ..	16
1.3.2. Блокирование энхансера .....	20
1.3.3. Инсулятор как пограничный элемент .....	22
1.3.4. CTCF как регулятор транскрипции .....	23
1.3.5. Участие белка CTCF в импринтинге генетической информации .....	25
1.3.6. CTCF и инактивация X-хромосомы .....	26
1.3.7. Котранскрипционная регуляция альтернативного сплайсинга.. ..	28
2. Материалы и методы .....	29
2.1. Материалы .....	29
2.2. Методы .....	35
3. Результаты и обсуждение.....	56
3.1. Выявление фрагментов $\alpha$ -глобинового локуса кур, способных связывать белок CTCF, методом двумерного EMSA .....	56
3.1.1. Получение и функциональный анализ фракции, обогащенной куриным белком CTCF. ....	57

3.1.2. Получение библиотеки CTCF-связывающих последовательностей $\alpha$ -глобинового локуса. Их идентификация и картирование.....	67
3.1.3. Выявление среди отобранных в результате двумерного EMSA фрагментов ДНК, последовательностей, способных связывать CTCF <i>in vivo</i> .....	78
3.2. Полногеномное выявление связывающих белок CTCF нуклеотидных последовательностей методом ChIP-seq .....	84
3.2.1. Снижение занятости белком CTCF-связывающих областей при индукции дифференцировки клеток HD3.....	87
3.2.2. Расположение сайтов связывания CTCF в $\alpha$ -глобиновом локусе кур .....	88
3.2.3. Оценка тканеспецифичности связывания CTCF .....	89
3.2.4. Сравнение результатов ChIP-seq анализа для взрослых и эмбриональных куриных клеток.....	90
3.2.5. Сравнение результатов поиска участков генома, связывающих CTCF <i>in vitro</i> и <i>in vivo</i> .....	92
ЗАКЛЮЧЕНИЕ И ВЫВОДЫ .....	98
Заключение .....	98
Выводы .....	98
СПИСОК СОКРАЩЕНИЙ.....	100
Список литературы .....	102
ПРИЛОЖЕНИЯ.....	114

## **ВВЕДЕНИЕ**

Фенотипическое разнообразие живого во многом определяется наследственной информацией, которая хранится в виде нуклеотидных последовательностей геномов.

Одной из важнейших задач молекулярной биологии является изучение механизмов, обеспечивающих реализацию наследственной информации, в том числе регуляции экспрессии генов. Регуляция экспрессии генов позволяет живым организмам поддерживать гомеостаз в условиях меняющейся окружающей среды, играет важнейшую роль в процессе клеточной дифференцировки. Нарушение регуляции экспрессии генов приводит к развитию различных заболеваний, а иногда делает организм нежизнеспособным уже на ранних стадиях развития.

Ключевую роль в регуляции экспрессии генов играют функциональные элементы геномов: промоторы, энхансеры, терминаторы транскрипции, участки начала репликации, участки связывания ДНК с транскрипционными факторами, и др. Определение полных нуклеотидных последовательностей геномов не предоставляет исчерпывающей информации об их расположении и функционировании. Таким образом, необходимо картирование и анализ активности областей геномов, ответственных за регуляцию экспрессии генов.

Белок CTCF – многофункциональный транскрипционный фактор позвоночных. К настоящему времени накоплен большой объем данных, свидетельствующих о его важной роли в различных ДНК-зависимых процессах. В некоторых случаях CTCF играет роль активатора транскрипции, в других – репрессора. Также белок CTCF обеспечивает активность инсуляторов, участвует в инактивации X-хромосомы, импринтинге генетической информации и, по-видимому, регулирует процесс сплайсинга РНК.

Особый интерес представляет участие белка CTCF в образовании доменной структуры геномов позвоночных. Образование функциональных хроматиновых доменов играет важную роль в регуляции экспрессии генов позвоночных. Функциональный хроматиновый домен – это протяженный геномный локус, хроматин которого независимо от хроматина других доменов может переходить в

открытую (транскрипционно активную) или закрытую (транскрипционно неактивную) конформацию. В одном хроматиновом домене, как правило, расположены гены, продукты которых функционально связаны, то есть, задействованы в общем биологическом процессе. Следствием колокализации этих генов являются общие механизмы регуляции их транскрипции, действующие на уровне хроматиновых доменов. В соответствии с гипотезой хроматиновых доменов, концы хроматинового домена сближены и он представляет собой петлевую структуру. Сайты связывания белка CTCF входят в состав инсуляторов, ограничивающих хроматиновые домены. Белок CTCF участвует в образовании хроматиновых петель, при этом, по-видимому, сближаются концы хроматинового домена, а также происходит взаимодействие регуляторных элементов внутри него.

Для выявления роли CTCF в образовании и функционировании хроматиновых доменов важно изучение расположения и активности сайтов связывания CTCF в отдельных протяженных геномных локусах. По сравнению с полногеномными исследованиями оно позволяет более детально исследовать взаимодействие белка CTCF с его сайтами, выявить возможную взаимосвязь между активностью сайтов связывания CTCF и регуляцией экспрессии генов в данной области генома. Выявление сайтов связывания CTCF, расположенных в одном локусе и активных в определенном типе клеток позволяет в дальнейшем выявлять CTCF-связывающие участки генома, задействованные в общих механизмах регуляции экспрессии генов. В этом преимущество изучения активности сайтов связывания CTCF протяженных геномных локусов перед изучением активности отдельных CTCF-связывающих фрагментов генома.

Примером одного из детально исследованных хроматиновых доменов является домен  $\alpha$ -глобиновых генов кур. Множество работ посвящено выявлению и анализу активности различных функциональных элементов этого домена.

В  $\alpha$ -глобиновом локусе кур гены  $\alpha$ -глобинов *HBZ*, *HBAD* и *HBAA* активно транскрибируются только в клетках эритроидного ряда дифференцировки. Экспрессия *HBZ* происходит в клетках полутора-пятидневных эмбрионов, а экспрессия *HBAD* и, в основном, *HBAA* в курином эмбрионе начинается с 4-5 дневного возраста. Кластер  $\alpha$ -глобиновых генов фланкирован генами,

конститутивно экспрессирующимися в разных типах тканей – *ggPRX* и *TMEM8*. При этом хроматин локуса находится в открытом состоянии как в клетках эритроидного пути дифференцировки, так и в других типах клеток. Поскольку не было выявлено пограничных элементов этого локуса, его причисляют к доменам открытого типа.

Однако, в эритроидных клетках отмечены некоторые изменения структуры хроматина в области  $\alpha$ -глобиновых генов, такие как повышение уровня ацетилирования гистонов по ряду специфических позиций. По-видимому, этот процесс контролируется путем низкоуровневой транскрипции всего домена. В области MRE в направлении генов  $\alpha$ -глобинов начинается синтез длинного транскрипта, размером около 33 т.п.о. Возможно, домен  $\alpha$ -глобиновых генов все же имеет пограничные элементы, между которыми происходят эти изменения структуры хроматина.

В состав пограничных элементов хроматиновых доменов, как правило, входят сайты связывания CTCF. Таким образом, картирование CTCF-связывающих последовательностей  $\alpha$ -глобинового локуса может способствовать выявлению его пограничных элементов.

Помимо CTCF-связывающих последовательностей, которые входят в состав пограничных элементов хроматиновых доменов, существуют CTCF-связывающие последовательности, способствующие сближению регуляторных элементов внутри локусов, влияющие на альтернативный сплайсинг, участвующие в активации и подавлении транскрипции. Поэтому для выявления механизмов, регулирующих экспрессию генов этого локуса, необходимы картирование, анализ активности и последующее изучение функций CTCF-связывающих последовательностей, расположенных в одном домене.

Картирование сайтов связывания CTCF и исследование их активности в  $\alpha$ -глобиновом локусе кур важно для изучения роли, которую играет белок CTCF в различных ДНК-зависимых процессах, а также для изучения регуляции функционирования геномов позвоночных на уровне хроматиновых доменов.

Для поиска сайтов связывания CTCF ранее использовались различные подходы. В ранних исследованиях сайты связывания CTCF были найдены

случайным образом в составе некоторых регуляторных последовательностей геномов позвоночных [113; 7; 94]. В  $\alpha$ -глобиновом локусе кур для некоторых сайтов гиперчувствительности к ДНКазе [173] и GC-богатых участков ДНК [92] показана способность связывать CTCF *in vitro*. Эти исследования проводились методом сдвига электрофоретической подвижности (EMSA-Electrophoretic Mobility Shift Assay). Для некоторых, из найденных таким образом, сайтов связывания CTCF методом иммунопреципитации хроматина было показано связывание с белком CTCF *in vivo* – в составе хроматина живых клеток [173; 92].

Также в результате полногеномного выявления фрагментов ДНК, связывающих CTCF, был обнаружен ряд участков  $\alpha$ -глобинового локуса, взаимодействующих с CTCF *in vivo* в эритроблестах пятидневных и десятидневных эмбрионов кур.

Однако в литературе отсутствуют данные об исчерпывающем выявлении CTCF-связывающих фрагментов в  $\alpha$ -глобиновом локусе кур, активных во взрослых эритроидных и неэритроидных клетках.

Таким образом, **целью** диссертационной работы стало выявление в  $\alpha$ -глобиновом локусе кур CTCF-связывающих участков, сравнение связывания CTCF в  $\alpha$ -глобиновом локусе в клетках эритроидного и лимфоидного происхождения.

В ходе работы были поставлены следующие **задачи**:

1. Получение фракции, обогащенной куриным белком CTCF, способным специфично взаимодействовать с CTCF-связывающими последовательностями, для последующего ее использования в ходе двумерного EMSA (Electrophoretic Mobility Shift Assay).

2. Выявление методом двумерного EMSA рестриктных фрагментов  $\alpha$ -глобинового локуса кур, способных связывать транскрипционный фактор CTCF *in vitro*.

3. Анализ методом дополнительного сдвига электрофоретической подвижности способности обнаруженных фрагментов ДНК связывать белок CTCF *in vitro*.

4. Анализ связывания с обнаруженными фрагментами ДНК белка CTCF *in vivo* путем иммунопреципитации хроматина, с использованием трех типов клеток:

клетки линии HD3 (эритроидного происхождения), до и после индукции дифференцировки, клетки линии DT40 (лимфоидного происхождения).

5. Выявление методом ChIP-seq нуклеотидных последовательностей, связывающих белок CTCF *in vivo* в хроматине клеток линии HD3, индуцированных к дифференцировке и интактных, клеток линии DT40.

6. Сравнение распределения сайтов связывания CTCF в  $\alpha$ -глобиновом локусе кур и их связывания с белком CTCF в клетках линии HD3, индуцированных к дифференцировке и интактных, клеток линии DT40.

7. Сравнение полученных данных с данными, опубликованными ранее другими авторами.

## ОСНОВНАЯ ЧАСТЬ

### 1. Обзор литературы

#### *1.1. CTCF-многофункциональный высококонсервативный транскрипционный фактор*

CTCF (CCCTC binding factor) – эволюционно консервативный транскрипционный фактор позвоночных. Он связывается с различными функциональными элементами генома и выполняет разнообразные регуляторные функции [131; 142; 72]

Транскрипционный фактор CTCF был впервые идентифицирован как белок, который узнает последовательность трех прямых повторов CCCTC в регуляторной области гена *MYC* кур [113; 114]. Одновременно был описан белок NeP1, связывающийся с F1-элементом сайленсера гена лизоцима кур [7]. Как оказалось, NeP1 и CTCF представляют собой один и тот же белок [16].

Первичная структура CTCF высококонсервативна; ортологи CTCF были выявлены у различных позвоночных: человека, собаки, мыши, кур, лягушек и др. [90; 157; 12]. Аминокислотные последовательности белка CTCF человека и кур идентичны на 93% [43].

Полипептидная цепь CTCF человека состоит из 727 аминокислотных остатков. Центральный ДНК-связывающий домен содержит одиннадцать цинковых пальцев, фланкированных богатыми лизином и аргинином положительно заряженными N- и C-концевыми доменами, которые содержат 267 и 150 а.о. [132]. После ДНК-связывающего домена располагается глицин-богатый мотив, который характерен для АТР- и GTP-связывающих белков. Ближе к C-концу полипептида находится сигнал ядерной локализации, за ним располагаются сайты фосфорилирования [90; 43].

Разнообразие выполняемых белком CTCF функций предполагает существование процесса регуляции активности самого CTCF, осуществляющегося в том числе и посредством его посттрансляционных модификаций. Из модификаций следует отметить фосфорилирование киназой 2 [90; 43], ADP-рибозилирование [88], сумоилирование [118; 87; 181].

Эволюционно CTCF появляется у билатеральных многоклеточных животных и отсутствует у растений и простейших [68]. CTCF экспрессируется конститутивно и необходим для функционирования многих типов клеток позвоночных в многоклеточном организме. Для жизни клеток млекопитающих в культуре наличие CTCF, по всей видимости, не является обязательным [131]. Тем не менее, на важное значение белка CTCF для развития позвоночных указывает тот факт, что мышинные эмбрионы, гомозиготные по поврежденному гену белка CTCF, гибнут еще до имплантации [166; 66]. Повреждение гена, кодирующего CTCF в ооцитах мышей, препятствовало нормальному развитию бластоцисты после оплодотворения [39].

## ***1.2. Функциональные свойства CTCF***

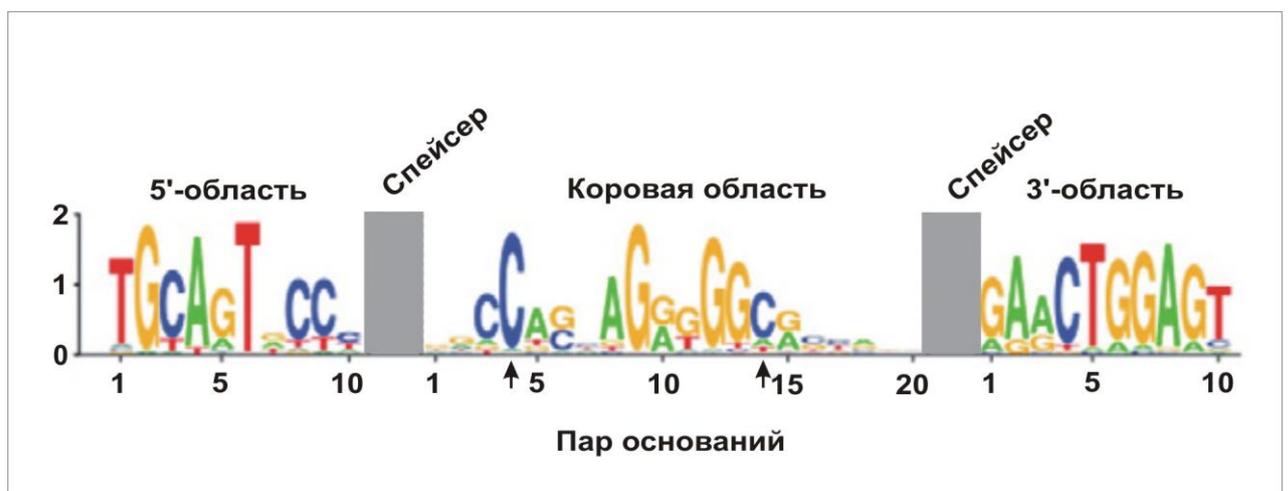
### *1.2.1. Сайт связывания CTCF*

С развитием технологий ChIP-chip (Chromatin Immunoprecipitation on chip) и ChIP-seq (Chromatin Immunoprecipitation-Sequencing) [9; 86; 19] был получен ряд консенсусных последовательностей ДНК, с которыми предпочтительно взаимодействует CTCF [9; 19; 186]. Оказалось, что не только аминокислотная последовательность белка CTCF, но и нуклеотидная последовательность его сайта связывания высококонсервативны для различных видов позвоночных.

Позднее процедура ChIP-seq была дополнена обработкой экзонуклеазой (ChIP-exo) [155], что позволило более точно локализовать сайты связывания CTCF в геноме. Было показано, что сайт связывания CTCF можно подразделить на четыре блока, каждый со своей консенсусной последовательностью. Около половины сайтов связывания CTCF содержало лишь два центральных блока.

Различные сайты связывания CTCF содержали либо все четыре блока, либо комбинацию двух или трех блоков, либо только один блок. Дальнейшие исследования привели к понятию коровой области связывания CTCF [126], лежащей в основе подавляющего большинства сайтов связывания (рисунок 1).

Среди CTCF-связывающих последовательностей всегда присутствует небольшое число тех, для которых нельзя выявить какой-либо консенсус [155; 126]. Предполагается, что эти последовательности связывают CTCF не напрямую, а через промежуточные белки [126]. Экспериментально обнаруженные CTCF-связывающие последовательности сведены в базу данных CTCF Binding Site Database [200].



**Рисунок 1. Общая нуклеотидная структура сайта связывания CTCF [126].** Стрелками обозначены потенциальные сайты метилирования CpG. Размер спейсера, отделяющего 5'-область от коровой области, 5-6 н.о., отделяющего коровую область от 3'-области – 6-8 н.о.

### 1.2.2. Взаимодействие CTCF с ДНК

Методом торможения в геле исследовали взаимодействие фрагментов домена цинковых пальцев CTCF с некоторыми известными CTCF-связывающими последовательностями, в частности сайтом из  $\beta$ -глобинового локуса кур [11], сайтами локуса *Igf2/H19* мышей [36; 154] и сайтом APV $\beta$ , расположенном в промоторе гена предшественника  $\beta$ -амилоида человека [176].

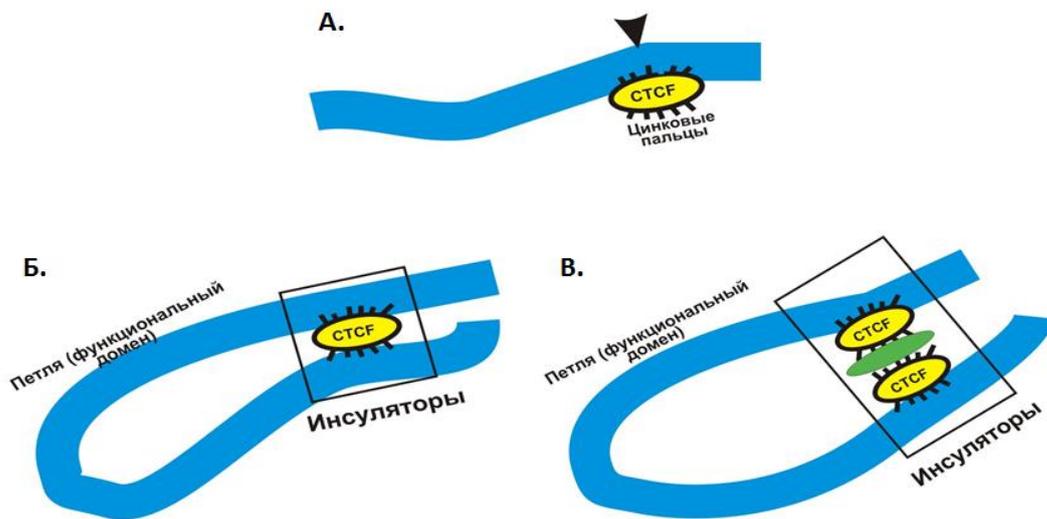
Поскольку цинковые пальцы – характерная ДНК-связывающая структура многих ядерных белков [93], предположили, что именно содержащий их домен

отвечает за взаимодействие белка CTCF с ДНК. Было показано, что 4 цинковых пальца (с 4 по 7) являются минимальным набором, необходимым для специфического взаимодействия CTCF с его сайтом связывания *in vitro*. При дальнейшем уменьшении количества цинковых пальцев ДНК-связывающая способность фрагмента CTCF резко уменьшалась [36; 154].

CTCF может использовать для связывания с ДНК различные комбинации цинковых пальцев. В частности, связывание с регуляторным сайтом гена *MYC* кур осуществляют цинковые пальцы со второго по седьмой, а с сайтом, расположенным рядом с промотором P2 гена *MYC* человека, CTCF связывается при помощи цинковых пальцев с третьего по одиннадцатый [43]. Для связывания с фрагментом F1 сайленсера гена лизоцима кур *in vitro* необходимы цинковые пальцы с 5 по 8 [16].

Позднее эти данные были подтверждены и расширены с использованием массивированного секвенирования [126]. Показано, что восьмой цинковый палец не обязателен для специфического узнавания CTCF его сайта связывания, но стабилизирует комплекс CTCF-ДНК за счет неспецифического взаимодействия. Цинковые пальцы 9-11 отвечают за взаимодействие с небольшим участком CTCF-связывающей последовательности, отделенным от корового участка спейсером. Этот фрагмент назвали U-элементом (upstream element), он соответствует блоку 1, идентифицированному в работе Rhee и соавторов [155]. Интересно, что при удалении цинковых пальцев 8-11 взаимодействие CTCF с сайтом связывания, содержащим U-элемент, полностью исчезает [126]. Возможно, что в том случае, когда коровая последовательность далека от консенсусной, ДНК-белковый комплекс стабилен только при дополнительном взаимодействии CTCF с U-элементом. Цинковый палец 3 вовлечен во взаимодействие с ДНК в отсутствие U-элемента, а цинковые пальцы 1 и 2, по-видимому, не участвуют в специфическом взаимодействии с ДНК, но отвечают за общую стабилизацию ДНК-белкового комплекса.

Предложены гипотетические механизмы образования петли (рисунок 2), посредством которых CTCF может участвовать в образовании доменов хроматина и функционирует в качестве белкового компонента инсуляторов [119; 63].



**Рисунок 2. Гипотетическая модель образования петли при помощи CTCF. А – связывание CTCF со своим сайтом (обозначен стрелкой) за счет части цинковых пальцев; Б – с помощью основной части цинковых пальцев белок связывается со своим сайтом, остальные цинковые пальцы остаются свободными для образования петли (при связывании CTCF ДНК изгибается, что позволяет свободным цинковым пальцам замкнуть петлю [119]); В – образование петли вследствие связывания двух молекул CTCF с двумя сайтами, расположенными на границах функционального домена (возможно, с участием или без него промежуточного белка (зеленый овал)); свободные цинковые пальцы могут принимать участие в белок-белковых взаимодействиях [119]; заключенные в квадрат области соответствуют инсуляторам или пограничным элементам домена).**

### 1.2.3. Влияние метилирования CpG ДНК на связывание CTCF

Было показано, что метилирование CpG в сайтах связывания CTCF приводит к подавлению их связывания с белком CTCF *in vitro* [10; 64; 44; 154]. Кроме того, CTCF не связывается с метилированной областью ICR локуса отцовского аллеля *in vivo* [81; 138], а связывание CTCF с ICR локуса *Igf2/H19* материнской хромосомы препятствует ее метилированию в ходе развития организма [138]. Wang и соавторы [180] исследовали дифференциальное метилирование *in vivo* CTCF-связывающих последовательностей на 19 типах клеток в культуре и тканей человека методами ChIP-seq и массивированного бисульфитного секвенирования. 36% сайтов

связывания CTCF присутствовали во всех 19 типах клеток. Для 67% дифференциально метилированных сайтов связывания CTCF была показана обратная зависимость между степенью их метилирования и способностью связывать CTCF. Таким образом, метилирование ДНК является важным фактором, определяющим, будет ли белок CTCF связываться с данной нуклеотидной последовательностью. Также было показано, что около 29% CTCF-связывающих последовательностей в геноме содержат CpG хотя бы в одном из двух положений -1 и 11, что соответствует положениям 4 и 14 коровой последовательности сайта связывания CTCF (рисунок 1, см. [126]) и положениям -5 и 5 [155].

#### 1.2.4. Взаимодействие CTCF с белками

При иммунопреципитации из клеточного лизата CTCF соосаждается с белком ядрышка нуклеофосмином. С другой стороны, методом иммунопреципитации хроматина с антителами к нуклеофосмину было показано, что нуклеофосмин взаимодействует с CTCF-зависимыми инсуляторами *in vivo* [195]. Содержащая хромодомен хеликаза CHD8 также связывается с CTCF и CTCF-зависимыми инсуляторами. В отсутствие CHD8 область ICR локуса *Ig2/H19* перестает выполнять инсуляторную функцию, хотя связывание CTCF с этой областью сохраняется [78]. В белковых комплексах CTCF-зависимых инсуляторов было показано присутствие большой субъединицы РНК-полимеразы II с фосфорилированным и дефосфорилированным С-концевым доменом [20].

Путем ко-иммунопреципитации было установлено, что белок CTCF взаимодействует с ДНК-связывающим белком YY1. По-видимому, во взаимодействии участвует N-концевой домен CTCF. Взаимодействие CTCF с YY1 необходимо во время инактивации одной X-хромосомы для поддержания второй X-хромосомы в активном состоянии [33].

К настоящему времени различными методами было обнаружено еще несколько белков, способных взаимодействовать с CTCF. Это YB1 (Y-box-binding protein 1) – мультифункциональный ДНК и РНК-связывающий белок, участвующий в регуляции процессов репликации и репарации ДНК, транскрипции, процессинга РНК, также способный взаимодействовать с YY1 [21], и Kaiso – транскрипционный

фактор, обладающий способностью связывать участки ДНК с повышенным содержанием метилированных CpG-сайтов. По-видимому, он может замещать CTCF в случае метилирования сайта связывания последнего [28]. С белком CTCF также взаимодействуют транскрипционный корепрессор Sin3A [115], гистон H2A.Z [195; 9], PARP [58], p68 (DDX5) [192] и другие белки [131; 201].

Еще одним важным белком, взаимодействующим с CTCF, является когезин. Когезин отвечает за удерживание вместе сестринских хроматид, необходимое для успешного протекания митоза и мейоза [70; 77; 141] и представляет собой белковый комплекс, состоящий из четырех субъединиц: Smc1, Smc3, Scc1 (Rad21) и Scc3 (SA1 или SA2). Четыре субъединицы образуют структуру в форме кольца, внутри которого, предположительно, оказываются удерживаемые вместе хроматиды [62]. Известно, что когезин отвечает не только за взаимодействие между собой сестринских хроматид, необходимое для успешного протекания митоза и мейоза, но и вовлечен в регуляцию экспрессии генов [61; 79; 30]. По результатам ChIP-seq анализа было выявлено, что примерно половина сайтов связывания когезина перекрывается с сайтами связывания CTCF [139; 158; 184]. Когезин связывается с участком в С-концевой области CTCF через свою субъединицу SA2 [188].

Методом 3C (Chromatin Conformation Capture) была установлена пространственная сближенность участков интерфазного хроматина, с которыми взаимодействуют одновременно когезин и CTCF. При этом участок хромосомы между сайтами связывания когезина образует петлю [61].

### ***1.3. Роль транскрипционного фактора CTCF в регуляции ДНК-зависимых процессов***

Связывание CTCF с ДНК может влиять на экспрессию генов различным образом: в некоторых случаях CTCF играет роль активатора транскрипции, в других – репрессора или же обеспечивает активность инсулятора. Инсуляторами называют фрагменты ДНК, препятствующие взаимодействию регуляторных элементов, между которыми они располагаются. В частности, находясь между

промотором и энхансером, инсулятор блокирует активирующее действие последнего, а также, будучи помещенным по краям генетической конструкции в составе геномной ДНК эукариотической клетки, препятствует эффекту положения [69]. Все, найденные на данный момент инсуляторы позвоночных, за редкими исключениями [120], связывают транскрипционный фактор CTCF.

Кроме того белок CTCF участвует в инактивации X-хромосомы, импринтинге генетической информации и, по-видимому, регулирует процесс сплайсинга РНК. Разработка новых методов изучения взаимодействий удаленных друг от друга областей эукариотического генома позволила получить множество свидетельств того, что CTCF играет важнейшую роль в образовании трехмерной структуры эукариотического генома [121; 17; 135; 41; 63; 183]. Понимание механизмов, которые лежат в основе столь многочисленных и различных функций белка CTCF, должно прояснить его роль в регуляции ДНК-зависимых процессов.

### *1.3.1. Гипотеза функциональных доменов хроматина. Инсуляторы*

В конце 80-х годов XX века была выдвинута гипотеза, в соответствии с которой весь хроматин эукариотической клетки разделен на структурно-функциональные домены. В соответствии с этой гипотезой хроматиновый домен представляет собой петлю, содержащую один или несколько генов, концы которой закреплены в ядерном матриксе. Для отдельных петель характерна независимая сверхспирализация ДНК. Хроматин одного домена может независимо от хроматина других доменов переходить в открытую (транскрипционно активную) или закрытую (неактивную) конформацию [14; 54].

Со времени формулировки гипотезы хроматиновых доменов было получено множество данных, подтверждающих, дополняющих и вносящих изменения в эту гипотезу. В настоящий момент установлено, что основным фактором, контролирующим декомпактизацию хроматиновой фибриллы, а, следовательно, и возможность начала транскрипции на данном участке хромосомы, является ацетилирование гистонов [99; 3]. Выявлены так называемые LCR (locus control regions) – участки ДНК, определяющие транскрипционный статус домена [57; 46; 149].

Изучение  $\beta$ -глобинового локуса кур показало, что в клетках, экспрессирующих  $\beta$ -глобина, хроматин локуса находится в декомпактизованном состоянии, уровень ацетилирования гистонов повышен. При этом за пределами локуса хроматин находится в конденсированном состоянии [40; 67; 112]. Концевые участки  $\beta$ -глобинового локуса взаимодействуют с ядерным матриксом и сближены между собой, при этом локус образует петлевую структуру [137; 194; 103; 182; 172]. Показано также, что петли могут формироваться не только при сближении границ функциональных доменов, но и при сближении отдельных участков ДНК, например, энхансеров и промоторов, внутри доменов [124; 59; 172].

В недавних работах установлено, что транскрипционная активность гена зависит не только от регуляторных элементов внутри домена, но и от того, в какой части ядра данный ген располагается [121; 17; 135; 41; 183]. При этом ген может подвергаться воздействию регуляторных элементов, расположенных в других доменах и даже на других хромосомах [73; 111; 140]. Таким образом, дополненная и уточненная гипотеза структурно-функциональных хроматиновых доменов остается актуальной.

Если существуют хроматиновые домены, то должны быть функциональные элементы, препятствующие влиянию на гены одного домена регуляторных элементов, принадлежащих другим доменам. Такими функциональными элементами генома являются инсуляторы [185; 179]. Инсуляторы предотвращают нежелательную активацию или репрессию генов под влиянием окружения. Нежелательная активация гена энхансером подавляется путем блокирования его действия на промотор только в том случае, когда инсулятор располагается между ними. Нежелательную репрессию гена инсулятор предотвращает путем ограничения распространения конденсированного хроматина вдоль хроматиновой фибриллы [185; 179]. Некоторые авторы называют инсуляторы пограничными элементами (border elements) в связи с тем, что они часто находятся на границах доменов (рисунок 2) [53]. Подавляющее большинство инсуляторов, обнаруженных в геномах позвоночных, обладают способностью связывать фактор транскрипции CTCF [11; 185; 120].

СТСФ играет важнейшую роль в образовании хроматиновых петель. Показано, что в  $\beta$ -глобиновых локусах позвоночных сайты связывания СТСФ на границах локусов находятся в контакте друг с другом [137; 22; 172]. При этом в экспрессирующих  $\beta$ -глобины клетках не происходит активации энхансерами локуса промоторов соседних доменов. Аналогично регуляторные элементы соседних доменов не влияют на экспрессию  $\beta$ -глобиновых генов в эритроидных клетках [40; 67; 15; 161; 112].

Анализ генома человека показал, что распределение СТСФ-связывающих сайтов коррелирует с плотностью распределения генов, но слабо коррелирует с длиной хромосом [86]. Несмотря на зависимость количества сайтов связывания СТСФ от количества генов, большая часть этих сайтов (46%) располагается очень далеко от промоторов, на расстоянии в среднем 48 т.п.о. [86; 189]. Это отличает распределение сайтов связывания СТСФ от распределения сайтов большинства других транскрипционных факторов и согласуется с инсуляторной функцией СТСФ.

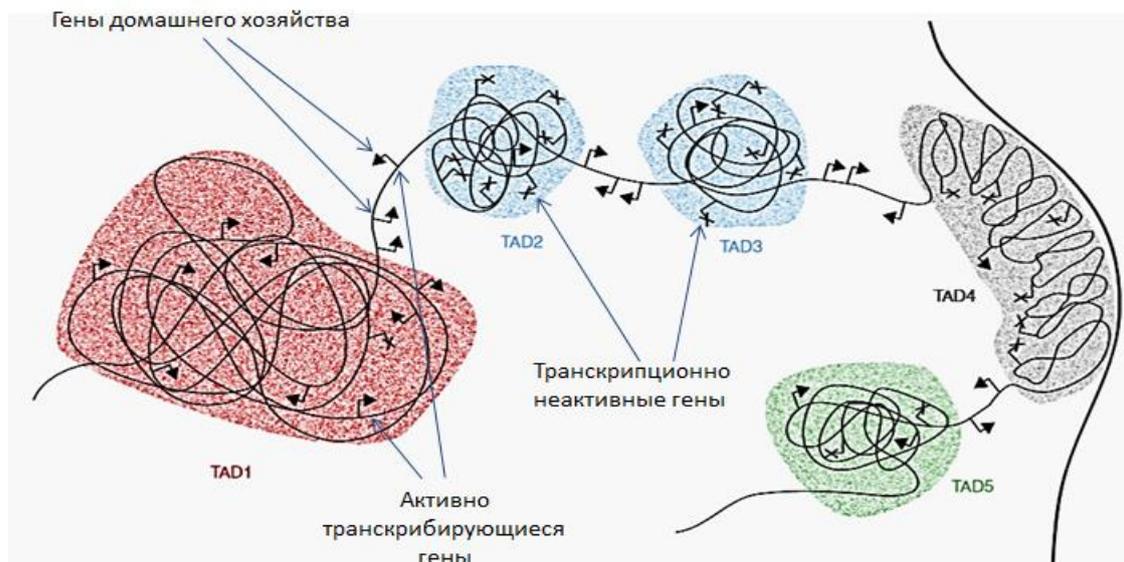
Обедненные сайтами связывания СТСФ участки генома содержат в основном семейства генов с общей регуляцией транскрипции, тогда как обогащенные сайтами связывания СТСФ домены содержат гены с альтернативно регулируемым промоторами. Эти наблюдения также согласуются с инсуляторной функцией СТСФ [86].

Последние данные, полученные преимущественно методом Hi-C, свидетельствуют о том, что трехмерную структуру геномов многоклеточных животных преимущественно составляют так называемые топологически ассоциированные домены, и разделяющие их участки хроматина (рисунок 3). Топологически ассоциированный домен (TAD - Topologically Associating Domain) представляет собой область хроматиновой фибриллы, участки которой преимущественно взаимодействуют между собой, а не с участками хроматина вне ее. При этом формируется глобула. Топологически ассоциированные домены ограничены инсуляторами, и участками хроматина, содержащими гены домашнего хозяйства. [24]

Многочисленные свойства топологически ассоциированных доменов соответствуют свойствам функциональных хроматиновых доменов, описанным ранее. Пограничные элементы хроматинового домена взаимодействуют между собой, обогащены сайтами связывания CTCF и, в большинстве случаев отделяют участки гетерохроматина от участков декомпактизованного хроматина. При этом, разные участки одного топологически ассоциированного домена имеют общие особенности структуры хроматина [24; 52]. Топологически ассоциированные домены, по-видимому, являются единицей репликации [145]. Таким образом, топологически ассоциированные домены могут соответствовать функциональным хроматиновым доменам, которые были выявлены и исследованы ранее другими методами.

В отличие от ранее описанных функциональных хроматиновых доменов, топологически ассоциированные домены обычно представляют в виде глобул, разделенных участками хроматиновой фибриллы, содержащей помимо инсуляторов гены домашнего хозяйства.

По-видимому, исследование топологически ассоциированных доменов скорее дополняет и немного видоизменяет первоначальную теорию функциональных хроматиновых доменов, нежели опровергает ее.



**Рисунок 3. Топологически ассоциированные домены. TAD – топологически ассоциированные домены, хроматин которых находится в различных состояниях: TAD1 - транскрипционно активном, декомпактизованном, TAD2 – TAD5 – в транскрипционно неактивных состояниях. [24]**

### 1.3.2. Блокирование энхансера

Первые данные об энхансер-блокирующих свойствах сайтов связывания CTCF были получены с использованием конструкций, содержащих находящийся под контролем промотора репортерный ген и энхансер. Между энхансером и промотором вводилась CTCF-связывающая последовательность [11; 151]. Анализ эффективности экспрессии репортерного гена при наличии и в отсутствие сайта связывания CTCF позволял судить об энхансер-блокирующей активности этого фрагмента ДНК. Большая часть CTCF-связывающих последовательностей проявляет энхансер-блокирующую активность в составе таких конструкций как при транзientной трансфекции, так и при интеграции в геном [2; 174; 31].

Метод оценки экспрессии репортерного гена при транзientной трансфекции имеет свои недостатки. Сайты связывания CTCF в норме функционируют в составе хроматина, а плазмидный вектор не имитирует полностью хроматинового окружения. Кроме того, во многих случаях плазмидный вектор представляет собой гетерологическую систему, в которой энхансер, промотор, сайт связывания CTCF и репортерный ген могут происходить из различных организмов.

В работе [198] при изучении механизма функционирования CTCF-зависимого инсулятора использовали искусственную минихромосому, содержащую энхансер из LCR локуса  $\beta$ -глобина человека и ген эpsilon-глобина с собственным промотором. CTCF-зависимый инсулятор 5'HS4 кур был клонирован между энхансером и промотором гена эpsilon-глобина, и для контроля – рядом с энхансером, с противоположной стороны от промотора. При расположении инсулятора между энхансером и промотором был продемонстрирован CTCF-зависимый энхансер-блокирующий эффект, а также CTCF-зависимое снижение количества РНК-полимеразы II, связанной с промотором и увеличение количества полимеразы, взаимодействующей с энхансером и CTCF-связывающей последовательностью. Этот эффект можно объяснить тем, что CTCF-содержащий ДНК-белковый комплекс мешает продвижению РНК-полимеразы от энхансера к промотору (функционирование энхансера по механизму tracking [185]). Отмечено также CTCF-зависимое снижение уровня ацетилирования гистонов H3 и H4 между

5'HS4 инсулятором и геном, включая область промотора, в том случае, когда инсулятор 5'HS4 располагался между энхансером и промотором [198].

Введение нескольких коровых последовательностей куриного 5'HS4 инсулятора в состав плазмидной конструкции между энхансером и промотором приводило к более осязаемому энхансер-блокирующему эффекту, нежели введение одной такой последовательности [23]. Такой же эффект наблюдается и при помещении между энхансером и промотором нескольких CTCF-зависимых инсуляторов HS1 или HS2 из ICR локуса *Igf2/H19* мыши [64; 80]. Таким образом, несколько CTCF-содержащих ДНК-белковых комплексов затрудняют перемещение РНК-полимеразы II от энхансера к промотору эффективней, чем один.

Однако передачу сигнала вдоль хроматиновой фибриллы вряд ли можно рассматривать как единственный механизм функционирования энхансеров. Многие энхансеры отделены от своих промоторов-мишеней миллионами нуклеотидных пар [65] и энхансер и промотор зачастую оказываются разделены несколькими хромосомными локусами с различной структурой хроматина. Кроме того, описана активация промотора энхансером, расположенным на другой хромосоме [73; 140].

Некоторые авторы считают, что этот механизм вообще не реализуется в случае настоящих эукариотических энхансеров, а только в случае энхансер-подобных элементов прокариот [102].

В последнее время с помощью технологии 3C (chromosome conformation capture) и ее производных [164] показано непосредственное взаимодействие энхансера и промотора, который он активирует [170; 171], хотя механизм, по которому формирование хроматиновой петли и сближение энхансера и промотора ведет к активации промотора точно неизвестен [101]. Возможно, за счет приближения промотора к энхансеру происходит привлечение промотора в тот компартмент ядра, где есть условия для активной транскрипции [101]. Способность стабилизировать хроматиновые петли была продемонстрирована методом ChIA-PET для различных белков, в том числе для CTCF [63; 111].

В соответствии с моделью ловушки [51; 69] CTCF-связывающие инсуляторы могут конкурировать с промоторами за взаимодействие с энхансерами, то есть перехватывать энхансер, не позволяя ему связаться с промотором. В работе [63]

показано, что значительная часть CTCF-связывающих последовательностей, участвующих в удаленных взаимодействиях, связана с последовательностями, обладающими свойствами энхансеров. С другой стороны, если инсулятор функционирует по модели ловушки, то между энхансером и промотором должна быть передача сигнала вдоль хроматиновой фибриллы, иначе ловушка будет обладать одинаковым действием, располагаясь с любой стороны от энхансера, что противоречит определению инсулятора. В  $\beta$ -глобиновых локусах позвоночных методом 3C было показано, что энхансеры и промоторы физически сближены между собой [137], что свидетельствует в пользу того, что энхансер функционирует по комбинированному механизму tracking-looping.

### 1.3.3. Инсулятор как пограничный элемент

При введении в геном клетки конструкций, содержащих трансген, экспрессия трансгена сначала происходит приблизительно на одном уровне во всех клетках, но через какое-то время большая часть встроившихся в геном конструкций становится неактивной из-за конденсации хроматина. Если у дрозофилы и дрожжей это ингибирование экспрессии обратимо, то у позвоночных становится необратимым из-за метилирования ДНК [144]. Инсулятор 5'HS4, расположенный на 5'-конце  $\beta$ -глобинового локуса, препятствует распространению расположенного рядом участка конденсированного хроматина на весь локус.

В работе [27] методом ChIP-seq показано, что часть сайтов связывания CTCF локализуется в местах перехода компактизованного хроматина в декомпактизованный. Такие места определялись по изменению уровня триметилирования остатка лизина 27 гистона H3 (H3K27me3) и ацетилирования остатка лизина 5 гистона H2A (H2AK5ac). Модификация H3K27me3 характерна для компактного хроматина, а H2AK5ac – для декомпактизованного. Доля CTCF-связывающих последовательностей, расположенных в местах перехода компактизованного хроматина в декомпактизованный, невелика, но с высокой значимостью показан неслучайный характер расположения этих последовательностей. Это указывает на возможную роль сайтов связывания CTCF в разделении хромосомных доменов *in vivo*.

В работе [32] показано значительное обогащение сайтами связывания CTCF пограничных областей так называемых топологически ассоциированных доменов. Авторы предположили, что пограничные области топологически ассоциированных доменов могут представлять собой инсуляторы, блокирующие распространение гетерохроматина. Действительно, в дифференцированных клетках пограничным элементам топологически ассоциированных доменов соответствуют области с пониженным уровнем содержания H3K9me, характерные для перехода гетерохроматина в менее компактный хроматин. По-видимому, при формировании окончательной структуры геномных доменов стабилизация состояния хроматина происходит после формирования петлевой структуры [32].

Возможно, CTCF-связывающие последовательности не задействованы напрямую в блокировании распространения гетерохроматина. Однако четкое разделение в геноме участков компактизованного и декомпактизованного хроматина может быть следствием образования CTCF-содержащими пограничными элементами петлевых структур.

Показано, что взаимодействующие между собой сайты связывания CTCF на границах петлевых доменов имеют определенную ориентацию. N-концы, связанных с ними молекул CTCF, направлены в сторону петлевой структуры [29; 52; 60; 175].

#### *1.3.4. CTCF как регулятор транскрипции*

CTCF способен как активировать, так и ингибировать транскрипцию в зависимости от целевого гена. Механизмы активации могут быть различными. Во-первых, CTCF может выступать как белок-регулятор транскрипции, способствуя или препятствуя образованию инициаторного комплекса. В других случаях связывание CTCF может приводить к образованию/разрушению хроматинового домена с соответствующим изменением структуры хроматина, что, в свою очередь, приводит к изменению уровня экспрессии находящихся в данном домене генов. Нужно отметить, что четкую границу между этими механизмами можно провести далеко не всегда. По первому механизму, скорее всего, осуществляется репрессия

гена *MYC* [114], по второму – репрессия гена *PUMA* [55]. Некоторые примеры регуляции с участием CTCF приведены ниже.

При связывании CTCF с 5'-регуляторной областью гена *APP* (amyloid  $\beta$  protein precursor), наблюдалась активация промотора этого гена при транскрипции *in vitro* в ядерном экстракте клеток HeLa [176; 177]. При добавлении к ядерному экстракту олигонуклеотидов, конкурирующих за связывание CTCF, активность промотора снижалась. Тот же эффект наблюдался при удалении CTCF из ядерного экстракта путем иммунопреципитации, однако при добавлении в обедненный ядерный экстракт белка CTCF, активность промотора восстанавливалась. За активацию ответственен участок белка CTCF между аминокислотными остатками 1 и 248 [177].

Показано, что гиперчувствительный к ДНК-азе I участок ДНК HS5-1, расположенный в кластере протокадгериновых генов, обладает свойствами энхансера [156] и, по данным иммунопреципитации хроматина, взаимодействует с CTCF в клетках головного мозга мышей. При делеции этой CTCF-связывающей последовательности из генома трансгенных мышей активность промотора снижается [83]. Методом иммунопреципитации хроматина с последующим масштабным секвенированием продемонстрировано, что CTCF и когезин взаимодействуют с промотором альтернативных форм  $\alpha$ -протокадгерина. Связывание прямо коррелирует с экспрессией альтернативных изоформ [125]. Методом 3C показано пространственное сближение промоторов  $\alpha$ -протокадгеринов 4, 8 и 12 с потенциальными энхансерами HS5-1 и HS7 [59].

Сайленсер гена лизоцима кур состоит из двух модулей F1 и F2. F2 связывает рецептор тиреоидных гормонов, тогда как F1 содержит сайт связывания CTCF, при этом оба модуля могут подавлять транскрипцию независимо друг от друга. Когда каждый из них связывает свой белковый фактор, их действие суммируется [7]. Возможно, в механизмы супрессии-активации гена лизоцима кур вовлечен косупрессор Sin3A и комплексы гистоновых ацетилаз и деацетилаз [115; 34; 116].

Взаимодействие CTCF с двумя CpG-островками в первом интроне гена *Bcl6* не позволяет этому гену активно экспрессироваться. При некоторых разновидностях лимфомы эти CpG-островки метилируются, что препятствует

связыванию CTCF с ДНК и приводит к увеличению количества мРНК *Bcl6* в клетках [106].

Связывание CTCF с участком в первом экзоне гена теломеразы человека *hTERT* подавляет транскрипцию этого гена. Метилирование CTCF-связывающей последовательности, по-видимому, играет важную роль в регуляции экспрессии этого гена. Первый экзон гена теломеразы в составе плазмидной конструкции оказывает репрессорное действие на ряд промоторов как в клетках, экспрессирующих теломеразу, так и в клетках, ее не экспрессирующих. При этом в клетках, экспрессирующих теломеразу, наблюдается метилирование CTCF-связывающей последовательности в первом экзоне этого гена и отсутствие взаимодействия CTCF с этой последовательностью [153].

После обработки 5-аза-2'-дезоксцитидином, приводящей к глобальному деметилированию ДНК, в клеточных линиях, экспрессирующих *hTERT*, наблюдали связывание CTCF с фрагментом первого экзона этого гена и подавление образования мРНК *hTERT*. Подавление экспрессии CTCF с помощью коротких шпилечных РНК приводило к увеличению количества мРНК *hTERT* в клеточных линиях [153; 154].

Отмечено также участие CTCF в регуляции транскрипции генов *Bax* [123], опухолевого репрессора p16 [148], и генов рибосомных РНК [74].

### 1.3.5. Участие белка CTCF в импринтинге генетической информации

Под генным импринтингом понимают способ наследования, при котором у потомства экспрессируется только один из родительских аллелей гена. Какой из аллелей будет экспрессироваться, определяется тем, от отцовского или материнского организма он был унаследован. CTCF взаимодействует с областями, контролирующими импринтинг (ICR - imprinting control region). Наиболее известным примером участия CTCF в импринтинге является регуляция генов локуса *Igf2/H19* [64; 80]. Ген *Igf2* (Insulin-like growth factor 2) кодирует эмбриональный митоген [64], с гена *H19* транскрибируется некодирующая РНК, замедляющая рост плода [109; 48; 69]. Экспрессия *Igf2* происходит только с отцовской, а *H19* – только с материнской хромосомы, что достигается при помощи

расположенного между этими генами ICR. ICR на материнской хромосоме мышей содержит два сайта, способных связывать CTCF и обладающих свойствами инсуляторов. На отцовской хромосоме ICR метилирован независимо от ткани и стадии развития и не способен связывать CTCF. Таким образом, на отцовской хромосоме CTCF-зависимые инсуляторы неактивны, промотор свободно взаимодействует с энхансерами и *Igf2* экспрессируется [110].

Взаимодействуя с ICR материнской хромосомы, CTCF предположительно защищает от метилирования близлежащие области локуса, в том числе промотор и внутренние области гена *H19*. Если материнский ICR мутирован и не связывает CTCF, метилирование распространяется на промотор и внутригенные участки *H19*, и уровень экспрессии этого гена снижается [160; 168]. Таким образом, метилирование отцовского ICR, подавляя связывание CTCF, приводит к метилированию промотора гена *H19* и подавлению экспрессии отцовского аллеля [49; 160].

### 1.3.6. CTCF и инактивация X-хромосомы

CTCF участвует в процессе инактивации X-хромосомы млекопитающих [18]. В состав X-хромосомы входит центр инактивации (Xic, X-chromosome inactivation center). Этот элемент контролирует инактивацию одной из X-хромосом в каждой клетке в ходе эмбриогенеза и дальнейшее поддержание ее неактивного состояния. В состав Xic входят гены *Xist* (X-inactive specific transcript), *Tsix* (название образовано от *Xist* путем перестановки букв в обратном порядке, что отражает обратную ориентацию гена *Tsix* по отношению к *Xist*) и *Xite* (X-inactivation intergenic transcription element), с которых транскрибируются некодирующие РНК. X-хромосома, на которой расположен активно экспрессирующийся аллель гена *Xist*, инактивируется. Транскрипция генов *Tsix* и *Xite* происходит с активной X-хромосомы. Транскрипт гена *Tsix* инактивирует экспрессию гена *Xist*, расположенного на той же хромосоме, а экспрессия гена *Xite* активирует экспрессию *Tsix*, что приводит к подавлению экспрессии гена *Xist*, в результате чего X-хромосома, на которой находятся транскрипционно активные аллели генов *Tsix* и *Xite*, остается активной [108; 201].

Процесс инактивации X-хромосомы происходит в несколько стадий [5; 25; 37; 108]. Это “подсчет” количества X-хромосом в клетке, а точнее определение отношения количества X-хромосом к количеству аутосом [5; 108]. На следующей стадии происходит гомологичное спаривание двух X-хромосом и выбор той, которая будет инактивирована. Взаимодействие X-хромосом происходит в районе *Xic*, причем для спаривания X-хромосом необходим и достаточен фрагмент *Xic*, включающий гены *Xite* и *Tsix* [6; 191]. Эта область изобилует сайтами связывания CTCF, а также белка YY1 [18; 33; 190]. Показано, что при отсутствии в клетке CTCF спаривание X-хромосом подавляется. Белок YY1, способный взаимодействовать с CTCF, не является необходимым для протекания этой стадии [190]. Существует модель, в соответствии с которой в результате спаривания X-хромосом происходит необратимый перенос белковых факторов, таких как CTCF и Oct-4, с одной X-хромосомы (в будущем неактивной) на другую (в будущем активную) [191; 33; 128; 129; 190; 127]. Если эта модель верна, то белок CTCF задействован и в выборе X-хромосомы, которая будет инактивирована. Последняя стадия - это сам процесс инактивации X-хромосомы. При этом происходит активная экспрессия гена *Xist*, транскрипт гена *Xist* покрывает ту хромосому, на которой он был синтезирован и хроматин этой X-хромосомы компактизируется [5; 25; 37].

В процессе инактивации X-хромосомы идет взаимодействие репрессорного комплекса PRC2 с 5'-областью гена *Xist* и обогащение хроматина этой области гистоном H3, триметилированным по остатку лизина 27 [167; 199]. Для инактивации X-хромосомы важна транскрипция небольшого участка гена *Xist*, в результате чего образуется РНК, которую называют RepA [187; 71]. По-видимому, RepA взаимодействует с комплексом PRC2 [199]. Показано, что X-хромосома, содержащая в промоторе гена *Xist* сайт связывания CTCF с повышенным сродством к этому белку, чаще подвергается инактивации, чем X-хромосома, на которой такой фрагмент ДНК обладает сниженным сродством к CTCF [146]. Вместе с YY1 CTCF, по-видимому, участвует в активации экспрессии гена *Tsix* [33]. CTCF также обеспечивает функционирование инсуляторных элементов,

отделяющих расположенные на инактивированной X-хромосоме активные гены от транскрипционно неактивного окружения[42].

### *1.3.7. Котранскрипционная регуляция альтернативного сплайсинга*

Недавно стали появляться свидетельства в пользу того, что CTCF может быть вовлечен в котранскрипционную регуляцию альтернативного сплайсинга. Было продемонстрировано, что взаимодействие CTCF с фрагментом ДНК, расположенным в пятом экзоне гена *CD45*, приводит к более частому включению этого экзона в состав зрелой мРНК. Уменьшение количества CTCF в клетке при помощи РНК-интерференции или метилирование сайта связывания CTCF в пятом экзоне гена *CD45* приводило к подавлению образования сплайс-формы, содержащей пятый экзон. Было показано, что связывание CTCF с этим сайтом приводит к тому, что РНК-полимераза II, достигая участка ДНК, взаимодействующего с CTCF, делает паузу. Было проведено сравнение полногеномных данных, полученных путем иммунопреципитации хроматина с антителами к CTCF и РНК-полимеразе II и данных анализа транскриптома и получено подтверждение того, что связанный с ДНК белок CTCF задерживает транскрибирующую РНК-полимеразу II [178; 163].

На основе этих данных была предложена гипотеза о роли белка CTCF в регуляции альтернативного сплайсинга. Начальные этапы сплайсинга РНК происходят во время транскрипции, при этом преимущественное образование той или иной сплайс-формы может зависеть от скорости элонгации. Если после появления в составе транскрипта относительно слабого сплайс-сайта, но до появления более сильного, РНК-полимераза II будет задержана связанным с ДНК белком CTCF, то более слабый сплайс-сайт получит преимущество. Таким образом, белок CTCF может влиять на скорость элонгации и регулировать альтернативный сплайсинг [163; 97].

## 2. Материалы и методы

### 2.1. Материалы

#### Реактивы

Этиловый спирт, изоамиловый спирт, изопропиловый спирт, ацетат аммония, ацетат натрия, фенол, хлороформ медицинский, ацетон, кислота соляная, кислота уксусная ледяная, хлорид магния, хлорид калия, хлорид натрия, хлорид цинка, гидроксид натрия, сульфат меди - квалификация “осч”, “хч”, “чда”, Реахим, Россия.

Дезоксирибонуклеозид-5'-трифосфаты, ДНК спермы лосося (Salmon Sperm DNA) – Pharmacia, Швеция.

Натрия додецилсульфат (SDS), тритон X-100, спермин, спермидин, Tween® 20 – Serva, Германия.

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

Обезжиренное сухое молоко, бромфеноловый синий, ксиленцианол, бромистый этидий, 2-меркаптоэтанол 98%, 2x буфер для нанесения проб для электрофореза по Лэммли, Bio-Rad Protein Assay Dye Reagent Concentrate (#5000006) – Bio-Rad, США.

Дезоксинуклеозидтрифосфаты – Евроген, Россия

Метиленбисакриламид, TEMED, персульфат аммония – Amersham Pharmacia Biotech, Швеция.

Хлорид магний (25 мМ), ДНК спермы лосося (100 нг/мкл) – Promega, США.

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

Протеин-А агароза, Lipofectamine 2000, HEPES – Gibco BRL, США.

Протеин-G магнитные частицы (Dynabeads Protein-G 10004D) – Life Technologies, США.

Ампициллина натриевая соль – Биосинтез, Россия.

Наборы для приготовления проявителя и фиксажа кислого – Випс-Мед, Россия.

Масло вазелиновое - Тверская фармацевтическая фабрика, Россия.

Коктейль ингибиторов протеиназ, бензидин 1-(5-изохинолинсульфонил)-3-метилпиперазиндигидрохлорид, 30% раствор перекиси водорода, имидазол, бетаин, диметилсульфоксид (ДМСО) – Sigma, США

Формальдегид, 36,5% - Reidel-de Haen, Германия.

Nonidet P-40, X-gal 20 мг/мл в диметилформамиде, IPTG 1М водный – Helicon, США.

Гликоген - Thermo Fisher Scientific, США.

### **Ферментные препараты**

Тақ ДНК-полимераза (1-5ед/мкл; hot-start modified) – Гос НИИ генетики, Россия.

Протеиназа К, РНКаза А из поджелудочной железы быка (10мг/мл) – ИБХ РАН, Россия.

Эндонуклеазы рестрикции *EcoRI*, *Csp6I* – MBI Fermentas, Литва.

Эндонуклеаза рестрикции *Sau3AI* – New England BioLabs, США.

Фрагмент Кленова ДНК-полимеразы I *E. coli* – MBI Fermentas, Литва.

Трипсин – Calbiochem, США.

Т4 ДНК лигаза (5 ед/мкл) – Promega, США.

### **Антитела**

Антитела к глицеральдегид-3-фосфатдегидрогеназе (sc-47724), антитела к N-концевому пептиду белка CTCF человека (sc-5916), конъюгированные с пероксидазой антитела к иммуноглобулинам-G козла (sc-2304)– Santa-Cruz, США.

Кроличыи поликлональные антитела к N-концевой части куриного CTCF (аминокислотные остатки 86-233), любезно предоставленные О.В. Яровой, ИБГ РАН.

Антитела к полигистидиновой последовательности аминокислот, гистидиновому тагу, – Sigma H1029, США.

Конъюгированные с пероксидазой антитела к иммуноглобулинам-G кролика (7074), к иммуноглобулинам-G мыши (7076) – Cell Signaling, США.

### **Растворы**

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

10x буфер для T4 ДНК лигазы, с 10mM АТФ – Promega, США.

10x буфер EcoRI, 10x буфер B для эндонуклеаз рестрикции – MBI Fermentas, Литва.

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

Буфер TE, pH 7,6: 10 мМ Tris-HCl pH 7,6, 1 мМ ЭДТА

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

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

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

Буфер А: 10 мМ HEPES–KOH, pH 7,9, 10% глицерин, 25 мМ KCl, 1 мМ EDTA, 0,15 мМ спермин, 0,5 мМ спермидин, 0,5 мМ AEBSF (Sigma, США), 0,5 мМ ДТТ.

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

Трис-глициновый буфер: 25 mM Tris, 192 mM glycine, 0,1% SDS pH=8,3.

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  $\text{KH}_2\text{PO}_4$ , 5,2 мМ  $\text{Na}_2\text{HPO}_4$ , 150 мМ NaCl.

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

Буфер PBS – ICN, США.

Лизирующий буфер: 50 мМ Трис-НСl pH 8,0, 1% SDS, 10 мМ ЭДТА.

IP буфер: 16,7 мМ Трис-НСl, pH 8,0, 16,7 мМ NaCl, 1,2 мМ ЭДТА, 1% Тритон X-100, 0,01% SDS.

Буфер элюции: 1% SDS, 0,1 М NaHCO<sub>3</sub>.

Растворы, используемые для иммунопреципитации хроматина:

1. 0,1% SDS; 1% Тритон X-100; 2мМ ЭДТА; 20 мМ Трис-НСl, pH 8,1; 150 мМ NaCl.

2. 0,1% SDS; 1% Тритон X-100; 2мМ ЭДТА; 20 мМ Трис-НСl, pH 8,1; 500 мМ NaCl.

3. 0,25 М LiCl; 1% NP-40; 1% дезоксихолат натрия; 1 мМ ЭДТА; 10 мМ Трис-НСl, pH 8,1.

4. Буфер TE, pH 8,0.

### **Питательные среды**

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

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 агаризованной LB до 50°С добавляли раствор ампициллина до конечной концентрации 100 мкг/мл.

Сухая смесь для приготовления культуральных сред DMEM/F12 НАМ, OPTI-MEM, эмбриональная сыворотка теленка (FCS), куриная сыворотка – GibcoBRL (США).

### **Наборы реактивов**

Wizard Plus Minipreps DNA Purification System, Wizard Plus Midipreps DNA Purification System, Wizard SV Gel and PCR Clean-Up System, Wizard Genomic DNA Purification Kit – Promega, США.

Protino Ni-IDA 150 Packed Columns – Macherey-Nagel, Германия.

Immun-Star HRP Chemiluminescent detection kit – Bio-Rad, США.

Набор Encyclo Plus PCR kit – Евроген, Россия.

QIAquick PCR Purification Kit, Plasmid Midi Kit, QIAquick Gel Extraction Kit – Qiagen, США.

### **Готовые смеси для ПЦР**

Готовая смесь для ПЦР ScreenMix-HS, готовая смесь для ПЦР qPCRmix-HS SYBR – Евроген, Россия.

### **Маркеры длин ДНК**

100 п.о. + 1,5 т.п.о., 1 т.п.о. – СибЭнзим, Россия.

Fermentas mass ruler high range, Fermentas mass ruler low range – Fermentas, Литва.

### **Маркеры длин белков**

Набор биотинилированных белков с молекулярными массами 9-200 кДа, (7727) – Cell Signaling, США.

Преокрашенный маркер PageRuler Prestained Protein Ladder (SM0671) – Fermentas, Литва.

### **Оборудование**

Автоматические микропипетки – Gilson, Франция.

Автоматический CO<sub>2</sub> инкубатор IR 1500, ламинар Gelaire – Flow Laboratories, США.

Источники питания: Power Supply Model 500, Power Pac, электрофорезный прибор Mini Protein III, прибор для электроблоттинга белков Trans-Blot<sup>R</sup> SD Semi-Dry Transfer Gel Bio, система для визуализации и документирования VersaDoc MP 4000 System – Bio-Rad, США.

Ламинар (Laminar Flow) – Babcock-BSH, ФРГ.

Микроцентрифуга (5415, MiniSpin Plus) – Eppendorf, ФРГ.

Одноразовая стерильная пластиковая посуда – Corning, SSI, США.

Приборы (камеры) для горизонтального электрофореза – Pharmacia Biotech Швеция.

ПЦР-амплификатор PTC-150 – MJ Research, Inc., США.

Термостатируемый шейкер Certomat®H, – B.Braun, США.

Горизонтальный шейкер - Red Rocker, San Francisco, США.

Термостатируемый шейкер для микропробирок “Thermomixer comfort” – Eppendorf, ФРГ.

УФ-спектрофотометр GeneQuant – Amersham Pharmacia Biotech, Швеция.

Система гель-документирования G:BOX – Syngene, Великобритания.

Центрифуга (GPR Centrifuge) – Beckman J2-21, США.

Ультразвуковой дезинтигратор (Ultra Sonic Processor) – Cole Parmer Instrument Company, США.

УФ-трансиллюминатор – UVP, Канада.

Электронные весы (Sartorius E 5500 S) – Sartorius GmbH Göttingen, ФРГ.

Рентгеновская пленка – РМ-В, Россия.

Иммобилон – Millipore, США.

Гомогенизатор Даунса – Wheaton Scientific, США.

Спектрофотометр Cary 50 Bio – Varian, США.

Магнитный штатив для микропробирок 1,5 мл – Dynal AS, Норвегия.

Прибор для количественной ПЦР Mx3000P – Stratagene, США.

### **Клеточные линии**

В работе использовали клеточные линии HeLa (аденокарцинома шейки матки человека, ATCC CCL-2), COS-1 (клетки почки африканской зеленой мартышки, ATCC CRL-1650), HD3 (клон А6 линии LSCC куриных эритробластов [13]), DT40 (В-клеточная лимфома кур, ATCC CRL-2111), HEK293 (клетки почек эмбриона человека, трансформированные ДНК аденовируса Ad5, ATCC CRL-1573).

### **Штаммы бактерий**

В работе использовали следующие штаммы *E. coli*:

**DH-5 $\alpha$** - *F'*/endA1 hsdR17(*r<sub>K</sub>*<sup>-</sup>,*m<sub>K</sub>*<sup>+</sup>) supE44 thi-1 recA1 gyrA(*Nal*<sup>r</sup>) relA1  $\Delta$ (*lacZYA-argF*)<sub>U169</sub> ( $\phi$ 80dLacZ $\Delta$ M15)

## 2.2. Методы

### Стандартные методики

Получение компетентных клеток *E.coli*, трансформацию, рестрикцию, лигирование, гель-электрофорез и другие стандартные манипуляции проводили по описанным методикам [159]. Выделение плазмидной ДНК проводили из 5 или 200 мл ночной культуры *E.coli* с использованием наборов фирмы Promega, Wizard Plus Minipreps DNA Purification System и Wizard Plus Minipreps DNA Purification System, соответственно, как было описано в прилагаемой инструкции. Выделение геномной ДНК проводили из 5-10 млн клеток кур, клеточная линия HD3 или человека, клеточная линия HeLa, с использованием набора фирмы Promega, Wizard Genomic DNA Purification Kit, как было описано в прилагаемой инструкции.

### Получение плазмидной конструкции рНис-his

1 мкг плазмиды рНис расщепляли эндонуклеазой рестрикции Eco RI, фирмы Fermentas (Литва), в растворе соответствующего буфера этой же фирмы. К плазмидной ДНК добавляли по 5 ед. эндонуклеазы рестрикции Eco RI, буфер EcoRI до 1х разведения, и деионизованную воду до нужного объёма. Реакционную смесь инкубировали при 37°C в течение часа. Инактивацию эндонуклеазы рестрикции после проведения реакции осуществляли нагреванием при температуре 80°C в течение 20 мин.

Плазмидную ДНК очищали с помощью экстракции смесью фенол:хлороформ 1:1.

Для получения олигонуклеотидного дуплекса олигонуклеотиды His fw (AATTCGAAATGGGACACCACCACCACCAC) и His rv (AATTGGTGGTGGTGGTGGTGTCCCATTTTCG) были синтезированы

фирмой Евроген (Россия). На гибридизацию было взято по 215 пмоль каждого олигонуклеотида. К смеси олигонуклеотидов был добавлен буфер TNM до 1x разведения. Гибридизационную смесь инкубировали 5 мин при 95°C, после чего температуру снижали до 20°C со скоростью ~2°C/мин, затем немедленно переносили в ледяную баню.

Для лигирования использовали 60 нг плазмиды рННс, линейаризованной по сайту EcoRI, и 22 пмоля олигонуклеотидного дуплекса, а также 2 ед. ДНК-лигазы фага T4 (Promega, США) в соответствующем буфере (все манипуляции производились на льду). Инкубировали при 4°C в течение 16 часов, далее лигазу инактивировали прогреванием смеси в течение 15 мин при 70°C.

### **Приготовление компетентных клеток *E.coli* [76]**

10 колоний штамма DH-5α из чашки с LB-агаром переносили в колбу объемом 2 л с 250 мл среды SOB и инкубировали при 18-25° С до плотности 0,6–1,0 OЕ<sub>600</sub>. После достижения достаточной плотности клетки помещали в лед на 10 мин. Клетки собирали при 3000 об/мин и 4°C в течение 15 мин. Осадок ресуспендировали в 80 мл охлажденного во льду буфера ТБ (10mM PIPES, 55mM MnCl<sub>2</sub>, 15mM CaCl<sub>2</sub>, 250mM KCl, pH-6,7), инкубировали 10 мин на льду и осаждали центрифугированием при 3000 об/мин в течение 15мин. Осадок ресуспендировали в 20мл буфера ТБ и при медленном помешивании добавляли ДМСО до 7%. Инкубировали 10 мин, разделяли на аликвоты и быстро замораживали в жидком азоте. Хранили клетки при температуре –70°C.

### **Трансформация клеток *E.coli***

К 100 мкл суспензии компетентных клеток *E. coli* штамма DH-5α *E.coli*, предварительно размороженных во льду, добавляли 5 нг вектора или 10 мкл (половину) лигазной смеси, инкубировали во льду в течение 30 мин для адсорбции вектора на поверхности клеток. Затем проводили тепловой шок, инкубируя клеточную суспензию при 42°C 30 сек, и переносили в лёд на 2 минуты. К трансформированным клеткам добавляли среду LB до конечного объема 1 мл и инкубировали при 37°C в течение 60 минут, перемешивая раз в 20 мин.

Трансформированные плазмидой рННс или рННс-his клетки высевали на чашки Петри с агаризованной средой LB и ампициллином (100 мкг/мл). Клетки, трансформированные вектором рGEM-T, высевали на чашки Петри с агаризованной средой LB, ампициллином (100 мкг/мл) и X-gal/IPTG (20 мкл X-gal (50 мг/мл) и 100 мкл IPTG (100 мМ)). Чашки Петри инкубировали в термостате при 37°C в течение ночи.

### **Выделение плазмид**

После трансформации отбирали часть индивидуальной колонии *E.coli* с чашки Петри, суспендировали клеточный материал в 10 мкл воды или 200 мкл LB. 5 мкл полученной суспензии клеток в воде или 1 мкл суспензии клеток в LB использовали для ПЦР-анализа на наличие плазмиды со вставкой.

Оставшиеся 5 мкл суспензии клеток в воде переносили в LB с ампициллином (100 мкг/мл). Для получения плазмиды рННс-his в небольших количествах (до нескольких десятков мкг) использовали 3-5 мл LB с ампициллином, для получения нескольких сотен мкг плазмидной ДНК – 200 мл. Суспензию клеток в среде инкубировали в течение ночи на термостатируемой качалке при 180 об./мин. при 37°C.

Клетки *E. coli*, трансформированные конструкцией рGEM-T со вставками фрагментов библиотеки, ранжированные в лунки 96-луночных планшетов, высевали и выращивали в течение ночи при 37 °C в 3-5 мл среды LB ампициллином (100 мкг/мл).

Клетки ночной культуры осаждали центрифугированием (10 000 об./мин), из полученных осадков выделяли плазмидную ДНК в соответствии с протоколом фирмы Promega (США) наборами реактивов Wizard Plus Midipreps DNA Purification System и Wizard Plus Midipreps DNA Purification Systems. Концентрацию плазмидной ДНК определяли спектрофотометрически с помощью спектрофотометра GeneQuant по оптической плотности при 260 нм. Препараты выделенной плазмидной ДНК

хранили при  $-20^{\circ}\text{C}$ . Плазмиду, выделенную из ночной культуры клеток объемом 200 мл предварительно разводили в буфере TE в 50 раз.

### **Анализ клонов на наличие вставки путем ПЦР**

Для выявления плазмиды рННс со вставкой в нужной ориентации, клетки *E. coli*, ресуспендированные в 5 мкл воды разрушали нагреванием, инкубируя при  $95^{\circ}\text{C}$  в течении 5 минут. Суспензию разрушенных клеток вносили в реакционную смесь в качестве матрицы для ПЦР. Также в реакционную смесь добавляли по 1 мкл плазмидного праймера рSG5 fw (10мкМ) ACAGCTCCTGGGCAACGTGC и олигонуклеотида His rv(10мкМ), который использовался для введения в конструкцию рННс фрагмента ДНК, кодирующего гистидиноый таг, 2,5 мкл 10x буфер для Taq-полимеразы (ИБХ РАН, Россия),  $\text{MgCl}_2$  (25мМ), 2 мкл смеси dNTP (2,5мМ), 0,5-0,1 мкл Taq ДНК-полимераза (1-5ед/мкл; hot-start modified Гос НИИ генетики, Россия) и деионизированную воду до 25мкл.

Для проверки наличия и размера вставок 1 мкл суспензии клеток из каждой лунки ранжированной библиотеки помещали в 10 мкл воды, прогревали 5 мин при  $96^{\circ}\text{C}$ , добавляли по 2 мкл праймера 27 (10 пмоль/мкл), 5 мкл 5x готовой смеси для ПЦР ScreenMix (Евроген, Россия), деионизированную воду до 25 мкл.

ПЦР амплификацию (20-25 циклов) проводили по следующей схеме:  $94^{\circ}\text{C}$  - 30 с,  $60^{\circ}\text{C}$  - 30 с,  $72^{\circ}\text{C}$  - 50 с. Продукты реакции разделяли в 1,5% агарозном геле.

### **Трансфекция эукариотических клеток с использованием липофектамина**

Трансфекцию эукариотических клеток COS-1 осуществляли с использованием липофектамина (Lipofectamine 2000, Gibco-BRL, США) согласно рекомендациям фирмы Gibco-BRL. За 24 часа до трансфекции клетки рассевали по  $2 \times 10^6$  клеток COS-1 в культуральный флакон с площадью дна  $25 \text{ см}^2$  (малый) или  $12 \times 10^6$  клеток во флакон с площадью дна  $150 \text{ см}^2$  (большой). Для трансфекции клеток плазмидой рННс-his в малом флаконе использовали 3 мкг и 6 мкг плазмиды – на флакон. К раствору плазмиды добавляли 100 мкл среды Opti-MEM (Gibco, США) без сыворотки и около 120 мкл раствора Lipofectamine 2000 в той же среде

(15 или 20 мкл реагента Lipofectamine 2000 смешивали со 100 мкл среды и инкубировали в течение 5 мин при комнатной температуре). После инкубации в течение 20 мин при комнатной температуре к смеси липосом с ДНК добавляли 1,2 мл среды OPTI-MEM. Клетки промывали 1 мл той же среды. В каждый малый культуральный флакон добавляли по 1,4 мл полученной суспензии липосом и инкубировали 4-6 часов при 37°C в CO<sub>2</sub>-инкубаторе. По окончании инкубации добавляли 4 мл ростовой среды с 1,2% сывороткой.

Через 48 часов инкубации при 37°C в CO<sub>2</sub>-инкубаторе клетки симали раствором Версена. Подсчет снятых клеток производили в камере Горяева.

Для трансфекции клеток в большом флаконе использовали 36 мкг плазмиды рННс-his. Всех остальных компонентов на большой флакон было взято в 6 раз больше, чем на малый.

Клетки промывали буфером PBS. Затем клеточные осадки использовали для получения клеточных лизатов, выделения ядерного и цитоплазматического экстрактов, и очищенного белка CTCF.

### **Получение клеточных экстрактов и обогащенных куриным CTCF белковых фракций**

Клетки собирали центрифугированием, осадок ресуспендировали в 2 мл 10 мМ HEPES–KOH, pH 7,9, 10% глицерин, 25 мМ KCl, 1 мМ EDTA, 0,15 мМ спермин, 0,5 мМ спермидин, 0,5 мМ AEBSF (Sigma, США), 0,5 мМ ДТТ (буфер А) с добавлением коктейля ингибиторов протеиназ (Sigma, США).

Для получения ядерного и цитоплазматического экстрактов клетки разрушали в гомогенизаторе Даунса, ядра осаждали центрифугированием в течение 5 мин при 3000 об/мин и 4°C. Супернатант (цитоплазматический экстракт) собирали и хранили при -70°C. Осадок ядер ресуспендировали в 500 мкл буфера А и добавляли KCl до концентрации 0,3 М. Ядра инкубировали при 0°C при постоянном помешивании в течение 30 мин, суспензию центрифугировали при 12000 об/мин в течение 10 мин.

Супернатант диализовали 2 часа против 100х объема 12 мМ HEPES–KOH,

pH 7,9, 12% глицерин, 60 mM KCl, 0,3 mM EDTA, 0,6 mM ДТТ, 0,5 mM AEBSF с добавлением коктейля ингибиторов протеиназ (Sigma, США), центрифугировали при 12000 об/мин в течение 10 мин и супернатант (ядерный экстракт) хранили при -70°C.

Лизат клеток для выделения белка получали путем обработки суспензии клеток в буфере LEW 1x (Macherey-Nagel, Германия): 50 mM NaHPO<sub>4</sub> pH 8,0, 300 mM NaCl, с добавлением AEBSF до 2 mM, коктейля ингибиторов протеиназ (Sigma, США) - 8 мкл/мл буфера, леупептина (Sigma, США) до 2 мкг/мл и апротинина (Sigma, США) до 8 мкг/мл. Обработку ультразвуком проводили на льду с использованием Ultrasonic processor model CV303 (Cole Parmer, США) при следующем режиме: длительность импульса 3 с, пауза 9,9 с, суммарное время обработки - 2 мин. Лизат наносили на колонку Protino Ni-IDA 150 Packed Column, колонку промывали и элюировали белок в соответствии с протоколом производителя (Macherey-Nagel, Германия). Далее элюат разводили в 10 раз буфером для нанесения на Q-сефарозу без NaCl: 20 mM HEPES-KOH, pH 7,9, 10% глицерин, 0,2 mM EDTA, 0,5 mM ДТТ, 2 mM AEBSF с добавлением коктейля ингибиторов протеиназ (Sigma, США) – 8 мкл/мл буфера, леупептина (Sigma, США) до 2 мкг/мл и апротинина (Sigma, США) до 8 мкг/мл, - и наносили на предварительно уравновешенную колонку с Q-сефарозой (Sigma, США) объемом 80 мкл. Колонка была уравновешена буфером следующего состава: 20 mM HEPES-KOH, pH 7,9, 10% глицерин, 0,2 mM EDTA, 0,5 mM ДТТ, 2 mM AEBSF с добавлением коктейля ингибиторов протеиназ (Sigma) – 8 мкл/мл буфера, леупептина (Sigma, США) до 2 мкг/мл и апротинина (Sigma, США) до 8 мкг/мл. Колонку промывали 1 мл того же буфера, содержащего дополнительно 40 mM NaCl. Для элюции использовался такой же буфер, со ступенчато повышающейся концентрацией NaCl от 200 до 450 mM. В процессе элюции на колонку наносилось по 2 порции буфера объемом 60 мкл для каждой концентрации NaCl.

Наличие STCF в полученных фракциях выявляли путем Вестерн-блот анализа и электрофореза в полиакриламидном геле с последующим его окрашиванием Кумасси синим.

Анализ функциональной активности CTCF в составе полученных фракций проводился методом ретардации в геле (EMSA – Electrophoretic Mobility Shift Assay) и методом дополнительной ретардации в геле.

### **Определение концентрации белка в ядерном и цитоплазматическом экстрактах**

Определение концентрации белка по методу Бредфорда (Bradford, 1976) с использованием реактива Bio-Rad Protein Assay Dye Reagent Concentrate, #5000006 (Bio-Rad) проводилось в соответствии с протоколом фирмы к этому реактиву на спектрофотометре Cary 50 Bio (Varian, США).

### **Электрофорез белков в полиакриламидном геле и Вестерн-блот анализ**

Препараты белка разделяли в 10% ПААГ в присутствии SDS [104; 105]. Перед нанесением на гель к белковым фракциям или суспензии клеток в деионизированной воде добавляли 10 мкл 2х буфера для нанесения (Bio-Rad), 2-меркаптоэтанол 98% (Bio-Rad) до 2% концентрации и деионизованная вода до объема 20 мкл. Затем инкубировали пробы в течение 5 минут при температуре 100°C. Электрофорез проводили при силе тока 10 мА. Гель окрашивали Кумасси синим, либо использовали для электропереноса белков на ПВДФ-мембрану (Immobilon P, Millipore) с помощью прибора Trans-Blot SD cell (Bio-Rad), в соответствии с протоколом производителя.

Электроперенос белковых зон, полученных в результате электрофореза, проводили в буфере для переноса (30 мМ Tris, 0.2 М глицина, 20% метанола) в течение 30 минут при силе тока 1 А. Для забивки неспецифической сорбции мембрану инкубировали в 5% растворе молока (not fat milk) в PBS с добавлением 0,1% Tween 20 (PBS-T) в течение 1 часа при комнатной температуре. Мембрану отмывали три раза по 5 минут в растворе для отмывки (0.1 % Tween-20, PBS) перед добавлением первичных и вторичных антител и непосредственно перед визуализацией сигнала. Инкубацию проводили как в случае первичных, так и вторичных антител в

растворе 5% обезжиренного молока в PBS-T в течение часа на горизонтальном шейкере (Red Rocker, San Francisco) при комнатной температуре. Первичные антитела, представляли собой антитела к глицеральдегид-3-фосфатдегидрогеназе (GAPDH, Santa Cruz, США, sc-47724, разведение 1:1000), антитела к N-концевому пептиду белка CTCF человека (Santa-Cruz, США, sc-5916, разведение 1:500), кроличьи поликлональные антитела к N-концевой части куриного CTCF (любезно предоставлены О.В. Яровой, ИБГ РАН, разведение 1:2000), антитела к гистидиновому тагу (Sigma H1029, разведение 1:2000). Вторичные антитела, конъюгированные с пероксидазой (Cell Signaling, США, разведение 1:1000, 7074, 7076 и Santa-Cruz, США, sc-2304, разведение 1:500). В зависимости от источника первичных антител использовали антитела к иммуноглобулинам кролика, мыши или козы. Затем мембрану промывали PBS-T и визуализировали сигнал при помощи набора Immun-Star HRP Chemiluminescent detection kit (Bio-Rad) и прибора VersaDoc MP4000 (Bio-Rad).

### **Получение фрагментов ДНК с помощью ПЦР**

F1-модуль лизоцимного сайленсера кур (F1) [7; 94], кор 5'HS4-инсулятора из  $\beta$ -глобинового локуса кур (5'HS4) [23; 11] и фрагмент промотора *c-тус* гена человека (MYC) [43], фрагмент энхансера  $\beta$ -глобинового локуса кур (Enh) [92] амплифицировали с использованием в качестве матрицы 20 нг геномной ДНК куриных клеток линии HD3 или клеток человека линии HeLa. Праймеры, которые использовали для амплификации: AAGCCCCCAGGGATGTAAT и AGGTGTCTGCAGGCTCAAAG – кор 5'HS4 инсулятора; CAGCACAGTTCTGGCTATGAAA и CTCAGCTGGGGTCAATAAGT – F1-модуль лизоцимного сайленсера; GGGATCGCGCTGAGTATAAA и GGATCTCCCTTCCCAGGAC – фрагмент промотора *c-тус* гена человека; GGGCAGGTTGCAGATAAACA и TAACCCCTCTCTTCCCTCA – энхансер из  $\beta$ -глобинового локуса кур. Режим амплификации: 94°C– 30 с, 60°C– 30 с, 72°C– 90 с в течение 26–29 циклов. Для амплификации фрагментов 1-13, отобранных в результате двумерного EMSA, в качестве матрицы использовали 0,1 нг плазмидной ДНК, выделенной из соответствующих клонов ранжированной библиотеки.

Приготовление реакционных смесей проводилось с использованием набора Encyclo Plus PCR kit (Евроген, Россия), в соответствии с рекомендациями производителя. ПЦР амплификацию проводили по следующей программе: 94°C – 30 с; 60°C – 30 с; 72°C – 90 с, 15 циклов, с использованием библиотечного праймера 27 (см. “Получение библиотеки фрагментов ДНК локуса FXVD5-COX7A 19 хромосомы человека”, раздел 2.1). ПЦР-продукт очищали электрофорезом в 1,2%-ном агарозном геле и использовали в качестве матриц для радиоактивного мечения с помощью ПЦР описанным далее методом.

### **Радиоактивное мечение фрагментов ДНК при помощи ПЦР и очистка меченой ДНК**

Для ПЦР-мечения смешивали:

1 мкл очищенного фрагмента ДНК (0,5-1,0 нг/мкл) или библиотеку фрагментов в том же количестве (1,0 нг/мкл), 1 мкл библиотечного праймера (10 пмоль/мкл), в случае мечения библиотеки или принадлежащих ей фрагментов 1-13, или 2 мкл соответствующих пар праймеров (по 5 пмоль/мкл каждого праймера), 0,6 мкл смеси dCTP, dGTP, dTTP (5 мМ каждого), 0,6 мкл 0,5 мМ dATP, 5 мкл alpha-<sup>32</sup>P-dATP (3000 Ки/ммоль), при амплификации библиотеки также добавлялись 4,5 мкл 5 М бетаина и 0,75 мкл ДМСО, для приготовления реакционных смесей использовался набор Encyclo Plus PCR kit (Евроген, Россия) остальные компоненты реакционных смесей добавлялись в соответствии с рекомендациями производителя, деионизированная вода добавлялась до 15 мкл. Для мечения фрагментов F1, HS4, Muc, Enh использовались те же праймеры, что и для их амплификации на матрице геномной ДНК (см. “Получение фрагментов ДНК с помощью ПЦР”, раздел 2.1).

Амплификацию (15 циклов) проводили по следующей программе: 94°C - 30с, 55(в случае библиотеки) либо 60°C – 30с, 72°C - 90 с.

В случае мечения отдельных фрагментов ДНК, их очищали в 10% полиакриламидном геле с 0,17% N,N-метилен-бис-акриламидом (сшивка

30:1) в буфере: 45 мМ трис-борат, 45 мМ борная кислота, 1 мМ ЭДТА, pH 8,0 (0,5xTBE). После проведения электрофореза и радиоавтографии меченой ДНК часть геля, соответствующую искомому фрагменту ДНК, вырезали из геля, измельчали и элюировали ДНК в 150 мкл буфера для элюции (0,5 М ацетат аммония, 1% SDS, 1 мМ ЭДТА) при 37<sup>0</sup>С в течение ночи. Затем отфильтровывали частицы геля через стекловату, добавляли к раствору 2 объема 96% этанола и выдерживали 1 час при –70<sup>0</sup>С. После центрифугирования (14500 об/мин, 10 мин), удаляли супернатант, осадок промывали 250 мкл 75% этанола, высушивали и растворяли в 10 мкл Н<sub>2</sub>О.

Библиотеку меченых фрагментов ДНК очищали с помощью колонок QIAquick PCR Purification Kit (QIAGEN) согласно рекомендациям производителя.

### **Сдвиг и дополнительный сдвиг электрофоретической подвижности в полиакриламидном геле (Electrophoretic mobility shift assay, EMSA)**

Эксперименты по сдвигу электрофоретической подвижности проводили по ранее описанной методике [47].

В качестве CTCF-связывающих зондов использовали следующие радиоактивно меченые фрагменты ДНК: F1, 5'HS4, MYC, в качестве отрицательного контроля – не связывающий CTCF фрагмент Enh, фрагменты 1-13, отобранные в результате двумерного EMSA.

Реакционная смесь, в которой проводилось образование ДНК-белковых комплексов, объемом 20 мкл содержала 2 мкл ~ 5нг (30000–50000 имп/мин) меченого фрагмента ДНК, 1 мкл (1 мкг) альтернирующего двунитевого гетерополимера poly(dI-dC), 2 мкл (1,2 или 2,4 мкг суммарного белка) ядерного экстракта или цитоплазматического экстракта клеток COS-1 (трансфицированных рННс-his или интактных) или 2 мкл фракции, обогащенной белком CTCF с N-концевым гистидиновым тагом, 12 мМ HEPES–КОН, pH 7,9, 12 % глицерин, 60 мМ КСl, 0,3 мМ EDTA, 0,6 мМ дитиотрейтол (буфер В). Белковый образец добавляли в последнюю очередь. После добавления всех реагентов, смесь инкубировали в течение 20 минут при комнатной температуре. Буфер для нанесения в образцы не добавляли.

Реакционную смесь для дополнительного сдвига электрофоретической подвижности [100] в присутствии антител к CTCF готовили аналогично, за исключением того, что в смесь вносили 3 мкг антител, специфичных к гистидиновому тагу (Sigma H1029), либо 4,5 мкг кроличьих поликлональных антител, специфичных к фрагменту CTCF (аминокислотные остатки 86-233), любезно предоставленных О.В. Яровой, ИБГ РАН.

По окончании инкубации, смесь наносили на 10%, 7,5% или 5% полиакриламидный гель (сшивка 30:1) приготовленный на буфере 0,5XTBE. В свободные дорожки геля наносили красители бромфеноловый синий и ксиленцианол для контроля миграции во время прохождения электрофореза. Электрофорез проводили при силе тока в 12-15 мА. По прохождении ксиленцианолом 2/3 (10-12 см) геля электрофорез останавливали, гель закрывали лавсановой пленкой и радиоавтографировали в течение 16-40 часов при 4°C.

### **Получение библиотеки фрагментов ДНК локуса FXVD5-COX7A 19 хромосомы человека**

ДНК клона CH261-75C12 искусственной бактериальной хромосомы (BAC, получена из CHORI BACPAC Resource Center, <https://bacpac.chori.org>), содержащую вставку  $\alpha$ -глобинового локуса кур длиной 227366 п.н., очищали при помощи набора Plasmid Midi Kit (Qiagen) и обрабатывали нуклеазой Plasmid-Safe ATP-Dependent DNase (Epicentre) в соответствии с рекомендациями производителей.

2 мкг ДНК искусственной бактериальной хромосомы исчерпывающе расщепляли 10 ед. рестриктазы *Sau3AI* или *Csp6I* в буфере, рекомендованном производителями. Полноту расщепления контролировали электрофорезом в 1% агарозном геле. ДНК экстрагировали смесью фенола и хлороформа (1:1), осаждали этанолом и растворяли в 30 мкл буфера TE (10 мМ Трис-HCl pH 8,0, 1 мМ ЭДТА).

Фрагменты ДНК, полученные после расщепления каждой из рестриктаз лигировали с праймером и соответствующим адаптором в следующей смеси:

10 мкл раствора фрагментов ДНК, 2 мкл 10х буфера для T4 ДНК лигазы, с 10mM АТФ (Promega, США), по 0,5 мкл (250 пмоль) соответствующего адаптора и праймера (см. ниже), 1,6 мкл (8 ед.) ДНК-лигазы фага T4, в конечном объеме 20 мкл. Реакцию проводили в течение ночи при 13°C для *Sau3AI* и 4°C для *Csp6I* гидролизатов.

Структуры праймера и адапторов приведены ниже:

5' GATCTGTTCATGG3'                      адаптор для *Sau3A*  
*Sau3A*

5' GТАСТGTTCATGG3'                      адаптор для *Csp6I*  
*Csp6-I*

5' АСТGAGСТСGAGTATССАТGААСА3'      универсальный библиотечный праймер (праймер 27).

Полученные суббиблиотеки амплифицировали при помощи ПЦР (20–25 циклов) с использованием библиотечного праймера и набора Encyclo Plus PCR kit («Евроген»), в соответствии с рекомендациями производителя, в присутствии 1.5 М бетаина и 5% диметилсульфоксида по следующей схеме: 95°C, 30 с; 55°C, 30 с; 72°C, 90 с.

Суббиблиотеки очищали с помощью набора QIAquick PCR Purification Kit (Qiagen) и объединяли в равных количествах.

Полученную библиотеку фрагментов ДНК радиоактивно метили (см. “Радиоактивное мечение фрагментов ДНК при помощи ПЦР и очистка меченой ДНК”, раздел 2.2) и очищали с помощью колонок QIAquick PCR Purification Kit (QIAGEN).

### **Селекция фрагментов, содержащих участки, способные связываться с транскрипционным фактором СТСF (метод двумерного EMSA)**

2,5 мкл (~20 нг) смеси радиоактивно меченых фрагментов ДНК (300000-500000 имп/мин) инкубировали, в случае первого раунда двумерного EMSA, с 10 мкл (около 0,5 пмоль), в случае второго раунда двумерного EMSA, с 1 мкл (около 0,05 пмоль), белка СТСF, с N-концевым гистидиновым тагом, в присутствии 1 мкл (1мкг) poly(dI-dC) в течение 20 минут при комнатной температуре и разделяли в

5% полиакриламидном геле (сшивка 30:1, гель размером 15×15 см, толщина 0,7 мм), приготовленном в 0,5×TBE-буфере. Как и при EMSA, в свободные дорожки геля наносили красители бромфеноловый синий и ксиленцианол для контроля миграции во время прохождения электрофореза. Электрофорез проводили при силе тока в 12-15 мА. По прохождении ксиленцианолом 2/3 (10-12 см) геля электрофорез останавливали, гель закрывали лавсановой пленкой и проводили радиоавтографию в течение 1 часа при температуре 4° С.

После радиоавтографии фрагмент геля, соответствующий полученной на радиоавтографе полосе разделенных в геле фрагментов ДНК и ДНК-белковых комплексов, размером ~ 0,3 см на 12-14 см, вырезали из геля и инкубировали в 40 мл буфера, содержащего 0,1×TBE, 0,5% SDS, в течение 20-30 минут при 50°С. Полоску геля размещали горизонтально на поверхности 5% полиакриламидного геля (сшивка 30:1), размером 20×20 см и толщиной 1 мм, содержащем 0,5×TBE-буфер и 0,1% SDS, для разделения фрагментов ДНК в направлении, перпендикулярном первому. Для контроля миграции во время прохождения электрофореза наносили красители бромфеноловый синий и ксиленцианол. Электрофорез во втором направлении проводили при силе тока 15мА. При прохождении ксиленцианола 1/2 геля (8-9 см) ток останавливали, гель закрывали лавсановой пленкой и проводили радиоавтографию в течение 16-40 часов при 4°С. Фрагменты, связывающиеся с CTCF, элюировали и подбирали оптимальное количество циклов в ПЦР-реакции для второго раунда селекции.

### **ПЦР-анализ полученной библиотеки фрагментов ДНК**

Для анализа полученной библиотеки использовали ПЦР-амплификацию с отбором реакционной смеси через 3 цикла.

В качестве матрицы использовалась библиотека рестриктных фрагментов  $\alpha$ -глобинового локуса кур, с которой проводилось связывание белка в первом раунде двумерного EMSA, разведенная в 16 раз, библиотека,

амплифицированная после первого раунда двумерного EMSA, разведенная в 128 раз и библиотека, амплифицированная после второго раунда двумерного EMSA, разведенная в 256 раз.

Для амплификации фрагмента M9 использовали праймеры CCTGCGTTTTAGCTGATTGG и TCAGGAAGAAAGAATGGGAAA, для амплификации фрагмента CDS – TCCCAGCACCTCGCAGTGCA и GCACAAGGCTCAAAGGTGAGACA. Амплификация фрагмента гена  $\alpha$ -глобина D (фрагмент HBAD) проводилась с праймерами, которые использовались для ПЦР в реальном времени (см. Таблицу 1).

Приготовление реакционных смесей проводилось с использованием набора Encyclo Plus PCR kit (Евроген, Россия), в соответствии с рекомендациями производителя. В реакционные смеси вносили по 1 мкл матрицы и по 1 мкл каждого праймера (10 мкМ) из соответствующей пары.

Температурный режим амплификации:

94<sup>0</sup>С-20 сек, 60<sup>0</sup>С-30 сек, 72<sup>0</sup>С-60 сек, 24-33 цикла

### **Клонирование библиотеки в вектор pGEM-T**

ПЦР-амплифицированные фрагменты ДНК после 2-го раунда селекции клонировали в плазмидный вектор pGEM-T (Promega) в соответствии с рекомендациями фирмы-изготовителя. Лигирование проводили при молярном соотношении вектора и вставки 1:1. Инкубацию с ДНК-лигазой проводили при 14<sup>0</sup>С в течение ночи. Далее проводили трансформацию клеток *E.coli* DH-5 $\alpha$  продуктом реакции.

### **Ранжирование библиотеки**

230 белых колоний были отобраны и ранжированы в 96-луночные планшеты, содержавшие по 200 мкл среды LB со 100 мкг/мл ампициллина в каждой лунке. Клетки подращивали в течение 3-5 час при 37<sup>0</sup>С, добавляли в лунки стерильный глицерин до 15% и хранили при –70<sup>0</sup>С. Наличие и размер вставок проверяли путем ПЦР.

## **Секвенирование и картирование нуклеотидных последовательностей**

Определение первичной структуры плазмидных клонов проводили на автоматическом секвенаторе ABI Prism model 3100 с использованием набора для сиквенса Big Dye Terminator Kit (ABI) в центре коллективного пользования «Геном».

Полученные последовательности сравнивали с имеющимися последовательностями в международных базах данных нуклеотидных последовательностей. Сравнение проводили при помощи программы BLAST с сервера Национального института здравоохранения США ([www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/)). Для картирования использовали UCSC Genome Browser: <http://genome.ucsc.edu/> [84; 85; 82].

## **Культивирование клеток**

Клетки линий COS-1 (ATCC CRL-1650, клетки почки африканской зеленой мартышки), HD3 (клон А6 линии LSCC куриных эритробластов), DT40 (CRL-2111, линия лимфоидных клеток кур) культивировали в условиях, рекомендованных ATCC (American Type Culture Collection) для этих клеточных линий.

Клетки COS-1 инкубировали при 37°C в атмосфере 5% CO<sub>2</sub> с частотой пассирования три раза в неделю с разведением 1:4. При пассировании использовали обработку 0,05% трипсином в PBS. Состав культуральной среды: DMEM/F12 HAM (1:1) (Gibco, США), 10% FCS (Gibco, США). Клетки линии HD3 (клон А6 линии LSCC) культивировали в среде DMEM/F12 HAM (1:1), содержащей 8% фетальной бычьей сыворотки, 2% куриной сыворотки (Gibco, США). Куриные лимфоидные клетки линии DT40 (CRL-2111, ATCC) культивировали в той же среде с добавлением до 50 мкМ β-меркаптоэтанола.

Терминальную эритроидную дифференцировку клеток HD3 индуцировали инкубацией клеток в течение 12 ч в присутствии 20 мкМ ингибитора протеинкиназ iso-H-7 (1-(5-изохинолинсульфонил)-3-метилпиперазиндигидрохлорид, Sigma, США) при pH 8.0 и 42°C в 100%

воздушной атмосфере как описано ранее [130]. Для контроля дифференцировки использовали окрашивание клеток бензидином [133]. К 25 мкл 0.4% (w/v) раствора бензидина (Sigma, США) в 4% уксусной кислоте добавляли 1 мкл 30% раствора H<sub>2</sub>O<sub>2</sub>, смешивали с 25 мкл суспензии клеток, выдерживали в течение 10 мин и с помощью светового микроскопа выявляли бензидин-положительные клетки, окрашенные в темно-синий цвет. Доля гемоглобинсодержащих (бензидин-положительных) клеток составляла 21% после 12 ч инкубации. При этих условиях уровень транскрипции гена  $\alpha$ -глобина близок к максимальному, но продолжает возрастать [50].

### **Иммунопреципитация хроматина**

Метод иммунопреципитации выполняли согласно методике, описанной в работе Орландо [134]. Для иммунопреципитации хроматина в ходе ChIP-seq в данную методику были внесены изменения в соответствии с рекомендациями фирмы производителя используемых для иммунопреципитации протеин-G магнитных частиц Life Technologies. Для иммунопреципитации хроматина клетки на стадии логарифмического роста ( $\sim 3 \times 10^7$ ) фиксировали 1 % раствором формальдегида в течение 8 минут при комнатной температуре. В случае ChIP-seq клеток использовалось в три раза больше, пропорционально были увеличены объемы всех используемых реактивов. Начиная со стадии лизиса до стадии отмывки, все процедуры проводились при 4<sup>0</sup>С в присутствии ингибиторов протеиназ (Sigma). Клетки собирали центрифугированием при 4000 g, лизировали в буфере (1% SDS, 10 мМ ЭДТА, 50 мМ Tris-HCl pH 8.1), и обрабатывали ультразвуком на приборе Ultra Sonic Processor (США) в режиме 30 сетов по 3 секунды с интервалом между импульсами по 10 секунд при амплитуде 30%. В случае ChIP-seq проводились 40 сетов ультразвуковой обработки. Клеточный дебрис удаляли центрифугированием при 4<sup>0</sup>С 10 минут на 13000g. 1/10 либо 1/30(в случае ChIP-seq) часть от всего объема отбирали для использования в качестве контроля (input chromatin, или Input). К супернатанту добавляли IP-буфер (1:10) и инкубировали со 120 мкл протеин-A-агарозы (Invitrogen, США), содержащей 40 мкг ДНК спермы лосося и 1 мг/мл БСА (бычьего сывороточного альбумина),

содержащей 20 мкг ДНК спермы лосося в течение 1 часа при 4°C при постоянном перемешивании. В случае ChIP-seq использовали 160 мкл протеин-G магнитных бусин (Life Technologies, США), содержащих 1 мг/мл БСА. Осаждали агарозу центрифугированием на максимальной скорости в течение 10 секунд; и далее, к надосадочной жидкости добавляли 10 мкг кроличьих поликлональных антител к белку CTCF (любезно предоставленных О.В. Яровой, ИБГ РАН) либо 10 мкг кроличьих поликлональных антител к растительному белку тауматину (для получения отрицательного контроля) и инкубировали в течение ночи при 4°C с перемешиванием. Добавляли 120 мкл протеин-A-агарозы, содержащей 40 мкг ДНК спермы лосося и 1 мг/мл БСА в течение 1 часа при 4°C при постоянном перемешивании. При проведении ChIP-seq, использовали 160 мкл протеин-G магнитных бусин, содержащих 1 мг/мл БСА. Связанные с протеин-A агарозой ДНК-белковые комплексы собирали центрифугированием, осадок последовательно промывали в течение 5 минут по 1 мл каждого раствора в следующем порядке:

- 1) 1x 1мл Low salt buffer (0.1% SDS, 1% Triton X-10, 2мМ ЭДТА, 20 мМ Tris-HCl pH8.1, 150мМ NaCl)
- 2) 1x 1мл High salt buffer (0.1% SDS, 1% Triton X-10, 2мМ ЭДТА, 20 мМ Tris-HCl pH8.1, 500мМ NaCl)
- 3) 1x 1мл LiCl buffer (0.25мМ LiCl, 1% NP-40, 1% дезоксихолат натрия, 1мМ ЭДТА, 10мМ Tris-HCl, pH 8.1)
- 4) 2x 1мл TE pH 8.0

Элюировали комплекс 250 мкл буфера элюции (1% SDS, 0.1М NaHCO<sub>3</sub>) в течение 20 минут при постоянном вращении, повторяли элюцию и объединяли элюаты. Для удаления «кросс-сшивок» в элюат и хроматин, обработанный ультразвуком (input chromatin), добавляли 20 мкл 5М NaCl и инкубировали в течение ночи при 65°C, далее добавляли 10мкл 0.5М ЭДТА, 20 мкл 1М Tris-HCl pH 6.5 и 2 мкл протеиназы К (10 мг/мл) и продолжали инкубировать ещё 1 час при 45°C. В ходе ChIP-seq перед добавлением протеиназы К к образцам добавляли 2 мкл РНКазы А (10 мг/мл) и инкубировали 1 час при 56°C. Полученную ДНК

экстрагировали смесью фенол/хлороформ 1:1 и осаждали спиртом в присутствии 2 мкг гликогена (Thermo Fisher Scientific, США). ДНК собирали центрифугированием, перерастворяли в стерильной воде и использовали для дальнейшего ПЦР-анализа. В случае ChIP-seq вместо экстракции смесью фенола и хлороформа проводилась очистка образцов на колонках QIAquick Gel Extraction Kit (Qiagen, США) в соответствии с рекомендациями производителя.

### **Количественная ПЦР в реальном времени**

В качестве матрицы для ПЦР использовали по 1 мкл геномной ДНК, полученной в результате иммунопреципитации хроматина. Образец Input был разведен в 10 раз. Также реакционная смесь содержала по 1 пмоль либо 0,5 пмоль каждого из соответствующих фрагменту праймеров (таблица 1). Реакцию проводили с использованием готовой смеси для ПЦР в реальном времени, содержащей флуоресцентный краситель, qPCRmix-HS SYBR (Евроген, Россия) в соответствии с рекомендациями производителя. Каждая реакция проводилась в трех повторах.

ПЦР в реальном времени проводили на амплификаторе Mx3000P (Stratagene, США) в следующем режиме: 95°C–7 мин, далее 38 циклов 95°C–30 с, (температура отжига для соответствующей пары праймеров)–30 с, 72°C–50 с, с дальнейшей записью кривой плавления продукта. Анализ флуоресцентного сигнала проводили с помощью программного обеспечения, прилагаемого к прибору. Для определения эффективности ПЦР использовали программу LinRegPCR.

**Таблица 1. Праймеры для ПЦР в реальном времени.**

Аmplи- фицируемый фрагмент ДНК	Прямой праймер	Обратный праймер	Темпе- ратура отжига, °C.
F1	CAGCACAGTTCTGGCTATGAAA	CCTCAGCTGGGGTCAATAAGT	61
Мус	AAGCAGCGAGGAGCGCCCTTT	TACTACAAGGAGAGGTTCGGAAGT	65
Enh	GGGCAGGTTGCAGATAAACA	TAACCCCTCTCTCCCTCA	61
HBAD	CCCAGACCAAGACCTACTTCC	GCTGAGGTTGTCCACGTTCTT	61
1	АТААСТТGGCAGCAAАСТAGCA	ТТТGGAAAGTGCTGTGGGTAAAG	61
2 (пара праймеров 1)	ТТСТАСАСТТGTCCCTCCTTTTCA	ССТАТТТТGTGGCTGCATTCTTC	63
2 (пара праймеров 2)	TGTTCCACACCCACSTTATCTCT	AACCGTGGTCAGGATGACTTTAC	63
3	GGAGCTCAGCAGGCAGAAACTA	GCTAAGGCAAAGGCTCTGTTGT	63
4	СТСТGCATTGCTGTGTGTGTTTT	ATGGTGGTTATCTCAGGGGTTTT	63
5	GGTACGTTCTCAGTGCCCAAAC	CCACCTGCAGACCTAACCTGTC	63
6	CAGCTCTTCTGGCTCATTTGTCT	ATCTCCCTTTCAGTCCCTTCTC	63
7	ТТТCACCСCAGAAGTTCATGCT	CCCAGTGTGGAAGCCATTTATC	63
8, 5d1, 10d2	GCTCTTCCTCACCCAGGTTTCT	CATCCAGCCCTCTCCAAACA	63
9	CATGGGCAGCAAACACACAG	TCCATTTCCAGCGGTTCTTATC	63
10 (пара праймеров 1)	GTGTGTGCATGGAAAGAAAAGTG	ACACTGTCCCTGCTGAGTCCTAC	61
10 (пара праймеров 2)	AGGTAGGACTCAGCAGGGACAG	GGGACAAGTAGCTGGGACAAAA	61
11	СТGGAGATACCCATGGCAGAAC	ТТТGTGGCCAACGTCAAACCTAC	63
12	GGTTTGCCTTCTTGCTCTG	ATGCCCATCTCACTTGCTCT	61
13	CGTACCAGCACCAGACAAACAG	TCGACTGTTGAAGGAGGCATAA	63
5d2	АТАGGAСТТCCCTGCTTCCATCT	GTGGAGTGTTGTGGTCTTCTCC	61
5d3	GTGAGGAGAGGGCGAAGTTTATT	GCTCCCTGAGCTCCTCACCT	63
10d1	TGTGGTCATCCATGTCCTCAATC	GGAAGCTTTTGGCCAAGGAGAA	65
10d3	TGACCCATCTTGCAATGGATACT	GTTTGGGAACCTCTCTCCATCC	63

### **Масштабное секвенирование образцов ДНК, полученных в результате иммунопреципитации хроматина**

Массовому секвенированию были подвергнуты образцы ДНК, полученные в результате иммунопреципитации хроматина с антителами к CTCF клеток линии HD3, интактных (образец HD3 CTCF) и индуцированных к дифференцировке (образец HD3-ind CTCF), линии DT40 (образец DT40 CTCF). Также было проведено массивное секвенирование образца тотальной ДНК, выделенной из порции клеток HD3 на стадии ультразвуковой дезинтеграции, который далее использовался для иммунопреципитации хроматина (Образец HD3 Input).

Масштабное секвенирование проводилось фирмой Евроген (Россия, [www.evrogen.ru](http://www.evrogen.ru)) с помощью прибора Illumina HiSeq (США), для секвенирования

использовалось по 10 нг каждого образца ДНК. В результате для каждого образца ДНК было получено 41-50 млн. прочтений. Также фирмой Евроген была проведена начальная биоинформатическая обработка полученных данных.

### **Картирование прочтений, полученных в результате ChIP-seq**

Картирование прочтений, проводилось на геном *Gallus gallus* сборки 2011 года (galGal4) с использованием программы Bowtie2 [107] в составе платформы BioUML [95; 96]. При этом использовался файл формата FASTQ, содержащий нуклеотидные последовательности, полученные при секвенировании и информацию об их качестве.

Далее был применен фильтр качества полученных прочтений: значения величины MAPQ, характеризующей качество прочтений, не менее 10.

Прочтения, полученные и картированные на геном *Gallus gallus* сборки 2006 года, Martin D и соавторами [122], были извлечены из файла формата elandmulti в виде файла формата fasta с помощью компонентов платформы Galaxy [1; 95; 96] в составе BioUML. Далее по причине низкого качества определения последних нуклеотидных остатков нуклеотидные последовательности были сокращены до первых 40 нуклеотидных остатков из 50. Картирование прочтений, полученных Martin D и соавторами, далее проводилось программой Bowtie2 [107]. Затем был применен фильтр качества прочтений: значения MAPQ не менее 10.

### **Выявление пиков**

Выявление пиков было проведено с помощью программы MACS1.4.2 [196] в составе платформы BioUML [95; 96]. Параметры, отличные от параметров по умолчанию для образцов, полученных нашим коллективом в результате ChIP-seq-эксперимента: effective genome size 1003053316, tag size 100. Параметры, отличные от параметров по умолчанию при обработке данных, полученных Martin D и соавторами [122]: effective genome size 1003053316, tag size 40.

### **Анализ обогащения нуклеотидными мотивами**

Выявление мотивов, которыми обогащены нуклеотидные последовательности, соответствующие пикам, проводилось с помощью программы MEME-chip [117]. При этом использовались центральные 500 либо 800 н.о. каждого пика (средняя длина пиков, выявленных при обработке данных Martin и соавторов [122], и средняя длина пиков, выявленных в результате обработки данных, полученных нашей лабораторией, соответственно). Параметры, отличные от параметров по умолчанию: MEME motif count: 5, MEME motif sites: 24-600.

### **Оценка тканеспецифичности связывания CTCF в области пиков**

Сравнение занятости белком CTCF его сайтов связывания в разных типах клеток проводилось с помощью программы MAnorm [162]. При этом использовались файлы с прочтениями, картированными Bowtie2, переведенные в формат BED с помощью компонентов платформы Galaxy [1; 147]. Также использовались файлы с координатами пиков, полученные с помощью программы MACS 1.4.2, и значения сдвига координат прочтений (d), рассчитанные этой программой.

### 3. Результаты и обсуждение

#### *3.1. Выявление фрагментов $\alpha$ -глобинового локуса кур, способных связывать белок CTCF, методом двумерного EMSA*

В основе метода двумерного EMSA лежит изменение электрофоретической подвижности фрагментов ДНК в ПААГ после их связывания с белком. После инкубации с белком, радиоактивно меченые фрагменты ДНК наносятся на ПААГ и разделяются в неденатурирующих условиях. В результате не связавшиеся с белком фрагменты ДНК не меняют своей подвижности. Те же фрагменты ДНК, которые образовали комплексы с белком, движутся медленнее и оказываются выше в дорожке геля, нежели другие фрагменты такой же длины. Затем дорожка геля, содержащая комплекс фрагментов ДНК с белком вырезается и инкубируется в буфере, содержащем SDS, в течение 30 минут при 50°C. После разрушения ДНК-белкового комплекса полоска геля помещается на поверхность ПААГ второго направления для разделения фрагментов ДНК в направлении, перпендикулярном первому. Фрагменты ДНК электрофоретически разделяются во втором направлении, но уже в денатурирующих условиях. При этом электрофоретическая подвижность фрагментов ДНК соответствует их длине. На радиоавтографе такого геля регистрируется диагональ, которую образует основная часть фрагментов ДНК, не связавшаяся с белком, также видно дополнительную диагональ, которая идет под основной. Ее образуют фрагменты ДНК, взаимодействующие с белком при инкубации. Дополнительная диагональ вырезается из геля, из нее выделяется ДНК. ДНК ПЦР-амплифицируется и с ней проводится второй раунд двумерного EMSA, который обеспечивает дополнительную специфичность отбора. Затем проводится определение нуклеотидных последовательностей фрагментов ДНК, полученных после второго раунда двумерного EMSA, а так же определяется их расположение на карте локуса. [174].

Мы использовали метод двумерного EMSA для выявления CTCF-связывающих участков  $\alpha$ -глобинового локуса кур. Перед проведением двумерного EMSA необходимо было получить очищенный белок CTCF, специфично узнающий CTCF-связывающие фрагменты ДНК, либо белковую фракцию, в составе которой только белок CTCF специфично взаимодействует с ДНК.

### *3.1.1. Получение и функциональный анализ фракции, обогащенной куриным белком CTCF.*

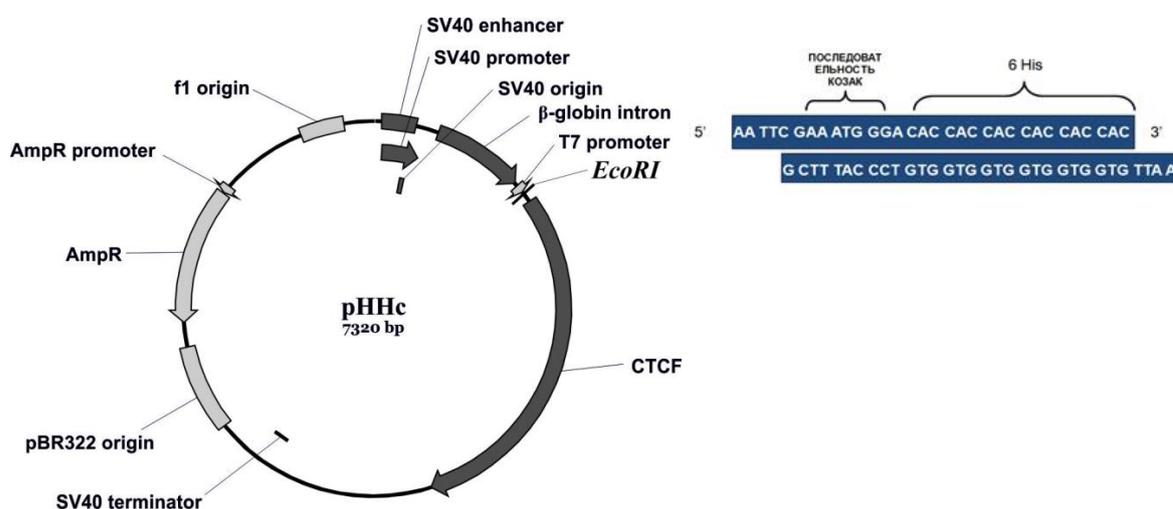
Существуют многочисленные данные о высокой консервативности структуры и функций белка CTCF среди различных видов позвоночных [122; 43; 136]. Белок CTCF подвергается ряду функционально важных посттрансляционных модификаций [38; 89; 193]. Представляется более вероятным, что белок CTCF, образующийся в клетках позвоночных содержит необходимые посттрансляционные модификации и имеет характерную для нативного CTCF укладку полипептидной цепи, чем полученный в каких-либо других системах экспрессии. Для экспрессии гена куриного CTCF и последующего получения обогащенной куриным CTCF фракции из клеточного лизата была выбрана система экспрессии с использованием клеток млекопитающих.

**Получение конструкции для экспрессии гена куриного CTCF, с N-концевым гистидиновым тагом в клетках позвоночных.** Получение куриного белка CTCF осуществляли с использованием гетерологичной системы экспрессии в клетках млекопитающих COS-1. Выделение CTCF планировалось проводить методом никель-аффинной хроматографии. Такой способ очистки требует наличия у белковой молекулы аминокислотной последовательности из шести остатков гистидина (H6) (гистидиновый таг).

Конструкция рННс-his для повышенной экспрессии гена куриного CTCF с дополнительной N-концевой аминокислотной последовательностью из шести остатков гистидина была получена путем модификации плазмиды рННс [91] (рисунок 4). Плазмида рННс была любезно предоставлена коллегами из Лаборатории структурно-функциональной организации хромосом Института биологии гена РАН. Эта конструкция представляла собой коммерческий вектор

pSG5 (Stratagene), со вставкой кДНК куриного CTCF. Конструкции рННс и рННс-his, будучи производными вектора pSG5, содержат промотор, энхансер и точку начала репликации вируса SV40, благодаря чему обеспечивают повышенную экспрессию целевого гена в клетках, содержащих Т-большой антиген вируса SV40. Одной из таких клеточных линий является линия COS-1, которую мы использовали в этой работе.

При получении рННс-his в начало открытой рамки считывания гена *CTCF* плазмиды рННс была добавлена последовательность, кодирующая гистидиновый таг. Клонирование этой последовательности осуществлялось по сайту рестрикции *EcoRI*, расположенному между промотором SV40 и ОРС гена *CTCF*. Клонированный фрагмент ДНК содержал инициаторный кодон в составе последовательности Козак, и кодировал 6 остатков аминокислоты гистидина.



**Рисунок 4.** Получение плазмиды рННс-his, предназначенной для экспрессии полноразмерного гена куриного белка CTCF с N-концевым гистидиновым тагом. А – Схема плазмиды рННс, содержащего ген полноразмерного куриного белка CTCF. Плазмида получена на основе вектора SG5 [91]. SV40 enhancer – энхансер вируса SV40, SV40 promoter – промотор вируса SV40, SV40 origin – точка начала репликации вируса SV40,  $\beta$ -globin intron – интрон  $\beta$ -глобинового гена кролика, T7 promoter – промотор бактериофага T7, *EcoRI* – сайт, узнаваемый эндонуклеазой рестрикции *EcoRI*, CTCF – ген белка CTCF, SV40 terminator – терминатор транскрипции вируса SV40, pBR322 origin – точка начала репликации из плазмиды pBR322, AmpR – ген  $\beta$ -

лактамазы, экспрессия которого обеспечивает устойчивость к ампициллину, *AmpR promoter* – промотор гена  $\beta$ -лактамазы, *f1 origin* – точка начала репликации бактериофага *f1*. Б – Олигонуклеотидный дуплекс, клонированный по сайту *EcoRI* в вектор рННс. Дуплекс содержит последовательность Козак, а также участок, кодирующий 6 последовательных гистидиновых остатков (гистидиновый таг).

Был проведен Вестерн-блот анализ клеточных лизатов, с антителами к белку СТСФ и антителами к гистидиновому тагу. Содержание белка СТСФ в клетках COS-1, трансфицированных плазмидой рННс-his, значительно превышало его содержание в интактных клетках (рисунок 5). Содержание СТСФ в трансфицированных клетках превышало его содержание в интактных как минимум в 20 раз.

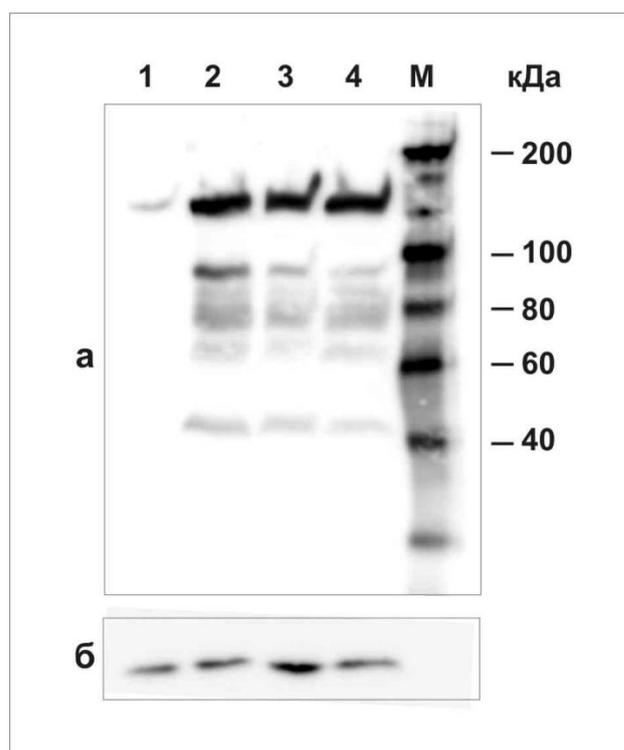


Рисунок 5. Вестерн-блот анализ лизатов клеток COS-1, трансфицированных плазмидой рННс-his. Дорожки: 1 - нетрансфицированные клетки; 2 - 6 мкг плазмиды, 15 мкл липофектамина; 3 - 3 мкг плазмиды, 20 мкл липофектамина; 4 - 3 мкг плазмиды, 10 мкл липофектамина. (а) - окраска антителами к куриному СТСФ, (б) - окраска антителами к глицеральдегид-3-фосфатдегидрогеназе. М - маркер молекулярных масс.

**Анализ способности белка CTCF с N-концевым гистидиновым тагом взаимодействовать с CTCF-связывающими последовательностями.** Белок CTCF, с N-концевым гистидиновым тагом, может быть использован для отбора CTCF-связывающих фрагментов в ходе двумерного EMSA, только если он сохранил способность специфично узнавать CTCF-связывающие нуклеотидные последовательности. Для того, чтобы проверить, узнает ли модифицированный CTCF те же последовательности, что и нативный, мы использовали методы ретардации и дополнительной ретардации в геле [47; 100].

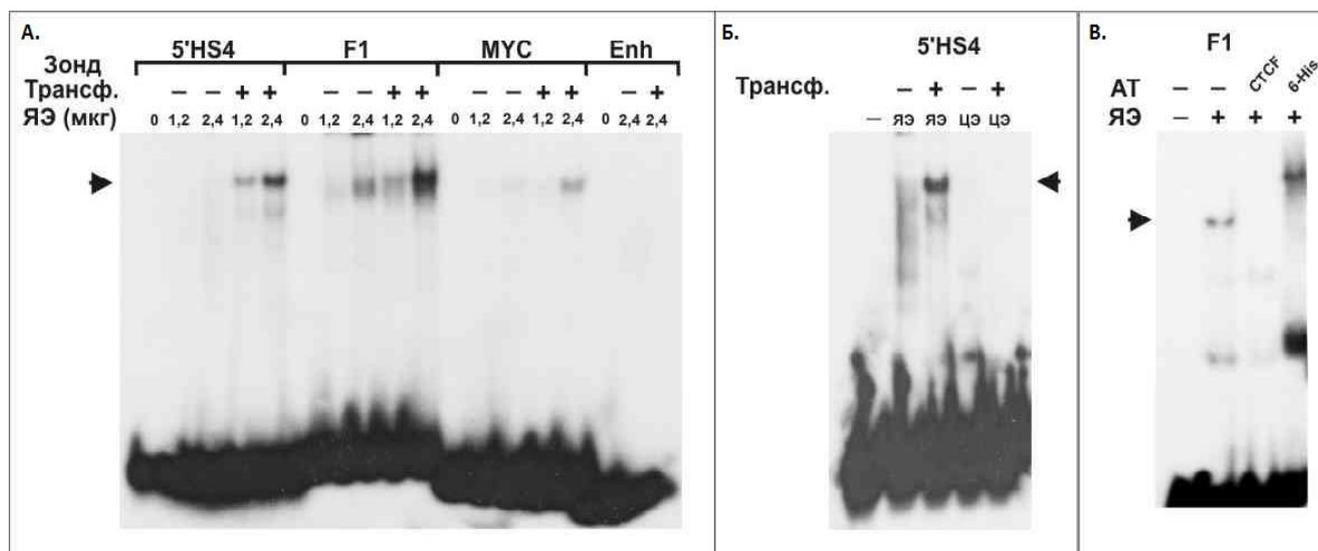
Радиоактивно меченые фрагменты ДНК, заведомо связывающие CTCF, инкубировались с ядерными экстрактами клеток COS-1, трансфицированных рННС-his, и интактных. В качестве CTCF-связывающих последовательностей нами были выбраны F1-модуль лизоцимного сайленсера кур [7; 94], коровый участок 5'HS4-инсулятора из  $\beta$ -глобинового локуса кур [23; 11] и фрагмент промотора *c-myc* гена человека [114; 43]. В качестве отрицательного контроля был взят участок ДНК, заведомо не связывающий белок CTCF, – это фрагмент энхансера из  $\beta$ -глобинового локуса кур [92]. В случае инкубации с ядерным экстрактом из трансфицированных клеток CTCF-связывающих фрагментов ДНК с последующим EMSA-анализом на электрофореграмме наблюдалось появление полос торможения, что соответствовало образованию ДНК-белкового комплекса. Образование ДНК-белковых комплексов при инкубации этих фрагментов с ядерным экстрактом из интактных клеток COS-1 не было выявлено (рисунок 6А). Таким образом, CTCF-связывающие последовательности взаимодействовали с белком, содержание которого в ядрах клеток увеличилось в результате трансфекции. Таким белком является CTCF.

Для энхансера из куриного  $\beta$ -глобинового локуса наблюдалось отсутствие дополнительной полосы во всех дорожках. Это говорит о том, что данный фрагмент ДНК не образует детектируемых ДНК-белковых комплексов при инкубации с ядерными экстрактами как интактных, так и трансфицированных клеток.

Чтобы показать, что в составе ядерных экстрактов именно белок CTCF с гистидиновым тагом взаимодействует с CTCF-связывающими

последовательностями, был применен метод дополнительной ретардации в геле. Радиоактивно меченый фрагмент F1 инкубировался с ядерным экстрактом трансфицированных рННС-his клеток в присутствии антител к куриному белку CTCF, а также антител к гистидиновому тагу (рисунок 6В). После добавления антител к куриному белку CTCF, практически отсутствовали какие-либо полосы, соответствующие ДНК-белковым комплексам. По-видимому, взаимодействие антител с белком CTCF препятствовало образованию ДНК-белковых комплексов. Подобный эффект при использовании поликлональных антител к белку CTCF наблюдался в работах [44; 174]. В первом случае антитела использовались к С-концевому фрагменту, во втором – к N-концевому, как и в нашей работе. После добавления антител к гистидиновому тагу полосы, соответствующие ДНК-белковым комплексам, располагались выше в дорожках ПААГ. Смещение дополнительных полос вверх по дорожке геля наблюдается, если антитела взаимодействуют с белком в составе ДНК-белкового комплекса. Таким образом, белок CTCF в составе ядерного экстракта клеток COS-1 способен специфично взаимодействовать с CTCF-связывающими фрагментами ДНК *in vitro*.

**Выявление ядерной локализации CTCF с N-концевым гистидиновым тагом.** Известно, что белок CTCF обладает сигналом ядерной локализации [90; 43]. Эндогенный CTCF после синтеза транспортируется в ядро, что соответствует его роли транскрипционного фактора. Для анализа локализации рекомбинантного CTCF в трансфицированных клетках проводилось связывание радиоактивно меченного корового участка 5'HS4-инсулятора с цитоплазматическими экстрактами клеток, трансфицированных плазмидой рННС-his, и интактных (рисунок 6Б). Видно, что полоса, соответствующая комплексу ДНК с белком, отсутствует в цитоплазматическом экстракте как из интактных так и из трансфицированных рННС-his клеток. Таким образом, значительная часть белка CTCF экспрессированного с плазмиды рННС-his, локализуется в ядре, как и эндогенный CTCF.



**Рисунок 6. Анализ взаимодействия экспрессированного CTCF с фрагментами ДНК методом сдвига электрофоретической подвижности. Зонды: 5'-HS4 - фрагмент 5'-HS4-инсулятора из  $\beta$ -глобинового локуса кур; F1 – F1-элемент лизоцимного сайленсера кур; MYC – фрагмент промотора гена MYC человека; Enh – фрагмент энхансера из  $\beta$ -глобинового локуса кур. ЯЭ - ядерный экстракт, ЦЭ - цитоплазматический экстракт клеток COS-1. Трансф. +/- указывает на наличие/отсутствие трансфекции клеток плазмидой рННс-his. Стрелки указывают положение в геле комплекса CTCF с ДНК. А. – связывание CTCF из трансфицированных и нетрансфицированных клеток COS-1 с различными последовательностями ДНК. Цифры над дорожками указывают количество (в мкг) белка ядерного экстракта. Б. – связывание фрагмента 5'-HS4-инсулятора из  $\beta$ -глобинового локуса кур с ядерным и цитоплазматическим экстрактами клеток COS-1. В. – связывание комплексов CTCF с F1-элементом лизоцимного сайленсера кур и антителами к куриному CTCF или полигистидиновой последовательности.**

**Оптимизация условий очистки транскрипционного фактора CTCF, с N-концевым гистидиновым тагом.** Были оптимизированы условия выделения белка CTCF с N-концевым гистидиновым тагом методом металл-афинной хроматографии с помощью колонок с иммобилизованными Ni-IDA группами (Protino Macherey-NaGel). Для этого трансфицированные конструкцией рННс-his клетки COS-1 обрабатывали ультразвуком. Полученную суспензию наносили на колонку и

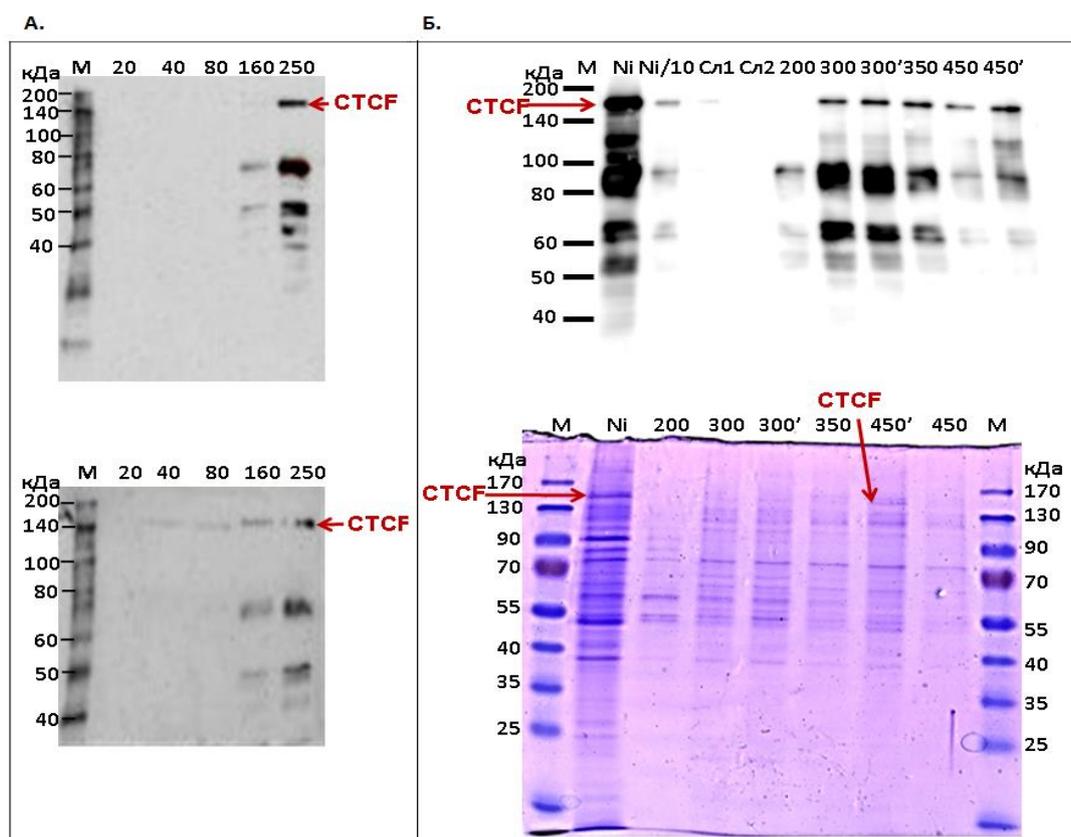
производили промывки буфером для уравнивания колонок (без имидазола) в соответствии с протоколом фирмы-производителя. Далее постепенно начинали градиентную промывку буфером с увеличивающейся концентрацией имидазола. Фракции, полученные при пропускании через колонку буфера, содержащего 20, 40, 80, 160 и 250 мМ имидазола, собирали и подвергали электрофоретическому разделению в SDS-ПААГ. Наличие белка в собранных фракциях анализировали с помощью антител к белку CTCF и к гистидиновому тагу. По результатам Вестерн-блот анализа этих фракций были выбраны условия промывки колонки и элюции белка (рисунок 7А).

На рисунке 7А видно, что наиболее интенсивная полоса на уровне 160 кДа, соответствующая полноразмерному белку CTCF, наблюдается во фракции, собранной при промывке буфером, содержащем 250 мМ имидазола. Также полоса слабой интенсивности, соответствующая полноразмерному CTCF, наблюдается при промывках буфером, содержащим от 40 до 160 мМ имидазола в своем составе. Таким образом, для очистки белка CTCF из клеточного лизата была выбрана 20 мМ концентрация имидазола в промывочном буфере. Его концентрация в буфере элюции была определена как 250 мМ.

Полоса с молекулярной массой около 70 кДа, также выявляемая антителами, не изменяла интенсивности при добавлении различных типов ингибиторов протеолитических ферментов, и ее происхождение может объясняться либо преждевременной терминацией полипептидной цепи [91], либо представлять собой продукт специфичного протеолиза. Предположение о наличии в клетках млекопитающих специфической CTCF-узнающей протеазы было высказано ранее [34].

Анализ полученной фракции при помощи SDS-электрофореза в полиакриламидном геле показал недостаточную чистоту белка (рисунок 7Б). Для дополнительной очистки мы применили хроматографию на Q-сефарозе, аналогично использованной в работе [4]. Элюат, полученный с Ni-IDA колонки, был разведен в 10 раз и нанесен на колонку с Q-сефарозой, с которой последовательно проводили элюцию буфером, содержащим 150, 250, 350 и 450 мМ NaCl. Полученные в результате дополнительной очистки белковые фракции были

проанализированы при помощи SDS-электрофореза в полиакриламидном геле с окраской Coomassie R-250 (рисунок 7Б). Основной примесный белок с молекулярной массой 70 кДа и другие примеси в основном элюируются при 250 мМ NaCl, тогда как белок CTCF сходит с колонки при 350-450 мМ. Как видно из рисунка 7Б, очистка в две стадии позволила нам получить фракцию, содержащую около 10% полноразмерного белка CTCF (второй элюат буфером, содержащим 450 мМ NaCl). Смывы с колонки буфером, содержащим 450 мМ NaCl, были объединены и использовались для получения библиотеки CTCF-связывающих последовательностей методом двумерного EMSA.



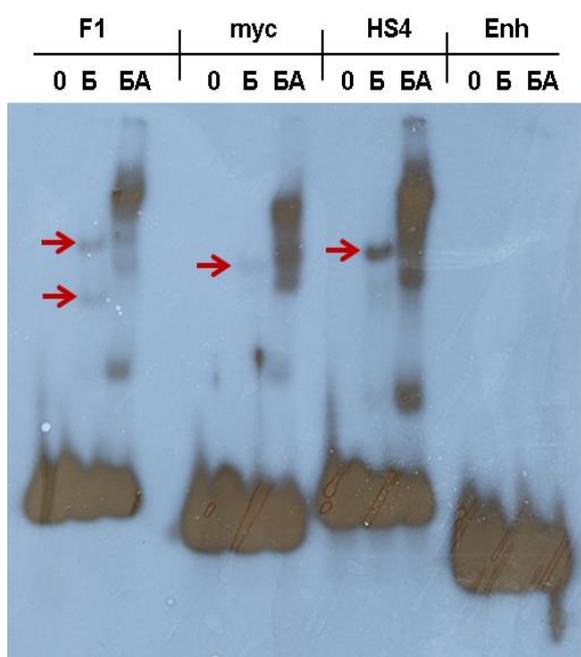
**Рисунок 7. Оптимизация условий очистки белка CTCF с N-концевым гистидиновым тагом из лизата клеток COS-1. CTCF – полоса, соответствующая полноразмерному белку CTCF. М – маркер молекулярных масс. А. – Вестерн-блот анализ фракций, полученных при очистке CTCF с N-концевым гистидиновым тагом с помощью колонок с иммобилизованными Ni-IDA группами. Дорожки: 20, 40, 80, 160, 250 – смывы с колонки буфером содержащим 20, 40, 80, 160 и 250 мМ имидазола, соответственно. Вверху – визуализация с помощью первичных антител к гистидиновому тагу. Внизу –**

визуализация с помощью первичных антител к СТСФ человека. Б. – Анализ фракций, полученных при дополнительной очистке СТСФ с N-концевым гистидиновым тагом с помощью колонок с Q-сефарозой. Дорожки: Ni – элюат с колонок с иммобилизованными Ni-IDA группами буфером, содержащим 250 мМ имидазол, Ni/10 – он же, после десятикратного разведения, 200, 300, 350, 450, – смывы с Q-сефарозы буфером, содержащим 200, 300, 350, 450 мМ NaCl, соответственно. 300', 450' – повторный смыв с Q-сефарозы буфером, содержащим 300, 450 мМ NaCl, соответственно. Вверху – Вестерн-блот анализ, визуализация с помощью первичных антител к гистидиновому тагу. Внизу – окраска полиакриламидного геля Кумасси.

**Функциональный анализ белковой фракции, обогащенной куриным белком СТСФ.** Был проведен анализ функциональной активности частично очищенного СТСФ с N-концевым гистидиновым тагом (элюат с Q-сефарозы буфером, содержащим 450 мМ NaCl, рисунок 7) методом дополнительного сдвига электрофоретической подвижности. СТСФ-связывающие последовательности, такие как F1-блок лизоцимного сайленсера кур [7; 94], фрагмент FII HS4 инсулятора кур [23; 11], фрагмент промотора гена с-тус человека [43] инкубировались с фракцией, обогащенной СТСФ с гистидиновым тагом. После инкубации этих фрагментов ДНК с белковой фракцией и электрофоретического разделения, на радиоавтографе ПААГ в соответствующих дорожках наблюдалось появление дополнительной полосы (рисунок 8). Это свидетельствовало об образовании ДНК-белковых комплексов. Появление нескольких дополнительных полос в случае фрагмента F1, по-видимому, является следствием взаимодействия с ДНК не только полноразмерного белка СТСФ, но и его короткой формы либо продуктов его частичной протеолитической деградации, происходившей во время очистки (продукты протеолиза СТСФ детектировались в обогащенной СТСФ фракции путем Вестерн-блот анализа обогащенной СТСФ фракции (рисунок 7Б)). Добавление антител к гистидиновому тагу приводило к дополнительному сдвигу полос, соответствующих ДНК-белковым комплексам. После добавления антител к гистидиновому тагу наблюдалось снижение электрофоретической подвижности

значительно большего количества ДНК, чем в случае добавления только СТСФ-содержащей белковой фракции. При этом было показано, что антитела к гистидиновому тагу не обладают неспецифической ДНК-связывающей активностью. Инкубация фрагмента энхансера из  $\beta$ -глобинового локуса кур, не способного связывать СТСФ [92], с обогащенной СТСФ фракцией, как без антител, так и в их присутствии не приводила к торможению этого фрагмента во время электрофореза в ПААГ, а значит, к образованию каких-либо ДНК-белковых комплексов. Таким образом, снижение электрофоретической подвижности дополнительных количеств ДНК в результате добавления антител к гистидиновому тагу является специфичным для фрагментов ДНК, способных связывать СТСФ. По-видимому, это объясняется особенностями взаимодействия между антителами и белком СТСФ с гистидиновым тагом, а также СТСФ с гистидиновым тагом и ДНК.

Таким образом, белок СТСФ, содержащий гистидиновый таг, после очистки специфично узнает СТСФ-связывающие фрагменты ДНК. Это свидетельствует в пользу сохранения белком нативной укладки и возможных посттрансляционных модификаций, необходимых для узнавания белком определенных последовательностей ДНК.



**Рисунок 8. Функциональный анализ фракции обогащенной куриным СТСФ с N-концевым гистидиновым тагом методом дополнительного сдвига электрофоретической подвижности. F1- F1-блок лизоцимного сайленсера кур, mys – фрагмент промотора гена *c-mys* человека, HS4 – FII-фрагмент HS4-**

инсулятора из  $\beta$ -глобинового локуса кур. 0 – фрагмент без белка, Б – фрагмент, инкубированный с белком, БА – белок, инкубированный с белком и антителами к гистидиновому тагу. Стрелками отмечены полосы, соответствующие ДНК-белковым комплексам.

*3.1.2. Получение библиотеки CTCF-связывающих последовательностей  $\alpha$ -глобинового локуса. Их идентификация и картирование*

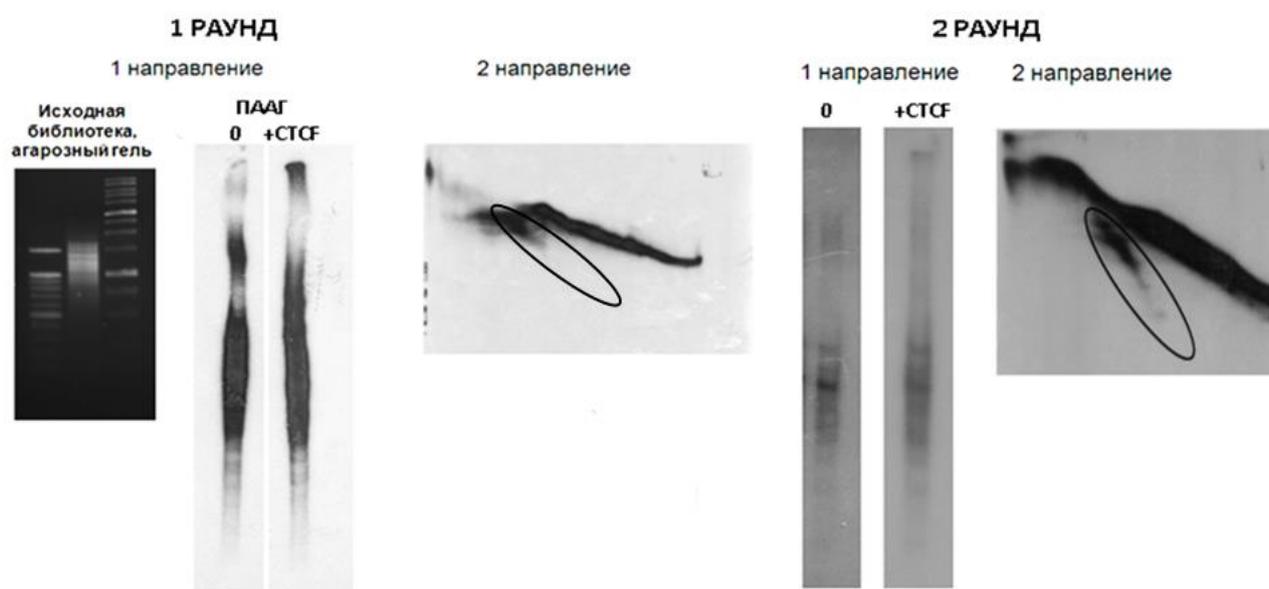
**Получение библиотеки перекрывающихся фрагментов  $\alpha$ -глобинового локуса кур.** Два образца искусственной бактериальной хромосомы, содержащей вставку локуса  $\alpha$ -глобиновых генов кур, длиной 230 т.н.о. были независимо обработаны эндонуклеазами рестрикции *CspBI* и *Sau3AI*. К каждому из двух полученных наборов рестриктных фрагментов был лигирован библиотечный адаптор и проведена ПЦР амплификация. Все ПЦР-амплификации библиотек проводились в присутствии бетаина и ДМСО для подавления ПЦР-селекции. Количества ДНК в полученных после амплификации образцах были выровнены, и две библиотеки фрагментов  $\alpha$ -глобинового локуса были объединены в одну. Полученная библиотека перекрывающихся фрагментов  $\alpha$ -глобинового локуса (рисунок 9) использовалась для последующего отбора CTCF-связывающих последовательностей методом двумерного EMSA.

**Получение библиотеки CTCF-связывающих фрагментов  $\alpha$ -глобинового локуса с помощью двумерного EMSA (Electrophoretic Mobility Shift Assay).**

Получение библиотеки связывающих CTCF фрагментов ДНК с помощью двумерного EMSA проводилось в два раунда. В первом раунде отбор CTCF-связывающих последовательностей проводился среди фрагментов библиотеки  $\alpha$ -глобинового локуса. Во втором раунде вместо полной библиотеки фрагментов  $\alpha$ -глобинового локуса использовалась библиотека обогащенная CTCF-связывающими последовательностями, полученная после первого раунда. Второй раунд проводился для повышения специфичности отбора.

Библиотека перекрывающихся фрагментов  $\alpha$ -глобинового локуса кур была радиоактивно мечена путем ее ПЦР-амплификации в присутствии дАТФ, содержащей в составе  $\alpha$ -фосфатной группы  $^{32}\text{P}$ . Затем меченая библиотека

использовалась в первом раунде двумерного EMSA. К ней был добавлен объем обогащенной CTCF фракции, содержащий порядка 0,5 пмоль полноразмерного CTCF и проведена инкубация. После инкубации образец был подвергнут электрофоретическому разделению в ПААГ, в неденатурирующих условиях. Дорожка, содержащая инкубированную с белковой фракцией библиотеку (рисунок 9), была вырезана из геля. После этого она использовалась для электрофоретического разделения во втором направлении в денатурирующих условиях. Фрагменты, связанные с белком во время электрофореза в первом направлении, образовали дополнительную диагональ, идущую под другим углом, нежели основная (рисунок 9).



**Рисунок 9. Получение набора CTCF-связывающих фрагментов  $\alpha$ -глобинового локуса кур методом двумерного EMSA. Слева направо. Первый раунд двумерного EMSA. Электрофоретическое разделение в агарозном геле библиотеки фрагментов  $\alpha$ -глобинового локуса кур. Электрофоретическое разделение в ПААГ этой же библиотеки без белка (0) и после инкубации с CTCF-содержащей фракцией (+CTCF). Электрофоретическое разделение инкубированной с белком библиотеки во втором направлении (дополнительная диагональ, которую образовали связавшие CTCF фрагменты ДНК, обведена). Второй раунд двумерного EMSA. Электрофоретическое разделение в ПААГ библиотеки, полученной в первом раунде EMSA, без белка (0) и после инкубации с CTCF-содержащей фракцией**

**(+CTCF). Электрофоретическое разделение инкубированной с белком библиотеки во втором направлении (дополнительная диагональ, которую образовали связавшие CTCF фрагменты ДНК, обведена).**

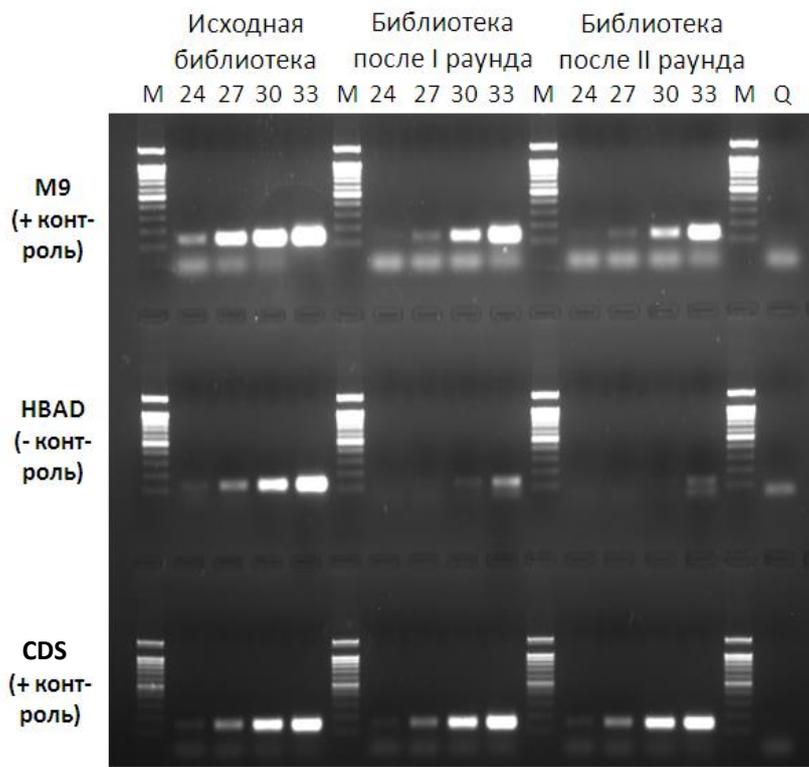
Далее дополнительная диагональ была вырезана из геля. Выделенная из нее ДНК использовалась для ПЦР-амплификации и второго раунда двумерного EMSA (рисунок 9).

Полученная во втором раунде двумерного EMSA дополнительная диагональ была вырезана из геля. Выделенная из нее ДНК была амплифицирована.

**Анализ полученной в результате двумерного EMSA библиотеки с помощью ПЦР с отбором через определенное количество циклов.** С помощью ПЦР с отбором проб для анализа через определенное количество циклов было показано обогащение библиотеки последовательностями  $\alpha$ -глобинового локуса кур, способными связывать белок CTCF в условиях экспериментов по сдвигу электрофоретической подвижности. Такими последовательностями являются CTCF-зависимый сайленсер, далее – CDS (CTCF Dependent Silencer), лежащий в промоторе гена *ggPRX* [92], и сайт связывания CTCF M9, лежащий в этом же гене [173]. В качестве отрицательного контроля был выбран случайный участок локуса – фрагмент гена *HBAD* (далее называемый фрагмент HBAD), для которого до этого не было показано какой-либо CTCF-связывающей активности. ПЦР-амплификация проводилась на матрице трех библиотек фрагментов ДНК: исходная библиотека, библиотека, полученная в результате первого раунда двумерного EMSA, библиотека, полученная после второго раунда двумерного EMSA. Электрофореграмма образцов, полученных при отборе из реакционных смесей через три цикла представлена на рисунке 10.

Как видно из рисунка 10, после двух раундов отбора продукты ПЦР областей CDS и M9 становятся видимыми после 24 и 27 циклов амплификации соответственно, тогда как продукт контрольного, не связывающего CTCF участка гена HBAD, становится видимым лишь после 33 циклов. С учетом того, что все три фрагмента амплифицируются из исходной библиотеки с примерно равной эффективностью, можно грубо оценить, что обогащение конечной библиотеки

СТCF-связывающими фрагментами составляет примерно 64–512 раз. По-видимому, обогащение библиотеки после двумерного EMSA фрагментом CDS более значительно, нежели фрагментом M9.



**Рисунок 10. Анализ обогащения библиотеки, полученной в результате двумерного EMSA, последовательностями, связывающими СТCF *in vitro*. ПЦР-амплификация с отбором через три цикла. Электрофореграмма отобранных аликвот реакционной смеси. M9 – амплификация СТCF-связывающего фрагмента M9, HBAD – амплификация фрагмента гена HBAD, CDS – амплификация фрагмента CDS. Исходная библиотека, библиотека после I раунда, библиотека после II раунда – матрицы в реакциях ПЦР-амплификации, I и II раунды – раунды двумерного EMSA. 24, 27, 30, 33 – цикл ПЦР, после которого отобрана аликвота. M-маркер масс, Q – контроль на контаминацию – в реакционной смеси отсутствует матрица.**

**Секвенирование полученной в результате двумерного EMSA библиотеки.** Полученная в результате двумерного EMSA библиотека была лигирована в вектор pGEM-T (Promega). Лигазная смесь использовалась для

трансформации клеток *E.coli*. Из полученных после трансформации клонов *E.coli* были выделены плазмиды, а затем секвенированы в области вставки.

В результате были определены нуклеотидные последовательности 230 фрагментов ДНК. 22 из них представляли собой фрагменты бактериальной ДНК либо были химерными последовательностями. Оставшиеся 208 являлись нехимерными фрагментами  $\alpha$ -глобинового локуса кур. Было определено их расположение на карте локуса (ПРИЛОЖЕНИЯ, таблица 2). В результате было выявлено 60 уникальных фрагментов, фланкированных сайтами рестрикции *Sau3AI*, и 32 фрагментов, фланкированных сайтами рестрикции *Csp6I*.

В ходе двумерного EMSA в результате взаимодействия CTCF с одним и тем же участком генома в большинстве случаев должны отбираться два рестриктных фрагмента: фланкированный сайтами рестрикции *Sau3AI* и сайтами рестрикции *Csp6I*. Далее два выявленных в конечной библиотеке рестриктных фрагмента, перекрывающихся между собой, мы будем называть уникальной CTCF-связывающей областью, подразумевая, что в большинстве случаев один общий сайт связывания CTCF находится в зоне перекрытия этих фрагментов. Один рестриктный фрагмент перекрывающийся с двумя, будет приравниваться к двум смежным CTCF-связывающим областям. Обнаруженные неперекрывающиеся с другими рестриктные фрагменты приравнивались к уникальным областям.

Всего было выявлено 79 уникальных областей.

По результатам секвенирования библиотеки был построен график зависимости количества выявленных уникальных областей от количества полученных сиквенсов (рисунок 11А). Этот график представляет собой кривую насыщения. График близок к выходу на плато, значит не обнаруженными осталась небольшая часть слабо представленных в библиотеке CTCF-связывающих областей. Так как на последних этапах секвенирования библиотеки на 18 полученных сиквенсов приходилась всего одна новая CTCF-связывающая область, дальнейшее секвенирование библиотеки представлялось нецелесообразным.

Также была построена кривая ранг-представленность (рисунок 11Б), которая отражает представленность различных фрагментов в библиотеке. Максимальное количество раз, которое при секвенировании библиотеки встретилась одна и та же

область – 16, минимальное – 1. 42 области встретились от 12 до 2-х раз. Остальные – единожды. Значительный разброс в количестве сиквенсов, соответствующих различным областям, по-видимому, говорит о том, что были выявлены последовательности, проявляющие *in vitro* как довольно сильное сродство к CTCF, так и относительно слабое.

Внутри обнаруженных последовательностей оказались расположены 6 из 7 (в соответствии с базой данных CTCF-binding site database, [8]) последовательностей  $\alpha$ -глобинового локуса, для которых ранее было показано связывание с CTCF в условиях экспериментов по сдвигу электрофоретической подвижности (последовательности C10 – C14 [173] и последовательность CDS из работы [92]). Обогащение седьмой (не выявленной в результате секвенирования библиотеки) последовательностью (сайт M9) было продемонстрировано нами с помощью ПЦР с отбором проб для анализа через определенное количество циклов, см. выше.

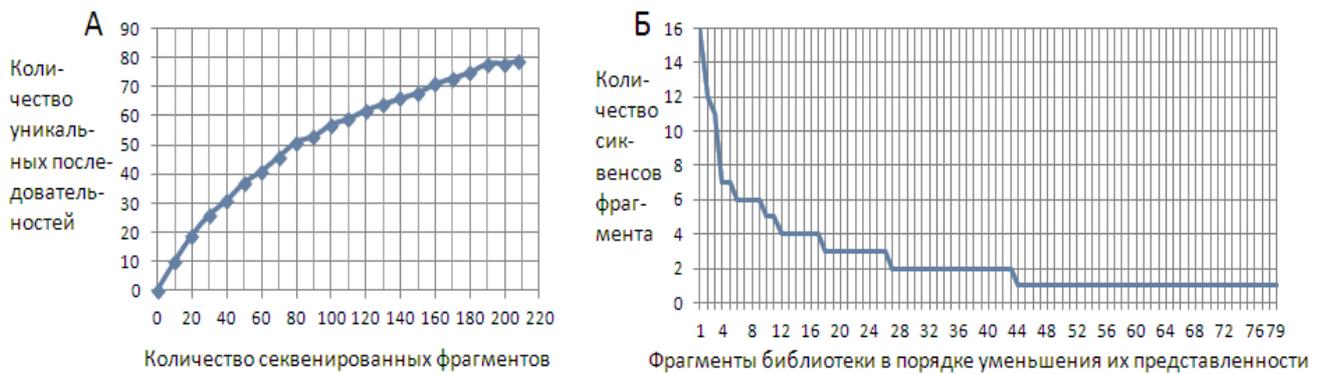
Ранее в исследуемой нами области куриного генома путем иммунопреципитации хроматина эмбриональных эритробластов с последующим масштабным секвенированием были выявлены шесть областей локуса, связывающих CTCF *in vivo* [122]. Трем областям, которые связывают белок CTCF в эритробластах пятидневного эмбриона кур мы дали обозначения 5d1-5d3 (5 days old embryo, сайт 1 - 5 days old embryo, сайт 3). Трем областям, которые связывают CTCF в эритробластах десятидневного эмбриона кур, были присвоены обозначения 10d1-10d3 (10 days old embryo, сайт 1 - 10 days old embryo, сайт 3). Их расположение в локусе  $\alpha$ -глобиновых генов кур показано на рисунке 12. Из шести сайтов связывания CTCF, выявленных Martin и соавторами в эмбриональных эритробластах, четыре располагаются в рестриктных фрагментах локуса, обнаруженных в результате двумерного EMSA. По результатам секвенирования библиотеки, полученной в результате двумерного EMSA, один сайт связывания CTCF (10d1) располагается в высоко представленном в библиотеке рестриктном фрагменте ДНК (11 сиквенсов), два сайта связывания CTCF (10d2 и 5d1) – в средне представленной последовательности (3 сиквенса) и один (10d3) – в слабо представленной (1 сиквенс). Два сайта связывания CTCF лежат вне рестриктных фрагментов, обнаруженных в результате двумерного EMSA. Один из них

расположен на небольшом расстоянии (100 п.о.) от выявленного в результате двумерного EMSA. Сайт связывания и место сшивки при иммунопреципитации хроматина могут не совпадать из-за изгиба ДНК [26, 27], т.е. фрагменты, идентифицируемые по EMSA и ChIP, не обязательно перекрываются, хотя и должны находиться на небольшом расстоянии друг от друга.

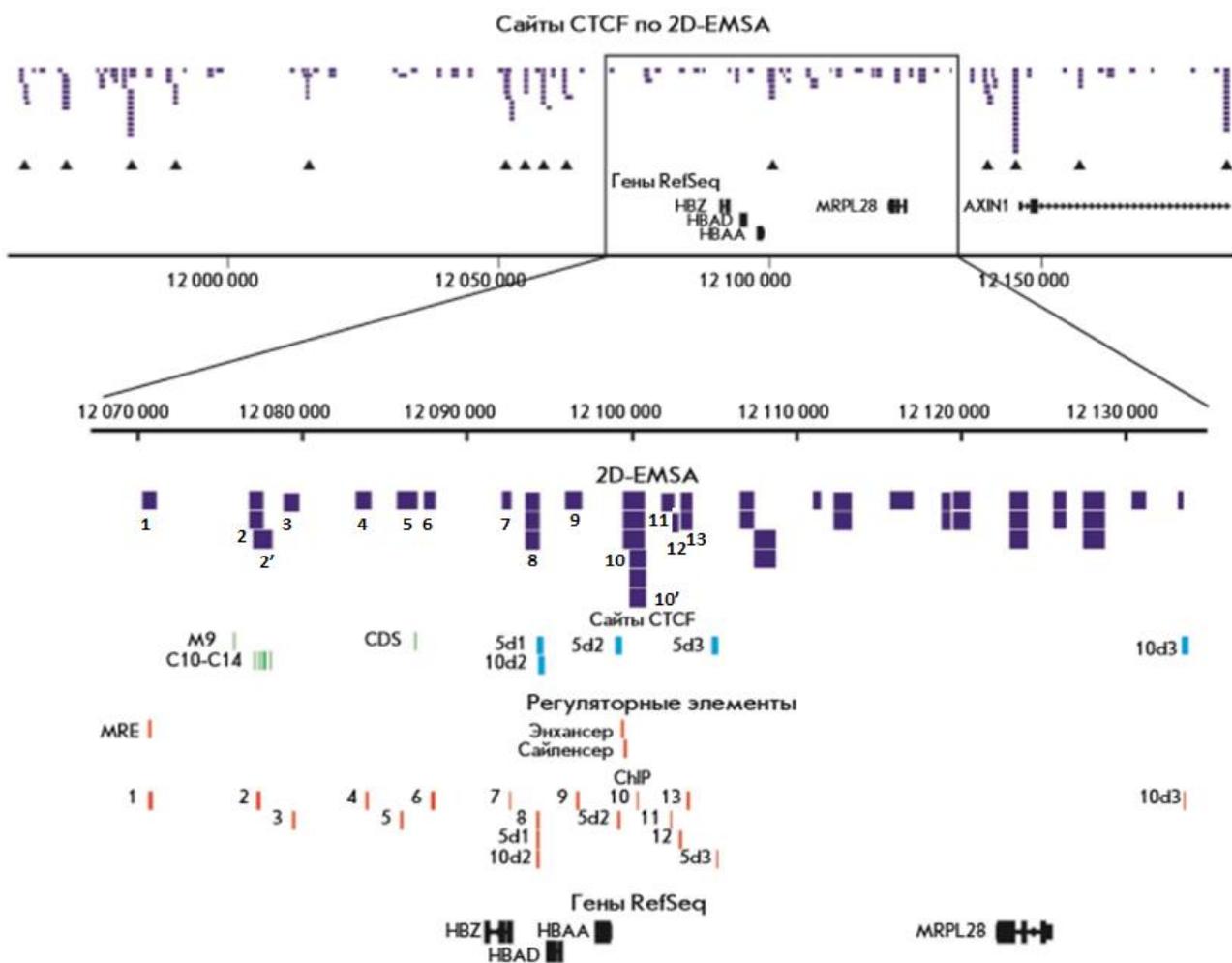
На карте локуса  $\alpha$ -глобиновых генов была выделена область, на изучении CTCF-связывающих последовательностей которой мы сосредоточили особое внимание. Это часть локуса, расположенная между MRE (Major Regulatory Element) и CTCF-связывающим сайтом эмбриональных эритробластов [122], отмеченным на рисунке 12 как 5d3. Эта область содержит гены  $\alpha$ -глобинов и большую часть обнаруженных в локусе регуляторных элементов [152]

Существуют свидетельства регуляции транскрипции на уровне всего домена  $\alpha$ -глобиновых генов. В эритроидных клетках происходят некоторые изменения структуры хроматина, такие как повышение уровня ацетилирования гистонов по ряду специфических позиций [26; 3; 202]. По-видимому, этот процесс контролируется путем низкоуровневой транскрипции всего домена. В области MRE в направлении генов  $\alpha$ -глобинов начинается синтез длинного транскрипта, размером около 33 т.п.о. [150]. Представляется возможным, что в области MRE может проходить граница функционального домена  $\alpha$ -глобиновых генов.

Фрагменты, выявленные в результате двумерного EMSA между MRE и сайтом связывания CTCF 5d3, были пронумерованы с возрастанием номера в сторону гена *TMEM8* от MRE локуса (рисунок 12). Фрагменты, перекрывающиеся между собой получили один и тот же номер без штриха и со штрихом. В области локуса, расположенной между MRE и сайтом связывания CTCF 5d3, в результате двумерного EMSA было выявлено 15 рестриктных фрагментов, которые с учетом перекрывания фрагментов, фланкированных сайтами *Sau3AI* и сайтами *CspBI*, составляют 13 неперекрывающихся между собой участков локуса. Был проведен анализ способности этих участков локуса связывать CTCF *in vitro* и *in vivo*.



**Рисунок 11. Графики, характеризующие размер библиотеки, полученной после двумерного EMSA, и представленность в ней различных фрагментов ДНК. А. Кривая накопления выявленных CTCF-связывающих областей. Б. Кривая ранг-представленность.**



**Рисунок 12. Расположение CTCF-связывающих участков и других регуляторных элементов в  $\alpha$ -глобиновом локусе генома кур. На верхней карте показано расположение всех полученных при отборе фрагментов ДНК. Стрелки указывают участки ДНК с повышенным сродством к CTCF. Ниже**

приведена увеличенная часть карты, представляющая непосредственное окружение глобиновых генов. В области «Сайты CTCF» показаны идентифицированные ранее сайты связывания CTCF M9, C10-C14 [173], CDS [92] и 5d1–5d3, 10d1–10d3 [122], в области «Регуляторные элементы» приведено расположение элементов MRE [45], энхансера и сайленсера [169]. В области ChIP указано расположение фрагментов ДНК, амплифицировавшихся в эксперименте по иммунопреципитации хроматина.

**Оценка специфичности проведенного отбора CTCF-связывающих последовательностей методом дополнительного сдвига электрофоретической подвижности.** Необходимо было подтвердить, что значительная часть отобранных в результате двумерного EMSA фрагментов ДНК действительно способна связывать белок CTCF *in vitro*. С этой целью был проведен анализ CTCF-связывающей активности десяти фрагментов локуса методом дополнительного сдвига электрофоретической подвижности. Это фрагменты, расположенные между MRE и сайтом связывания CTCF 5d3, номера 1, 2, 3, 4, 6, 7, 8, 9, 10' и 13 (рисунок 12). Эти фрагменты имели различную представленность в библиотеке. Среди них есть последовательности, встретившиеся при секвенировании библиотеки единожды и не перекрывающиеся с другими обнаруженными фрагментами, есть – обнаруженные два и три раза и перекрывающиеся при этом с другими фрагментами, выявленными в библиотеке.

Радиоактивно меченые фрагменты ДНК инкубировались с ядерным экстрактом трансфицированных плазмидой рННс клеток COS-1, содержащим куриный CTCF с N-концевым гистидиновым тагом. Дополнительно проводилась инкубация каждого фрагмента с этим же ядерным экстрактом в присутствии антител к куриному CTCF. После инкубации образцы наносились на ПААГ. Затем проводилось электрофоретическое разделение образцов. В случае каждого образца мы видим появление на дорожке дополнительных полос после инкубации с ядерным экстрактом (рисунок 13А). Это объясняется формированием ДНК-белковых комплексов, которые снижают подвижность меченой ДНК в геле. В дорожках, соответствующих фрагментам 2,3,9, после добавления антител,

наблюдается исчезновение полосы, соответствующей ДНК-белковому комплексу, при увеличении интенсивности полосы на уровне миграции фрагмента в гель. Это происходит в результате дополнительного торможения ДНК-белкового комплекса за счет связывания с ним антител. В случае других фрагментов, наблюдалось только исчезновение полос, соответствующих ДНК-белковым комплексам. Связывание антител с белком CTCF, по-видимому, препятствует формированию ДНК-белковых комплексов с этими фрагментами. Мы уже наблюдали такой эффект при использовании этих антител для дополнительной ретардации в геле (рисунок 6).

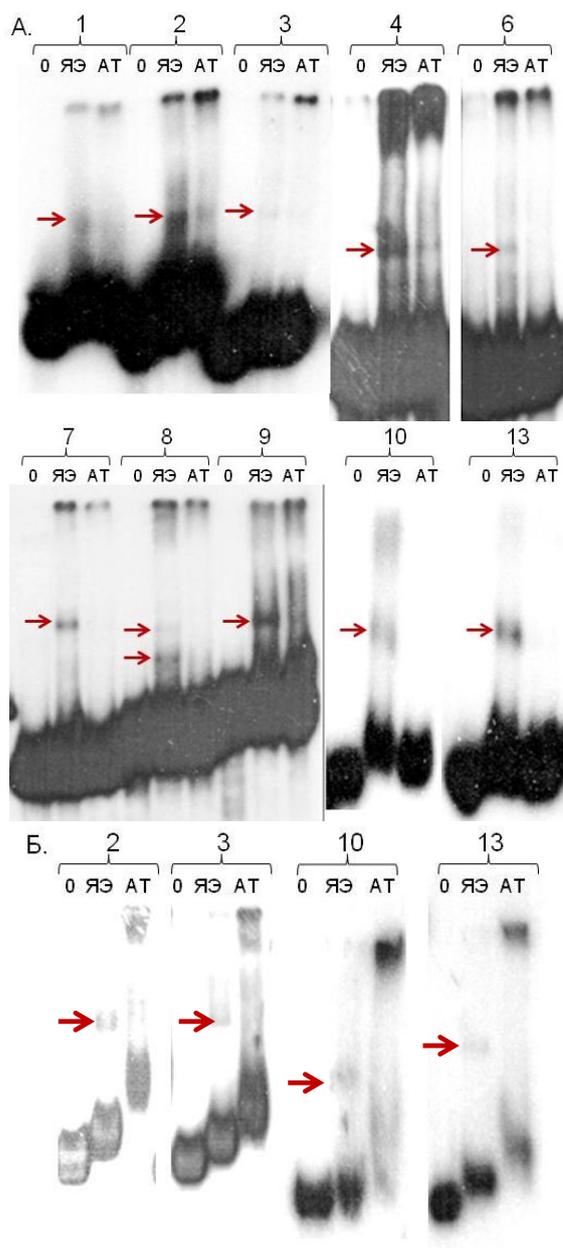
Дополнительно был проведен анализ CTCF-связывающей активности четырех обнаруженных фрагментов ДНК методом дополнительного торможения в геле с использованием антител к гистидиновому тагу (рисунок 13Б). Инкубация всех четырех фрагментов с ядерным экстрактом, содержащим куриный CTCF с гистидиновым тагом, приводила к образованию ДНК-белковых комплексов и, как следствие, дополнительных полос на радиоавтографе. Дополнительная инкубация с антителами к гистидиновому тагу приводила к сдвигу полос, соответствующих ДНК-белковым комплексам, на радиоавтографе. Вновь был отмечен своеобразный, но, по-видимому, специфичный для CTCF-связывающих последовательностей (см. “Функциональный анализ белковой фракции, обогащенной куриным белком CTCF”, раздел 3.1.1., рисунок 8), эффект торможения в геле значительно больших количеств ДНК при добавлении антител к гистидиновому тагу, нежели в их отсутствии.

Появление нескольких дополнительных полос после инкубации с ядерным экстрактом, исчезающих после добавления антител, возможно, является результатом связывания с фрагментом ДНК не только полноразмерного белка CTCF, но и его короткой формы или фрагментов CTCF, обладающих ДНК-связывающей активностью и появляющихся в результате его протеолиза (рисунок 7Б, Вестерн-блот анализ CTCF-содержащих фракций).

Смещение и исчезновение полос, соответствующих ДНК-белковым комплексам, после добавления антител к CTCF и к гистидиновому тагу

свидетельствует о том, что именно CTCF, присутствующий в ядерном экстракте взаимодействует с каждым из фрагментов ДНК.

Способность всех 10 фрагментов ДНК образовывать комплексы с белком CTCF говорит о достаточно высокой (около 100%) специфичности отбора, проведенного методом двумерного EMSA. Таким образом, даже фрагменты, встретившиеся во время секвенирования библиотеки 1 раз способны специфично образовывать комплекс с CTCF в условиях экспериментов по сдвигу электрофоретической подвижности.



**Рисунок 13. Анализ CTCF-связывающей активности десяти фрагментов ДНК, отобранных в результате двумерного EMSA методом дополнительного сдвига электрофоретической подвижности. Числа соответствуют номерам**

фрагментов, рисунок 12. 0 – фрагмент ДНК не инкубировался с белковыми фракциями, ЯЭ – фрагмент ДНК инкубировался с ядерным экстрактом клеток COS-1, содержащим белок CTCF с N-концевым гистидиновым тагом, АТ – фрагмент ДНК инкубировался с ядерным экстрактом, содержащим белок CTCF с гистидиновым тагом в присутствии антител. Стрелками показано расположение полос, соответствующих ДНК-белковым комплексам. А. Для анализа методом дополнительного сдвига электрофоретической подвижности использовались антитела к CTCF. Б. Для анализа методом дополнительного сдвига электрофоретической подвижности использовались антитела к гистидиновому тагу.

### 3.1.3. Выявление среди отобранных в результате двумерного EMSA фрагментов ДНК, последовательностей, способных связывать CTCF *in vivo*

Чтобы определить, какая часть выявленных в результате двумерного EMSA областей генома связывает белок CTCF *in vivo*, мы провели иммунопреципитацию хроматина [134] куриных клеток линии HD3 (интактных и после индукции дифференцировки) и DT40. Далее HD3 – интактные клетки линии HD3, HD3-ind – клетки HD3 после индукции дифференцировки.

Клеточная линия HD3 была получена благодаря вирусной трансформации эритробластов взрослых кур [13]. Интактные клетки линии HD3 делятся, в них происходит транскрипция генов взрослых  $\alpha$ -глобинов, но образование белка отсутствует. Индукция дальнейшей дифференцировки клеток линии HD3, делает их сходными с эритроцитами по форме и по синтезирующимся в этих клетках поверхностным антигенам [56]. Также в результате индукции происходит образование гемоглобина.

Благодаря этому, линия HD3 используется как модель для изучения процессов, идущих при эритроидной дифференцировке [13; 56].

При индукции клеток HD3 синтез гемоглобина активируется в значительной степени посттранскрипционно. Происходит смена ядерной локализации глобиновых транскриптов на цитоплазматическую. По-видимому в результате их процессинга [75].

Также после индукции повышается содержание транскриптов взрослых  $\alpha$ -глобиновых генов [56; 50; 143] и происходит изменение пространственной структуры  $\alpha$ -глобинового локуса [50; 143]. Вероятно при индукции происходит регуляция транскрипции глобиновых генов за счет изменения пространственной структуры локуса.

Множество работ демонстрирует участие белка CTCF в пространственных взаимодействиях геномных областей [121; 17; 135; 41; 63; 183]. Выявление участков  $\alpha$ -глобинового локуса, связывающих CTCF в индуцированных и интактных клетках HD3, и сравнение полученных данных может в дальнейшем привести к пониманию механизмов регуляции экспрессии генов  $\alpha$ -глобинового домена и выявлению роли, которую при этом играет белок CTCF.

Клеточная линия DT40, получена на основе В-клеточной лимфомы кур. В ее клетках, не детектируются транскрипты  $\alpha$ -глобиновых генов, что характерно для клеток неэритроидного происхождения. Между клетками HD3 и DT40 имеются различия в пространственной структуре  $\alpha$ -глобинового локуса. [50; 143] Различия в пространственной структуре локуса и экспрессии  $\alpha$ -глобиновых генов в клетках HD3 и DT40 могут быть также связаны с различиями в расположении сайтов связывания CTCF в этих клетках. Поэтому представлялось важным сравнить расположение и занятость белком фрагментов локуса, связывающих CTCF *in vivo* в клетках DT40, HD3 и HD3-ind.

В результате каждого эксперимента по иммунопреципитации хроматина было получено три образца ДНК: Input — тотальная ДНК, выделенная из клеточного лизата после ультразвуковой обработки, CTCF — образец, полученный в результате иммунопреципитации хроматина с антителами к белку CTCF, и Tau — образец, полученный с антителами к растительному белку тауматину — контроль на неспецифическое связывание.

Иммунопреципитация хроматина в случае клеток HD3 интактных и DT40 была проведена в трех повторностях, в случае клеток HD3 индуцированных к дифференцировке — в двух.

Анализ качества ультразвукового дробления в каждом эксперименте проводился путем электрофоретического разделения ДНК образца Input (рисунок 14).

Методом ПЦР в реальном времени был проведен анализ обогащения иммунопреципитов хроматина CTCF-связывающих фрагментов, выявленными в результате двумерного EMSA между MRE и сайтом 5d3 эмбриональных эритробластов. В случае перекрывания между собой рестриктных фрагментов (одна CTCF-связывающая область) приведены результаты, полученные с использованием пары праймеров к одному из фрагментов.

Также был проведен анализ обогащения иммунопреципитатов фрагментами генома, способными связывать белок CTCF *in vivo* по данным различных авторов. Такими фрагментами генома являются F1-блок лизоцимного сайленсера (фрагмент F1), фрагмент промотора гена *c-тус* кур (фрагмент MYC) [43], фрагменты  $\alpha$ -глобинового локуса, связывающие CTCF *in vivo* в эмбриональных эритроблестах кур по данным ChIP-seq анализа [122](расположение - см. рисунок 12). Результаты иммунопреципитации хроматина для исследуемого фрагмента сравнивались с результатами для фрагмента, не способного связывать CTCF, (энхансер из бэта-глобинового локуса кур — Enh), и случайно выбранного фрагмента  $\alpha$ -глобинового локуса (фрагмент гена *HBAD* — HBAD) — отрицательные контроли.

На рисунке 15А приведено содержание каждого фрагмента ДНК в иммунопреципитате относительно его содержания в образце Input, в процентах (% от Input). Значения, полученные для каждого фрагмента, были нормализованы на значения, полученные для положительного контроля – F1 (рисунок 15В). Обогащение определенным фрагментом генома в результате иммунопреципитации хроматина характеризует занятость этой геномной области белком в используемых клетках.

На рисунке 15 видно обогащение иммунопреципитатов хроматина всех трех типов клеток фрагментами F1 и MYC, а также фрагментами 10 и 5d3. Фрагмент F1, связывает CTCF *in vivo* в различных типах клеток, как в активном, так и в неактивном состоянии лизоцимного сайленсера [116], мы рассматривали его как потенциально конститутивный сайт и положительный контроль в эксперименте.

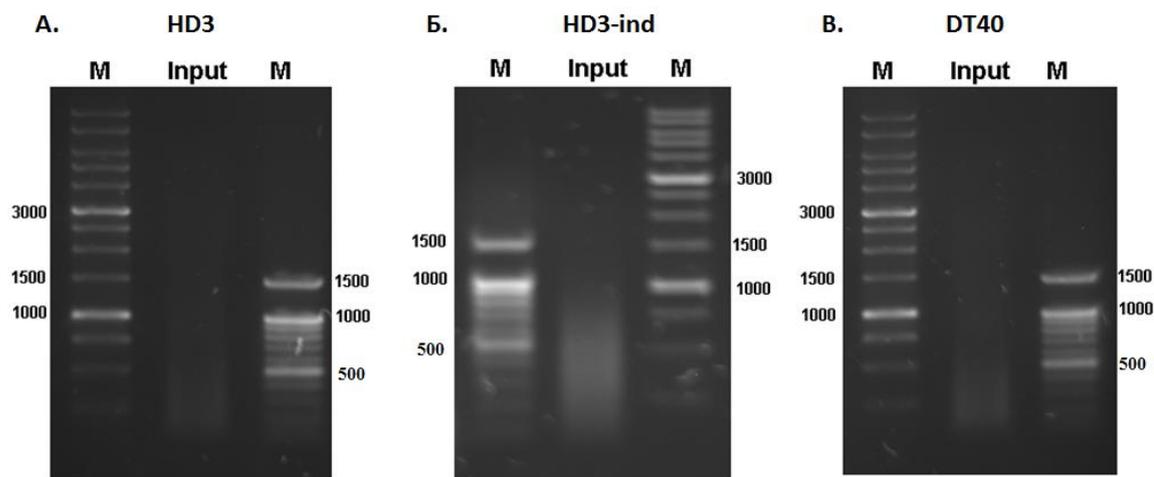
Высокая занятость белком фрагментов F1 и MYC, в сравнении с отрицательными контролями, показанная во всех образцах клеток, свидетельствует о специфичном отборе связанных с CTCF фрагментов ДНК в ходе иммунопреципитации хроматина. Столь же высокая занятость белком CTCF выявляется для фрагмента 10, но лишь в двух образцах клеток: HD3 и HD3-ind. В клетках DT40 степень связывания белком этого сайта значительно (в 2-3 раза) ниже. При сравнении занятости белком CTCF этого сайта с применением t-теста Стьюдента в случае клеток HD3 и DT40  $p < 0.05$ . Это говорит о тканеспецифичности связывания с ним белка CTCF. При другом попарном сочетании образцов уровень значимости был выше 0.05, что не позволяло сделать достоверные выводы о различиях. Степень связывания CTCF с 5d3 выше значений отрицательного контроля (Enh) для всех трех типов клеток, но существенно ниже, чем у фрагментов F1 и MYC, а также у фрагмента 10 в клетках HD3 и HD3-ind.

Таким образом, из тринадцати CTCF-связывающих областей локуса, расположенных между MRE локуса и эмбриональным сайтом связывания CTCF 5d3, лишь одна связывает CTCF *in vivo* в выбранных типах клеток. По-видимому, сродство CTCF к фрагменту генома с определенной нуклеотидной последовательностью является лишь одним из факторов, определяющих, будет ли данный участок ДНК связывать CTCF *in vivo*. Другим фактором, влияющим на связывание CTCF с фрагментом ДНК, является метилирование цитозина в составе динуклеотида CpG в некоторых CTCF-связывающих последовательностях, препятствующее взаимодействию с ними CTCF [10; 64]. Также связывание CTCF с ДНК *in vivo* определяется структурой хроматина/модификациями гистонов и/или наличием вблизи сайтов связывания других факторов транскрипции, облегчающих связывание CTCF [35].

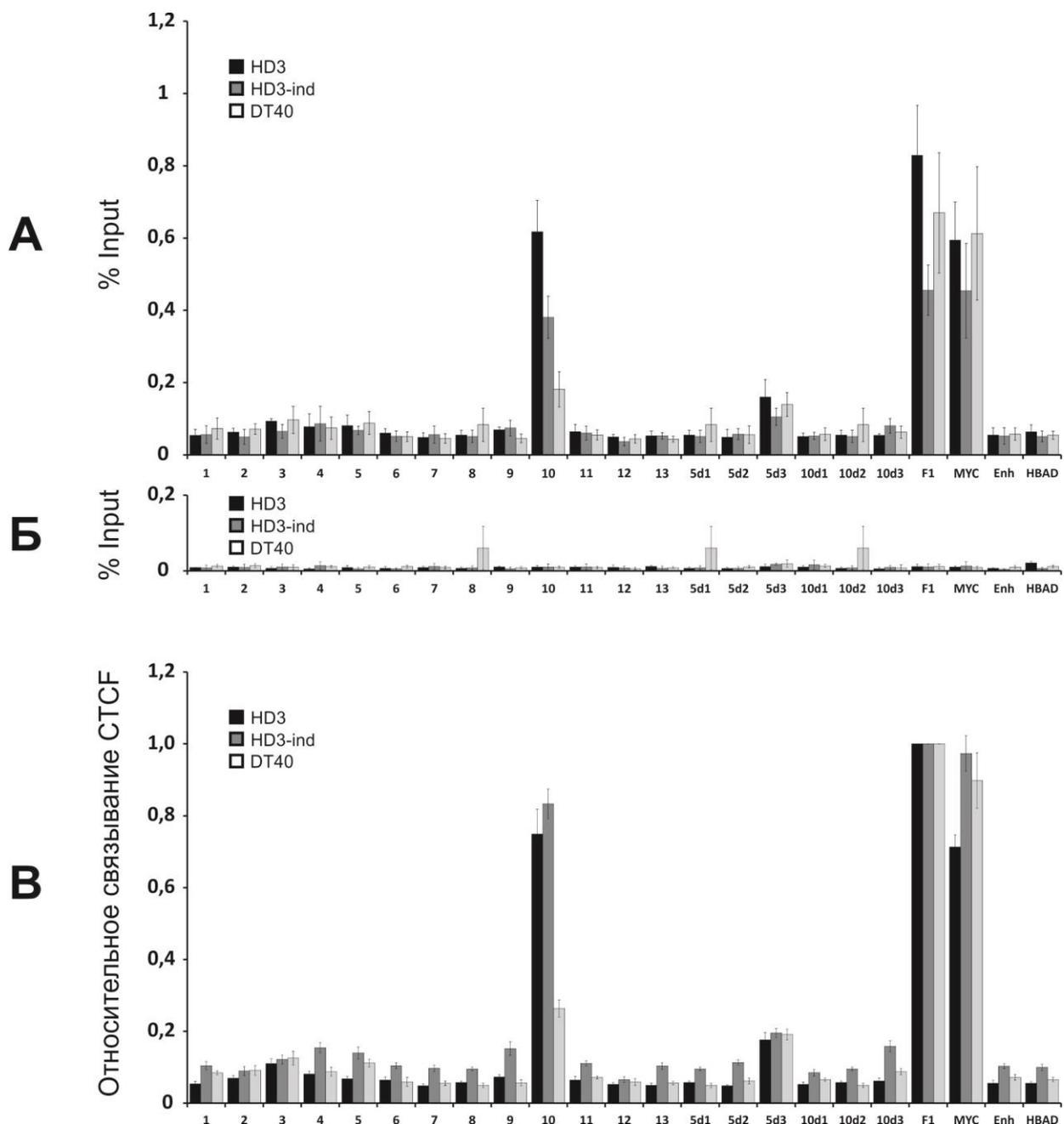
В случае большинства сайтов связывания CTCF, выявленных в эритроцитах эмбрионов кур, не было показано связывание с CTCF в клетках HD3 и HD3-ind. Лишь в случае одного сайта было выявлено слабое превышение связывания с белком CTCF в сравнении с отрицательным контролем. Это может свидетельствовать о тканеспецифичности связывания CTCF с этими областями локуса и о значительных различиях в распределении областей, связывающих CTCF

*in vivo* в эмбриональных и взрослых эритроблестах кур (фрагмент 10, связывающий CTCF в клетках HD3 и HD3-ind, напротив, не был выявлен в эмбриональных эритроблестах).

Сравнение образцов ДНК, полученных с использованием в ходе иммунопреципитации антител к белку тауматину, не выявило обогащение фрагментами F1, MYC, 10 и 5d3 в сравнении с фрагментами Eln и HBAD (рисунок 15Б). Таким образом, обогащение образца CTCF этими фрагментами не является результатом неспецифической сорбции.



**Рисунок 14. Электрофоретический анализ качества ультразвукового дробления ДНК в ходе иммунопреципитации хроматина. Дорожки: Input – образец ДНК Input, M-маркер длин фрагментов ДНК. 500, 1000, 1500 и 3000 – длины фрагментов ДНК, которым соответствуют полосы в дорожке маркеров, в н.о. Над каждой электрофореграммой указан тип клеток. А.-В. Образцы Input получены в ходе иммунопреципитации хроматина клеток HD3, HD3 индуцированных и DT40, соответственно.**



**Рисунок 15. Результаты иммунопреципитаций хроматина клеток HD3, интактных и после индукции дифференцировки и DT40, с последующим анализом иммунопреципитатов ПЦР в реальном времени. А.**

Иммунопреципитат хроматина получен с антителами к белку CTCF. По вертикальной оси отложен процент содержания данного фрагмента в иммунопреципитате от содержания данного фрагмента в образце Input (%Input). Б. Иммунопреципитат хроматина получен с антителами к белку тауматину. По вертикальной оси отложен процент содержания данного фрагмента в иммунопреципитате от содержания данного фрагмента в образце Input (%Input). В. Иммунопреципитат хроматина получен с антителами к

белку CTCF. Высота столбиков отражает занятость белком CTCF фрагмента локуса, относительно фрагмента F1. А.-В. Столбики черного цвета соответствуют клеткам линии HD3, серого – клеткам линии HD3 после индукции дифференцировки и светло-серого – клеткам линии DT40. 1 – 13 – фрагменты ДНК, выявленные в результате двумерного EMSA. 5d1 – 5d3 – сайты связывания CTCF, в хроматине эритробластов пятидневных эмбрионов кур. 10d1 – 10d3 - сайты связывания CTCF, в хроматине эритробластов пятидневных эмбрионов кур. F1-блок лизоцимного сайленсера (фрагмент F1), MYS – фрагмент промотора гена *c-тус* кур, Enh – энхансер из  $\beta$ -глобинового локуса кур, NBAD – последовательность экзона гена NBAD, Высота столбиков – среднее арифметическое, планки погрешностей – стандартная ошибка среднего по трем либо двум (в случае HD3-ind) экспериментам по иммунопреципитации хроматина.

### *3.2. Полногеномное выявление связывающих белок CTCF нуклеотидных последовательностей методом ChIP-seq*

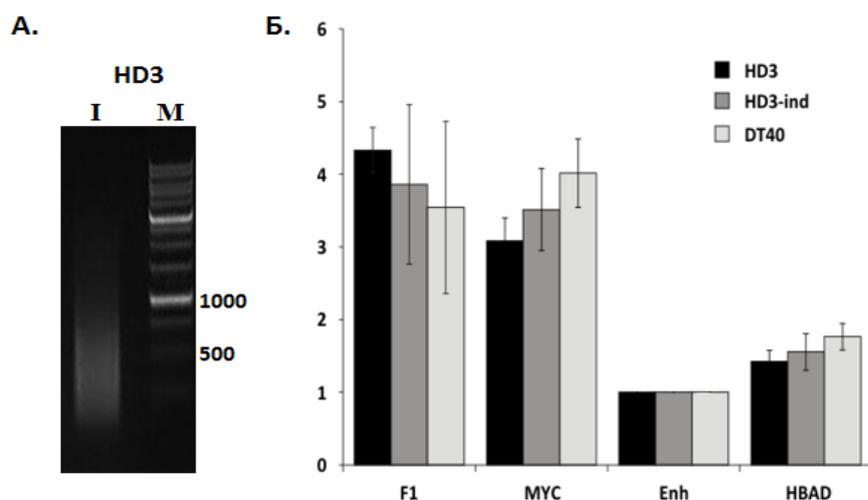
Методом ChIP-seq (Chromatin Immunoprecipitation Sequencing) [9] был проведен полногеномный поиск областей, связывающих CTCF *in vivo*, в клетках HD3, HD3-ind и DT40. В основе данного метода лежит иммунопреципитация хроматина. Анализ содержания в иммунопреципитате различных фрагментов ДНК при этом проводится путем массового секвенирования.

Для иммунопреципитации хроматина в ходе ChIP-seq использовались антитела к белку CTCF, ранее проявившие специфичность в экспериментах по иммунопреципитации хроматина (см. раздел 3.1.3.).

Массированному секвенированию были подвергнуты четыре образца ДНК: полученные с антителами к CTCF иммунопреципитаты хроматина клеток HD3 и HD3-ind, клеток DT40 и тотальная ДНК, выделенная из лизата клеток HD3 (образец Input).

Методом ПЦР в реальном времени было продемонстрировано обогащение иммунопреципитированной ДНК фрагментами, связывающими CTCF *in vivo* в данных клетках: F1 и MYS (см. главу 3.1.3). Для сравнения использовались не связывающие CTCF фрагменты ДНК: Enh и NBAD. Обогащение фрагментами, специфично связывающими CTCF, по сравнению с не связывающими его

фрагментами составило 3-4 раза. Это свидетельствует о специфичности отбора CTCF-связывающих фрагментов в ходе иммунопреципитации хроматина.



**Рисунок 16. Оценка качества иммунопреципитированной ДНК, полученной во время ChIP-seq-эксперимента. А. Дробление ДНК ультразвуком.**

**Электрофоретический анализ в агарозном геле. Дорожки: I – образец Input, M – маркер длин фрагментов ДНК. HD3 и DT40 – линии клеток, которым соответствует образец Input. Б. Анализ обогащения образцов иммунопреципитированной ДНК CTCF-связывающими фрагментами методом ПЦР в реальном времени. По вертикальной оси отложены значения обогащения данным фрагментом ДНК иммунопреципитата хроматина относительно отрицательного контроля – фрагмента энхансера из  $\beta$ -глобинового локуса кур, не связывающего белок CTCF (фрагмент Enh). Планки погрешностей соответствуют стандартной ошибке среднего по трем измерениям.**

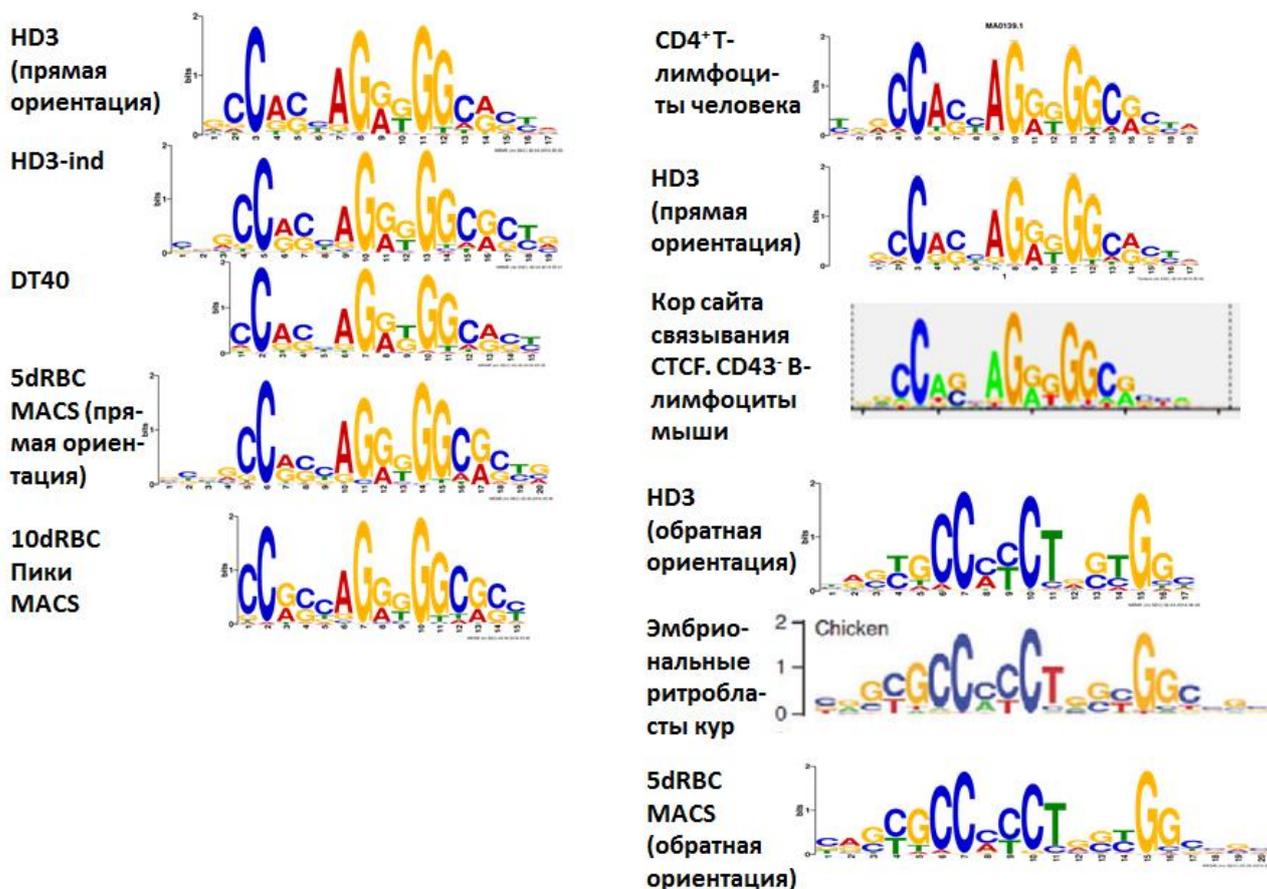
Для каждого образца было получено 41-50 млн прочтений. Было проведено картирование полученных прочтений на геном банкивской джунглевой курицы (*Gallus gallus*) сборки galGal4, 2011 г. Прочтения с показателем качества картирования MAPQ<10 были удалены. При этом доля прочтений, для которых было выявлено единственное местоположение на карте генома, составляла 86-88%.

Далее был проведен поиск пиков - областей генома, для которых наблюдается повышенная плотность картированных прочтений. Плотность картированных прочтений характеризует занятость белком данного участка генома. Таким образом, расположение пиков отражает расположение участков генома, связывающих исследуемый белок (в нашем случае CTCF). Для выявления

пиков использовалась программа MACS 1.4.2. Образец Input, полученный с использованием интактных клеток HD3, использовался в качестве контроля при выявлении пиков в случае иммунопреципитатов хроматина клеток HD3, интактных и индуцированных.

В случае образца HD3 было выявлено 10695 пиков, в случае образца HD3-ind – 3471 пик, в случае DT40 – 10770 пиков.

С помощью программы MEME-chip был проведен поиск мотивов, которыми обогащены нуклеотидные последовательности пиков (рисунок 17). Для всех иммунопреципитатов хроматина с наименьшей значимой вероятностью (p-value) были выявлены сходные мотивы размером 15-20 п.н. Выявленные мотивы имели высокое сходство с известными консенсусными последовательностями сайтов связывания CTCF позвоночных, в том числе кур. Обогащение нуклеотидных последовательностей пиков консенсусной последовательностью сайтов связывания CTCF указывает на специфичное выявление в результате ChIP-seq эксперимента областей, связывающих CTCF.



**Рисунок 17. Консенсусные последовательности сайтов связывания CTCF, выявленные при анализе данных ChIP-seq. HD3, HD3-ind, DT40, 5dRBC,**

**10dRBC – консенсусные последовательности, выявленные в случае клеток линии HD3, интактных, клеток линии HD3, после индукции дифференцировки, клеток линии DT40, эритробластов пятидневного эмбриона кур, эритробластов десятидневного эмбриона кур, соответственно. В случае клеток линии HD3 и эритробластов пятидневных эмбрионов кур выявленный мотив представлен в прямой и обратной ориентации (HD3 (прямая ориентация), HD3 (обратная ориентация), 5dRBC (прямая ориентация), 5dRBC (обратная ориентация)). CD4<sup>+</sup> Т-лимфоциты человека, CD43<sup>-</sup> В-лимфоциты мыши, эмбриональные эритробласты кур – консенсусные последовательности сайта связывания CTCF выявленные в результате ChIP-seq экспериментов с использованием данных типов клеток [122; 9; 126]. Кор сайта связывания CTCF – консенсусная последовательность соответствует кору сайта связывания CTCF [126].**

*3.2.1. Снижение занятости белком CTCF-связывающих областей при индукции дифференцировки клеток HD3*

Количество пиков, выявленных в случае индуцированных клеток HD3, оказалось в 3 раза меньше, чем для интактных. Большая часть пиков, выявленных в индуцированных HD3 (92%), совпадала с пиками, выявленными в интактных клетках HD3. При сравнении пиков, общих для двух образцов, обогащение относительно фонового уровня в образце HD3-ind оказалось в среднем по общим пикам в 1,5 раза меньше. По данным иммунопреципитации хроматина с последующим анализом иммунопреципитированной ДНК путем ПЦР в реальном времени (рисунок 15А) в случае CTCF-связывающих фрагментов MYC, F1 и фрагмента 10, выявленного путем двумерного EMSA, обогащение ими иммунопреципитированной с антителами к CTCF ДНК в случае интактных клеток HD3 было также примерно в 1,5 раза больше, чем индуцированных. По-видимому, в клетках HD3 в результате индукции была снижена средняя занятость белком сайтов связывания CTCF. Это может быть связано с процессами компактизации хроматина, сопровождающими заключительные стадии эритроидной дифференцировки у птиц [98]. Снижение занятости белком CTCF его сайтов

связывания также может быть следствием апоптотических изменений хроматина при дифференцировке клеток HD3 [75; 165].

### 3.2.2. Расположение сайтов связывания CTCF в $\alpha$ -глобиновом локусе кур

Расположение пиков, выявленных программой MACS 1.4.2 в  $\alpha$ -глобиновом локусе кур (геномные координаты chr14:11957588-12184953 в сборке куриного генома 2011) приведено на рисунке 18, а также в приложениях, таблицы 3 и 4.

В клетках линии DT40 было выявлено 4 пика, HD3 — 5 пиков. В индуцированных клетках HD3 программа не выявила в локусе  $\alpha$ -глобиновых генов ни одного пика. Однако распределение плотности прочтений по геному в случае этого образца имеет зоны значительно более высокой, по сравнению с окружающими областями генома, плотности прочтений, совпадающие с тремя пиками, выявленными программой MACS в образце HD3. Как было сказано выше (см. раздел 3.2.1), в индуцированных клетках HD3 была снижена средняя занятость белком сайтов связывания CTCF. Поэтому программой MACS часть связывающих CTCF участков локуса могла быть не выявлена.

Три пика, выявленные в случае клеток HD3 и DT40 перекрывались (были общими для обоих образцов). Два пика были выявлены только в образце HD3, один — только в образце DT40.

Один из пиков (номер 3, рисунок 18), общий для клеток HD3 и DT40, расположен ближе других к  $\alpha$ -глобиновым генам с 5' стороны. Данный пик расположен рядом с MRE локуса (Major Regulatory Element – основной регуляторный элемент), с противоположной стороны от глобиновых генов. Такое расположение делает его потенциальной границей хроматинового домена  $\alpha$ -глобиновых генов в клетках взрослых кур (возможно, только в клетках крови и гематопоэтических клетках). Сайт связывания CTCF в составе этого пика может препятствовать перекрестному влиянию цис-регуляторных элементов  $\alpha$ -глобинового локуса (в том числе MRE) и цис-регуляторных элементов, расположенных с другой стороны от него. В индуцированных клетках HD3 этому пику соответствует область с повышенной плотностью прочтений. По-видимому, в

индуцированных клетках HD3 занятость белком CTCF этого участка также повышена.

Один из пиков, выявленных только в образце HD3 (номер 6, рисунок 18), расположен с 3' стороны от  $\alpha$ -глобиновых генов. Он перекрывается с CTCF-связывающими фрагментами 10 и 10' (рисунок 12), выявленными методом двумерного EMSA. Программа MACS не выявила этого пика в других образцах, что указывает на тканеспецифичность связывания CTCF в этой области генома. Тканеспецифичность связывания CTCF с фрагментом 10 также была показана методом иммунопреципитации хроматина с последующим анализом иммунопреципитатов ПЦР в реальном времени (см. раздел 3.1.3).

С 3'-стороны от  $\alpha$ -глобиновых генов расположен ген *TMEM8*, уровень транскрипции которого возрастает в ходе эритроидной дифференцировки. Это указывает на действие общих механизмов, регулирующих их транскрипцию, и возможное расположение границы хроматинового домена ниже промотора гена *TMEM8*. Ближайший к генам  $\alpha$ -глобинов пик, расположенный при этом ниже промотора гена *TMEM8*, находится в интроне этого гена (пик 8, рисунок 18). Он выявляется в клетках HD3 и DT40. Ему соответствует область повышенной плотности прочтений в случае клеток HD3-ind. Он является ближайшим к  $\alpha$ -глобиновым генам пиком, лежащим ниже промотора гена *TMEM8* и, таким образом, может соответствовать второй границе функционального хроматинового домена  $\alpha$ -глобиновых генов.

### 3.2.3. Оценка тканеспецифичности связывания CTCF

С помощью программы MAout нами были оценены различия в занятости белком CTCF выявленных в разных типах клеток CTCF-связывающих областей (пиков). При сравнении клеток HD3 и DT40 в области  $\alpha$ -глобинового локуса, разница в занятости CTCF была показана ( $p < 10^{-3}$ ) только для пика, обозначенного на рисунке 18 номером 6. Он перекрывается с фрагментами 10 и 10', выявленными в результате двумерного EMSA (рисунок 12). При сравнении клеток HD3 и HD3-ind с использованием тех же параметров, дифференциально связывающих CTCF участков выявлено не было.

Таким образом, анализ данных ChIP-seq с помощью программы MAnorm, также, как и иммунопреципитация хроматина с анализом ПЦР в реальном времени, свидетельствует в пользу тканеспецифичного связывания белка CTCF с сайтом, находящимся в области фрагментов 10 и 10' (см. раздел 3.1.3.). По-видимому, связывание CTCF с данным сайтом играет роль в регуляции экспрессии глобиновых генов или генов, расположенных рядом с ними, на определенных стадиях эритроидной дифференцировки.

#### 3.2.4. Сравнение результатов ChIP-seq анализа для взрослых и эмбриональных куриных клеток

Мы сравнили данные проведенных нами ChIP-seq экспериментов с данными ChIP-seq, других авторов. Ранее Martin и соавторы опубликовали результаты ChIP-seq с антителами к CTCF для эритробластов пятидневных и десятидневных эмбрионов кур [122]. Мы извлекли из приложенных к данной публикации файлов нуклеотидные последовательности, полученные в результате секвенирования, и картировали их на геном *Gallus gallus* сборки 2011 года. Все прочтения с качеством картирования (MAPQ) меньше 10 были удалены. Было получено распределение плотности картированных прочтений по геному и проведен поиск пиков программой MACS 1.4.2. Координаты пиков приведены в таблицах 5 и 6 приложений.

К публикации Martin и соавторов также были приложены файлы с координатами пиков, выявленных программой SISSRs в геноме кур сборки 2006 г. Геномные координаты пиков, выявленных в  $\alpha$ -глобиновом локусе, были конвертированы в координаты генома сборки 2011 г. (сайты связывания CTCF в эмбриональных эритроблестах 5d1-5d3, 10d1-10d3, см. рисунок 12 и рисунок 18).

С помощью программы MEME мы провели поиск мотивов, которыми был обогащен пул нуклеотидных последовательностей, соответствовавших пикам, выявленным программой MACS1.4.2. С наименьшей значимой вероятностью (*p*-value) в случае обоих образцов эмбриональных клеток выявлялся мотив, сходный с консенсусными последовательностями сайтов связывания CTCF других позвоночных. Данный мотив практически не отличался от консенсусной

последовательности, сайта связывания CTCF кур, выявленной Martin и соавторами [122].

С помощью программы MACS 1.4.2 в обоих типах эмбриональных эритробластов в  $\alpha$ -глобиновом локусе было выявлено четыре пика, 3 из которых перекрывались. Таким образом, вблизи  $\alpha$ -глобиновых генов расположены три участка локуса, связывающие CTCF в эритроблестах эмбрионов разных возрастов (пики 4, 5, 8, рисунок 18).

Расположение пиков, выявленных в  $\alpha$ -глобиновом локусе в случае эритробластов эмбрионов кур, было нами сопоставлено с расположением пиков, выявленных в клетках HD3 и DT40.

Для пиков, выявленных Martin и соавторами, с помощью программы SISR<sub>s</sub>, не было показано перекрываний с пиками, выявленными нашим коллективом с использованием клеток взрослых кур. В то же время некоторые пики, полученные для эмбриональных эритробластов кур с помощью программы MACS1.4.2, перекрываются с пиками, выявленными в клетках взрослых кур. Так, пик, расположенный в интроне гена *TMEM8* в клетках HD3 и DT40, наблюдается и в эмбриональных эритроблестах (пик 8, рисунок 18). В индуцированных клетках HD3 в этом участке локуса также наблюдается повышение плотности прочтений. Таким образом, соответствующий этому пику сайт связывания CTCF взаимодействует с белком как в эмбриональных эритроблестах, так и в клетках кур эритроидного и лимфоидного происхождения.

Один из пиков, общих для эритробластов пятидневного и десятидневного эмбрионов, перекрывается с регуляторным элементом, состоящим из энхансера и сайленсера. На рисунках 18 и 19 он обозначен как пик номер 5. Координаты этого пика, определенные программой MACS1.4.2, в случае эритробластов пятидневного эмбриона, перекрываются также с координатами пика номер 6, выявленного в клетках HD3. Однако, на рисунке 19 видно, что середины пиков, как и участки с максимальной плотностью прочтений внутри пиков находятся вне области перекрытия пиков. В области перекрытия в случае обоих пиков наблюдается довольно невысокая плотность прочтений. Поэтому мы предположили, что данным пикам соответствуют различные сайты связывания CTCF. Путем анализа

иммунопреципитатов хроматина с помощью ПЦР в реальном времени было показано, что занятость белком CTCF падает от середины пика 6, выявленного в клетках HD3, в направлении области перекрытия с пиком 5, выявленным в эритроблестах пятидневных эмбрионов кур (рисунок 19). Следовательно, сайт связывания CTCF, в составе пика 6, выявленного в клетках HD3, не входит в состав пика 5, выявленного в эмбриональных эритроблестах. По-видимому, пики 5 и 6 соответствуют разным сайтам связывания CTCF.

Различия в расположении ближайших к генам  $\alpha$ -глобинов сайтов связывания CTCF, выявленных нами в клетках взрослых кур эритроидного происхождения и выявленных в эмбриональных эритроблестах могут свидетельствовать о важной роли белка CTCF в регуляции экспрессии генов  $\alpha$ -глобинового локуса в ходе индивидуального развития гематопоетической системы кур.

### 3.2.5. Сравнение результатов поиска участков генома, связывающих CTCF *in vitro* и *in vivo*

Мы провели сравнение расположения пиков, выявленных в результате ChIP-seq, с расположением фрагментов ДНК, связывающих CTCF *in vitro* в  $\alpha$ -глобиновом локусе кур.

Часть связывающих CTCF фрагментов ДНК, выявленных в результате двумерного EMSA, перекрывались с пиками, выявленными в результате ChIP-seq. По-видимому, эти участки локуса связывают CTCF и *in vitro* и *in vivo*.

Для многих фрагментов ДНК, связывающих CTCF *in vitro*, связывание *in vivo* показано не было. Возможно, часть этих фрагментов связывает CTCF *in vivo* в клетках других линий. Связывание CTCF *in vivo* определяется многими факторами, влияние которых отсутствует *in vitro*: метилированием цитозина в составе динуклеотида CpG в некоторых CTCF-связывающих последовательностях, структурой хроматина, в том числе модификациями гистонов, наличием вблизи сайтов связывания других факторов транскрипции, облегчающих связывание CTCF.

В то же время некоторая часть фрагментов, связывающих CTCF *in vivo*, не была выявлена при секвенировании библиотеки, полученной в результате

двумерного EMSA. Возможно, эти фрагменты имеют не очень высокую представленность в библиотеке, в результате меньшего сродства к CTCF по сравнению с другими фрагментами библиотеки. В этом случае они могли отбираться в ходе двумерного EMSA, но не выявляться в результате секвенирования конечной библиотеки, как, например, фрагмент M9 (обогащение этим фрагментом библиотеки было показано путем ПЦР, но он не попал в число секвенированных). Кроме того, часть фрагментов ДНК, связывающих CTCF, могла быть утеряна в результате ПЦР-селекции. Так, некоторые GC-богатые фрагменты ДНК не амплифицируются даже в присутствии ДМСО и бетаина [197], а часть сайтов связывания CTCF входит в состав GC-богатых участков генома.

Chr12:11957588-12184953 (galGal4)

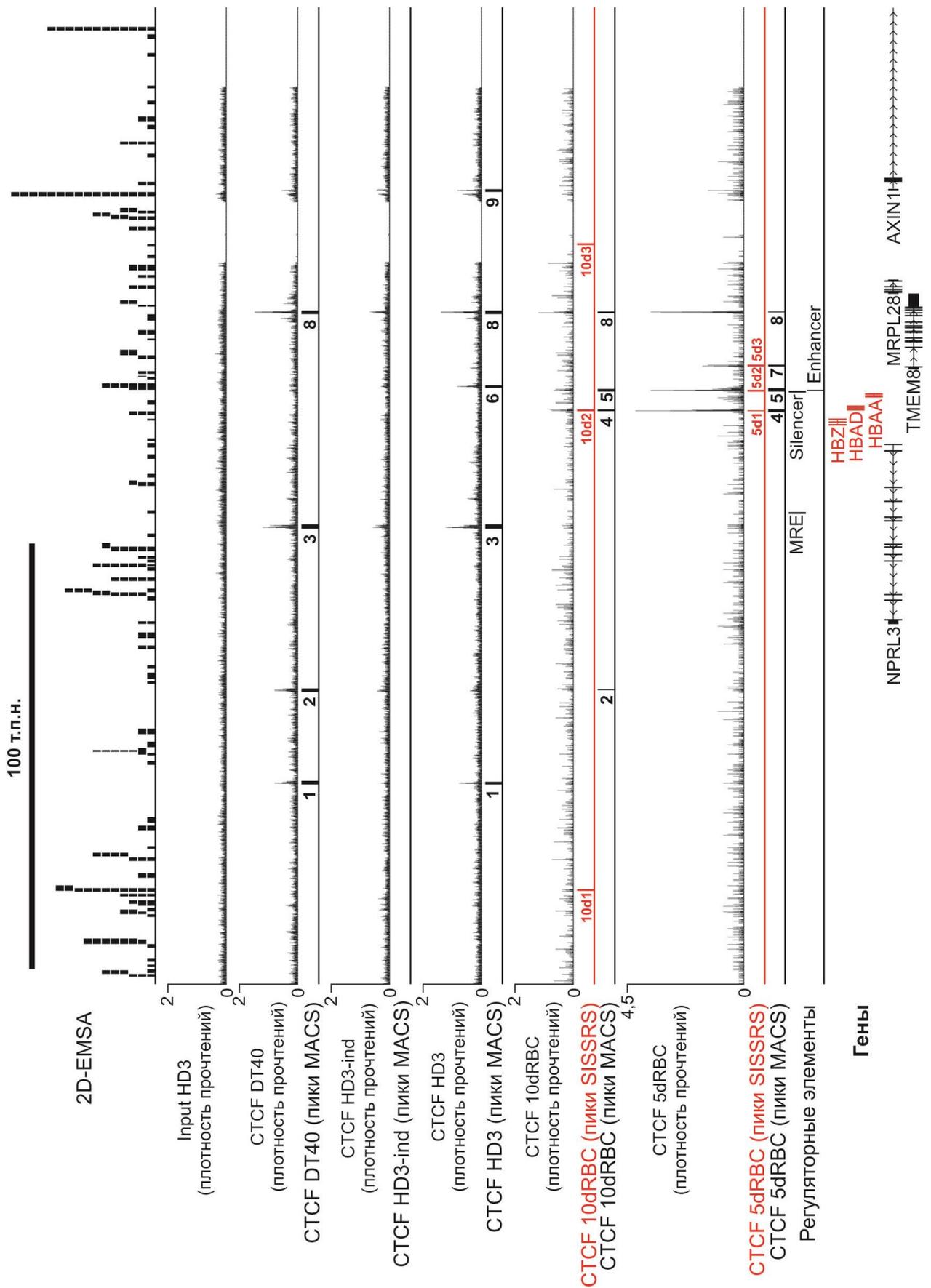
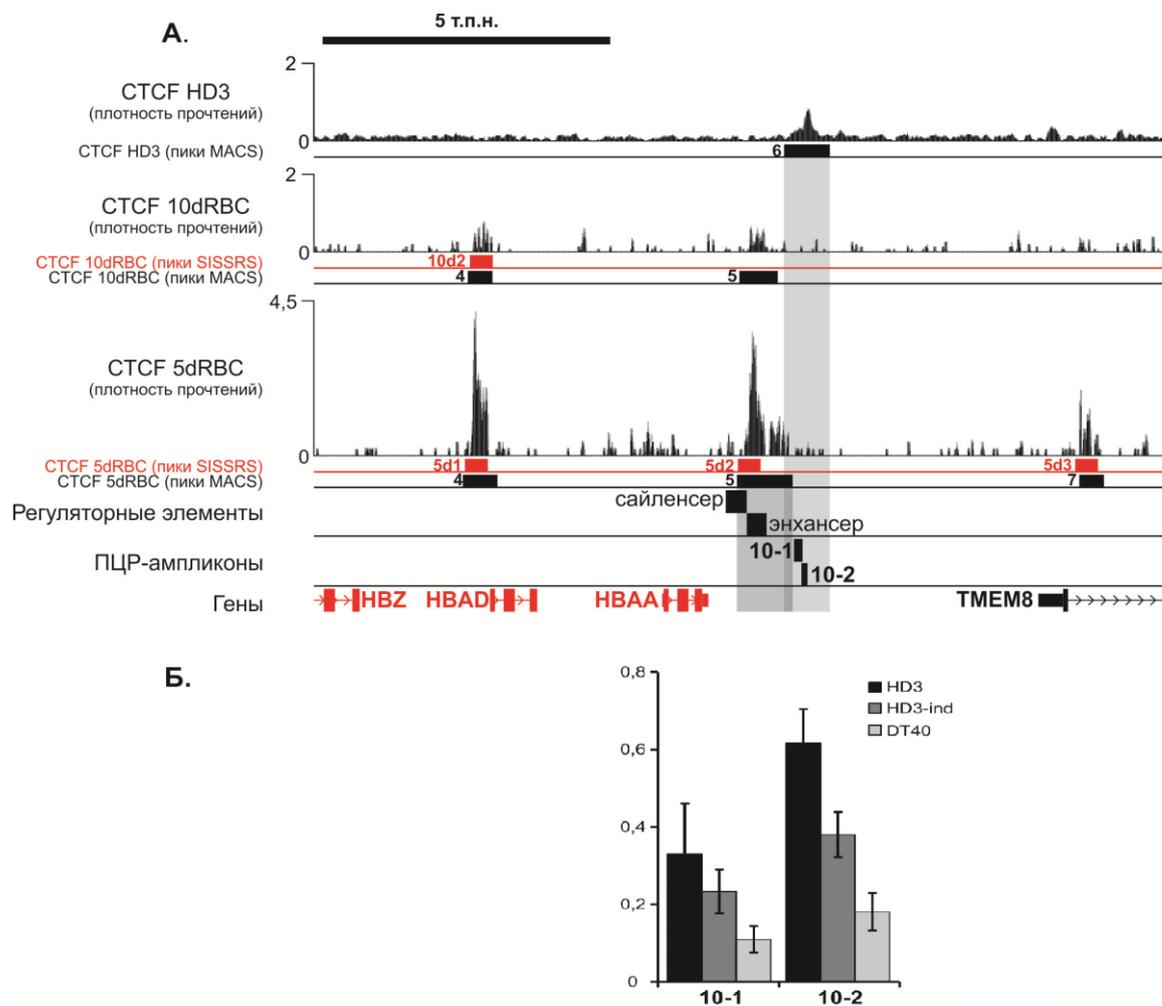


Рисунок 18. Расположение STCF-связывающих последовательностей в  $\alpha$ -глобиновом локусе кур. 2D-EMSA – STCF-связывающие последовательности, выявленные путем двумерного EMSA. Input – вся

выделенная после ультразвуковой дезинтеграции ДНК. CTCF –ДНК, преципитированная с антителами к CTCF. DT40, HD3-ind, HD3, 10dRBC, 5dRBC – клетки линии DT40, HD3, HD3 после индукции дифференцировки, эритробласты пятидневного и десятидневного эмбриона кур, соответственно. Данные для эритробластов из эмбрионов кур получены в результате обработки результатов ChIP-seq экспериментов, опубликованных ранее. Плотность прочтений – количество прочтений, перекрывающихся с данным нуклеотидом в геноме, на миллион прочтений, полученных при секвенировании образца. Пики MACS – расположение пиков, выявленных при анализе данных ChIP-seq программой MACS. Пики SISRAs – расположение пиков, выявленных программой SISRAs [122]. Пики, выявленные в  $\alpha$ -глобиновом локусе, в разных типах клеток с использованием программы MACS 1.4.2 пронумерованы с возрастанием номера в направлении возрастания геномных координат. Большинство перекрывающихся пиков имеет один и тот же номер, за исключением пиков 5 и 6, которым, по-видимому, соответствуют разные сайты связывания CTCF (см. “Сравнение результатов ChIP-seq анализа для взрослых и эмбриональных куриных клеток”, раздел 3.2.4.). Регуляторные элементы – регуляторные элементы локуса: MRE, энхансер и сайленсер. Гены – гены локуса.



**Рисунок 19. Связывание белка CTCF в области сайленсера- энхансера  $\alpha$ -глобинового локуса в клетках линии HD3 и эмбриональных эритроблестах кур. А. Карта участка  $\alpha$ -глобинового локуса, содержащего регуляторный элемент, состоящий из сайленсера и энхансера. Показано расположение генов (дорожка “Гены”), регуляторных элементов (дорожка “Регуляторные элементы”), распределение плотности прочтений (плотность прочтений) и пиков сигнала (пики MACS и пики SISR) по результатам ChIP-seq с антителами к CTCF. Обозначения дорожек, содержащих данные ChIP-seq и пики – см. рисунок 18. Также на карте отмечено расположение участков ДНК, которые амплифицировались во время анализа иммунопреципитатов хроматина ПЦР в реальном времени (“ПЦР-ампликоны”). Б. Результаты анализа иммунопреципитатов хроматина методом ПЦР в реальном времени. Расположение амплифицированных областей на карте локуса показано под буквой А – дорожка “ПЦР-ампликоны”. Высота столбиков — процент**

**содержания данного фрагмента в иммунопреципитате от содержания  
данного фрагмента в образце Input (тотальная геномная ДНК) - %Input.**

## ЗАКЛЮЧЕНИЕ И ВЫВОДЫ

### Заключение

В работе была получена белковая фракция, обогащенная куриным CTCF с N-концевым гистидиновым тагом, специфично взаимодействующим с CTCF-связывающими фрагментами ДНК. Был проведен поиск фрагментов  $\alpha$ -глобинового локуса кур, способных связывать CTCF *in vitro*, методом двумерного EMSA (Electrophoretic Mobility Shift Assay). Методом иммунопреципитации хроматина было изучено связывание CTCF с этими фрагментами *in vivo* в геноме эритроидных и лимфоидных клеток кур. Был проведен полногеномный поиск участков ДНК, связывающих CTCF *in vivo*, в геноме эритроидных и лимфоидных клеток кур методом ChIP-seq (иммунопреципитация хроматина с последующим масштабным секвенированием). В области  $\alpha$ -глобиновых генов было выявлено два сайта связывания CTCF, общих для эритроидных и лимфоидных неэмбриональных куриных клеток. Один из них также связывает белок CTCF в эмбриональных эритроблестах кур. Кроме того, был выявлен сайт связывания CTCF, имеющий высокую занятость белком только в клетках эритроидного происхождения взрослых кур.

### Выводы

1. Получена белковая фракция, обогащенная белком CTCF с N-концевым гистидиновым тагом, синтезированным в клетках позвоночных, способным специфично взаимодействовать с CTCF-связывающими фрагментами ДНК.
2. Методом двумерного EMSA выявлено 79 участков ДНК  $\alpha$ -глобинового локуса кур, способных связывать белок CTCF *in vitro*. При помощи иммунопреципитации хроматина показано, что из тринадцати участков ДНК,

непосредственно прилегающих к генам глобина, только один связывает CTCF *in vivo*. Таким образом, связывание CTCF с участком ДНК *in vitro* по большей части не означает, что этот участок будет занят CTCF в ядре клетки.

3. Методом ChIP-seq был проведен полногеномный поиск областей, связывающих CTCF *in vivo*, в клетках линии HD3, HD3 после индукции эритроидной дифференцировки, и в клетках лимфоидной линии DT40. Показано, что в индуцированных к эритроидной дифференцировке клетках HD3 средняя занятость белком CTCF его сайтов связывания примерно в 1,5 раза ниже, чем в интактных. В области  $\alpha$ -глобиновых генов было выявлено два сайта связывания CTCF, общих для эритроидных и лимфоидных неэмбриональных куриных клеток. Один из них также связывает белок CTCF в эмбриональных эритроблестах кур. Кроме того, был выявлен сайт связывания CTCF, имеющий высокую занятость белком только в клетках эритроидного происхождения взрослых кур.

## СПИСОК СОКРАЩЕНИЙ

- 3C (Chromosome Conformation Capture) – фиксация конформации хромосомы
- 5'HS4 (5' chicken  $\beta$ -globin DNase I Hypersensitive Site 4) – участок сверхчувствительный к расщеплению ДНКазой I, расположенный рядом с 5'-концом  $\beta$ -глобинового локуса кур, инсулятор
- APV $\beta$  – сайт связывания CTCF, расположенный в промоторе гена предшественника  $\beta$ -амилоида человека
- APP – ген, кодирующий предшественник  $\beta$ -амилоида (amyloid  $\beta$  protein precursor)
- CDS (CTCF Dependent Silencer) – CTCF-зависимый сайленсер
- ChIA-PET (Chromatin Interaction Analysis by Paired-End Tag Sequencing) – метод выявления хроматиновых взаимодействий
- ChIP (Chromatin ImmunoPrecipitation) - иммунопреципитация хроматина
- ChIP-chip (Chromatin ImmunoPrecipitation on chip) – иммунопреципитация хроматина с использованием чипов для анализа иммунопреципитированной ДНК
- Chip-ехо – иммунопреципитация хроматина, в ходе которой проводится обработкой образцов экзонуклеазой
- ChIP-seq (Chromatin ImmunoPrecipitation Sequencing) – иммунопреципитация хроматина с последующим масштабным секвенированием
- CTCF (CCTC-binding Factor) - фактор, связывающийся с последовательностью CCTC
- EMSA (Electrophoretic Mobility Shift Assay) - метод сдвига электрофоретической подвижности
- F1 – F1-блок лизоцимного сайленсера кур
- HBAD - фрагмент гена альфа-глобина D
- HD3-ind – клетки линии HD3, индуцированные к дифференцировке
- HEPES – N-2-гидроксиэтилпиперазин-N'-2-этилсульфоновая кислота
- Hi-C (High-throughput 3C) – метод полногеномного выявления хроматиновых взаимодействий

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

ICR (Imprinting Control Region) – регуляторный элемент, контролирующий геномный импринтинг

IPTG - изопропил-бета-D-1-тиогалактопиранозид

MRE (Major Regulatory Element) – главный регуляторный элемент

MYC – сайт связывания CTCF в составе промотора гена *c-myc* кур

mys – сайт связывания CTCF в составе промотора гена *c-myc* человека

PBS (Phosphate Buffered Saline) - фосфатно-солевой буфер

PMSF – фенилметилсульфотригид

SDS - додецилсульфат натрия

SV40 (Semian Virus 40) – обезьяний вирус 40

TAD (Topologically Associating Domain) – топологически ассоциированный домен

TBE - трис-боратный буфер с добавлением ЭДТА

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

TE - трис-ЭДТА буфер

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

BSA - бычий сывороточный альбумин

Да – атомная единица массы

ДДС – додецилсульфат натрия

ДТТ – дитиотрейтол

кДа – килодальтон

ПААГ – полиакриламидный гель

ПЦР – полимеразная цепная реакция

п.о. - пар оснований

т.п.о. - тысяч пар оснований

ЭДТА - этилендиаминтетрауксусная кислота

ЯЭ – ядерный экстракт

## СПИСОК ЛИТЕРАТУРЫ

1. Afgan, E., D. Baker, et al. "The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update." *Nucleic Acids Res.*
2. Akopov, S. B., V. M. Ruda, et al. (2006). "Identification, genome mapping, and CTCF binding of potential insulators within the FXYD5-COX7A1 locus of human chromosome 19q13.12." *Mamm Genome* **17**(10): 1042-9.
3. Anguita, E., C. A. Johnson, et al. (2001). "Identification of a conserved erythroid specific domain of histone acetylation across the alpha-globin gene cluster." *Proc Natl Acad Sci U S A* **98**(21): 12114-9.
4. Arnold, R., M. Burcin, et al. (1996). "DNA bending by the silencer protein NeP1 is modulated by TR and RXR." *Nucleic Acids Res* **24**(14): 2640-7.
5. Avner, P. and E. Heard (2001). "X-chromosome inactivation: counting, choice and initiation." *Nat Rev Genet* **2**(1): 59-67.
6. Bacher, C. P., M. Guggiari, et al. (2006). "Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation." *Nat Cell Biol* **8**(3): 293-9.
7. Baniahmad, A., C. Steiner, et al. (1990). "Modular structure of a chicken lysozyme silencer: involvement of an unusual thyroid hormone receptor binding site." *Cell* **61**(3): 505-14.
8. Bao, L., M. Zhou, et al. (2008). "CTCFBSDB: a CTCF-binding site database for characterization of vertebrate genomic insulators." *Nucleic Acids Res* **36**(Database issue): D83-7.
9. Barski, A., S. Cuddapah, et al. (2007). "High-resolution profiling of histone methylations in the human genome." *Cell* **129**(4): 823-37.
10. Bell, A. C. and G. Felsenfeld (2000). "Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene." *Nature* **405**(6785): 482-5.
11. Bell, A. C., A. G. West, et al. (1999). "The protein CTCF is required for the enhancer blocking activity of vertebrate insulators." *Cell* **98**(3): 387-96.
12. Bell, A. C., A. G. West, et al. (2001). "Insulators and boundaries: versatile regulatory elements in the eukaryotic genome." *Science* **291**(5503): 447-50.
13. Beug, H., A. von Kirchbach, et al. (1979). "Chicken hematopoietic cells transformed by seven strains of defective avian leukemia viruses display three distinct phenotypes of differentiation." *Cell* **18**(2): 375-90.
14. Bodnar, J. W. (1988). "A domain model for eukaryotic DNA organization: a molecular basis for cell differentiation and chromosome evolution." *J Theor Biol* **132**(4): 479-507.
15. Bulger, M., M. A. Bender, et al. (2000). "Comparative structural and functional analysis of the olfactory receptor genes flanking the human and mouse beta-globin gene clusters." *Proc Natl Acad Sci U S A* **97**(26): 14560-5.
16. Burcin, M., R. Arnold, et al. (1997). "Negative protein 1, which is required for function of the chicken lysozyme gene silencer in conjunction with hormone receptors, is identical to the multivalent zinc finger repressor CTCF." *Mol Cell Biol* **17**(3): 1281-8.

17. Chambeyron, S. and W. A. Bickmore (2004). "Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription." *Genes Dev* **18**(10): 1119-30.
18. Chao, W., K. D. Huynh, et al. (2002). "CTCF, a candidate trans-acting factor for X-inactivation choice." *Science* **295**(5553): 345-7.
19. Chen, X., H. Xu, et al. (2008). "Integration of external signaling pathways with the core transcriptional network in embryonic stem cells." *Cell* **133**(6): 1106-17.
20. Chernukhin, I., S. Shamsuddin, et al. (2007). "CTCF interacts with and recruits the largest subunit of RNA polymerase II to CTCF target sites genome-wide." *Mol Cell Biol* **27**(5): 1631-48.
21. Chernukhin, I. V., S. Shamsuddin, et al. (2000). "Physical and functional interaction between two pluripotent proteins, the Y-box DNA/RNA-binding factor, YB-1, and the multivalent zinc finger factor, CTCF." *J Biol Chem* **275**(38): 29915-21.
22. Chien, R., W. Zeng, et al. (2011). "Cohesin mediates chromatin interactions that regulate mammalian beta-globin expression." *J Biol Chem* **286**(20): 17870-8.
23. Chung, J. H., A. C. Bell, et al. (1997). "Characterization of the chicken beta-globin insulator." *Proc Natl Acad Sci U S A* **94**(2): 575-80.
24. Ciabrelli, F. and G. Cavalli "Chromatin-driven behavior of topologically associating domains." *J Mol Biol* **427**(3): 608-25.
25. Clerc, P. and P. Avner (2006). "Random X-chromosome inactivation: skewing lessons for mice and men." *Curr Opin Genet Dev* **16**(3): 246-53.
26. Craddock, C. F., P. Vyas, et al. (1995). "Contrasting effects of alpha and beta globin regulatory elements on chromatin structure may be related to their different chromosomal environments." *EMBO J* **14**(8): 1718-26.
27. Cuddapah, S., R. Jothi, et al. (2009). "Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains." *Genome Res* **19**(1): 24-32.
28. De La Rosa-Velazquez, I. A., H. Rincon-Arano, et al. (2007). "Epigenetic regulation of the human retinoblastoma tumor suppressor gene promoter by CTCF." *Cancer Res* **67**(6): 2577-85.
29. de Wit, E., E. S. Vos, et al. "CTCF Binding Polarity Determines Chromatin Looping." *Mol Cell* **60**(4): 676-84.
30. Degner, S. C., J. Verma-Gaur, et al. (2011). "CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the Igh locus and antisense transcription in pro-B cells." *Proc Natl Acad Sci U S A* **108**(23): 9566-71.
31. Didych, D. A., E. S. Kotova, et al. (2012). "DNA fragments binding CTCF in vitro and in vivo are capable of blocking enhancer activity." *BMC Res Notes* **5**: 178.
32. Dixon, J. R., S. Selvaraj, et al. (2012). "Topological domains in mammalian genomes identified by analysis of chromatin interactions." *Nature* **485**(7398): 376-80.
33. Donohoe, M. E., L. F. Zhang, et al. (2007). "Identification of a Ctfc cofactor, Yy1, for the X chromosome binary switch." *Mol Cell* **25**(1): 43-56.

34. Dunn, K. L., H. Zhao, et al. (2003). "The insulator binding protein CTCF associates with the nuclear matrix." *Exp Cell Res* **288**(1): 218-23.
35. ENCODE (2012). "An integrated encyclopedia of DNA elements in the human genome." *Nature* **489**(7414): 57-74.
36. Engel, N., A. G. West, et al. (2004). "Antagonism between DNA hypermethylation and enhancer-blocking activity at the H19 DMD is uncovered by CpG mutations." *Nat Genet* **36**(8): 883-8.
37. Erwin, J. A. and J. T. Lee (2008). "New twists in X-chromosome inactivation." *Curr Opin Cell Biol* **20**(3): 349-55.
38. Farrar, D., S. Rai, et al. "Mutational analysis of the poly(ADP-ribosylation) sites of the transcription factor CTCF provides an insight into the mechanism of its regulation by poly(ADP-ribosylation)." *Mol Cell Biol* **30**(5): 1199-216.
39. Fedoriw, A. M., P. Stein, et al. (2004). "Transgenic RNAi reveals essential function for CTCF in H19 gene imprinting." *Science* **303**(5655): 238-40.
40. Felsenfeld, G. (1993). "Chromatin structure and the expression of globin-encoding genes." *Gene* **135**(1-2): 119-24.
41. Ferrai, C., S. Q. Xie, et al. (2010). "Poised transcription factories prime silent uPA gene prior to activation." *PLoS Biol* **8**(1): e1000270.
42. Filippova, G. N., M. K. Cheng, et al. (2005). "Boundaries between chromosomal domains of X inactivation and escape bind CTCF and lack CpG methylation during early development." *Dev Cell* **8**(1): 31-42.
43. Filippova, G. N., S. Fagerlie, et al. (1996). "An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes." *Mol Cell Biol* **16**(6): 2802-13.
44. Filippova, G. N., C. P. Thienes, et al. (2001). "CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus." *Nat Genet* **28**(4): 335-43.
45. Flint, J., C. Tufarelli, et al. (2001). "Comparative genome analysis delimits a chromosomal domain and identifies key regulatory elements in the alpha globin cluster." *Hum Mol Genet* **10**(4): 371-82.
46. Forrester, W. C., E. Epner, et al. (1990). "A deletion of the human beta-globin locus activation region causes a major alteration in chromatin structure and replication across the entire beta-globin locus." *Genes Dev* **4**(10): 1637-49.
47. Fried, M. and D. M. Crothers (1981). "Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis." *Nucleic Acids Res* **9**(23): 6505-25.
48. Gabory, A., H. Jammes, et al. (2010). "The H19 locus: role of an imprinted non-coding RNA in growth and development." *Bioessays* **32**(6): 473-80.
49. Gao, Z. H., S. Suppola, et al. (2002). "Association of H19 promoter methylation with the expression of H19 and IGF-II genes in adrenocortical tumors." *J Clin Endocrinol Metab* **87**(3): 1170-6.
50. Gavrilov, A. A. and S. V. Razin (2008). "Spatial configuration of the chicken alpha-globin gene domain: immature and active chromatin hubs." *Nucleic Acids Res* **36**(14): 4629-40.

51. Geyer, P. K. (1997). "The role of insulator elements in defining domains of gene expression." *Curr Opin Genet Dev* **7**(2): 242-8.
52. Ghirlando, R. and G. Felsenfeld "CTCF: making the right connections." *Genes Dev* **30**(8): 881-91.
53. Gohl, D., T. Aoki, et al. (2011). "Mechanism of chromosomal boundary action: roadblock, sink, or loop?" *Genetics* **187**(3): 731-48.
54. Goldman, M. A. (1988). "The chromatin domain as a unit of gene regulation." *Bioessays* **9**(2-3): 50-5.
55. Gomes, N. P. and J. M. Espinosa (2010). "Gene-specific repression of the p53 target gene PUMA via intragenic CTCF-Cohesin binding." *Genes Dev* **24**(10): 1022-34.
56. Grdisa, M. and M. K. White (2003). "Molecular and biochemical events during differentiation of the HD3 chicken erythroblastic cell line." *Int J Biochem Cell Biol* **35**(4): 422-31.
57. Grosveld, F., M. Antoniou, et al. (1987). "The regulation of expression of human beta-globin genes." *Prog Clin Biol Res* **251**: 133-44.
58. Guastafierro, T., B. Cecchinelli, et al. (2008). "CCCTC-binding factor activates PARP-1 affecting DNA methylation machinery." *J Biol Chem* **283**(32): 21873-80.
59. Guo, Y., K. Monahan, et al. (2012). "CTCF/cohesin-mediated DNA looping is required for protocadherin alpha promoter choice." *Proc Natl Acad Sci U S A* **109**(51): 21081-6.
60. Guo, Y., Q. Xu, et al. "CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function." *Cell* **162**(4): 900-10.
61. Hadjur, S., L. M. Williams, et al. (2009). "Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus." *Nature* **460**(7253): 410-3.
62. Haering, C. H., J. Lowe, et al. (2002). "Molecular architecture of SMC proteins and the yeast cohesin complex." *Mol Cell* **9**(4): 773-88.
63. Handoko, L., H. Xu, et al. (2011). "CTCF-mediated functional chromatin interactome in pluripotent cells." *Nat Genet* **43**(7): 630-8.
64. Hark, A. T., C. J. Schoenherr, et al. (2000). "CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus." *Nature* **405**(6785): 486-9.
65. Harmston, N. and B. Lenhard (2013). "Chromatin and epigenetic features of long-range gene regulation." *Nucleic Acids Res* **41**(15): 7185-99.
66. Heath, H., C. Ribeiro de Almeida, et al. (2008). "CTCF regulates cell cycle progression of alphabeta T cells in the thymus." *EMBO J* **27**(21): 2839-50.
67. Hebbes, T. R., A. L. Clayton, et al. (1994). "Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain." *EMBO J* **13**(8): 1823-30.
68. Heger, P., B. Marin, et al. (2012). "The chromatin insulator CTCF and the emergence of metazoan diversity." *Proc Natl Acad Sci U S A* **109**(43): 17507-12.
69. Herold, M., M. Bartkuhn, et al. (2012). "CTCF: insights into insulator function during development." *Development* **139**(6): 1045-57.

70. Hirano, T. (2006). "At the heart of the chromosome: SMC proteins in action." *Nat Rev Mol Cell Biol* **7**(5): 311-22.
71. Hoki, Y., N. Kimura, et al. (2009). "A proximal conserved repeat in the Xist gene is essential as a genomic element for X-inactivation in mouse." *Development* **136**(1): 139-46.
72. Holwerda, S. J. and W. de Laat (2013). "CTCF: the protein, the binding partners, the binding sites and their chromatin loops." *Philos Trans R Soc Lond B Biol Sci* **368**(1620): 20120369.
73. Hu, Q., Y. S. Kwon, et al. (2008). "Enhancing nuclear receptor-induced transcription requires nuclear motor and LSD1-dependent gene networking in interchromatin granules." *Proc Natl Acad Sci U S A* **105**(49): 19199-204.
74. Huang, K., J. Jia, et al. (2013). "Ribosomal RNA gene transcription mediated by the master genome regulator protein CCCTC-binding factor (CTCF) is negatively regulated by the condensin complex." *J Biol Chem* **288**(36): 26067-77.
75. Iarovaia, O., S. V. Razin, et al. (2001). "In chicken leukemia cells globin genes are fully transcribed but their rnas are retained in the perinucleolar area." *Exp Cell Res* **270**(2): 159-65.
76. Inoue, H., H. Nojima, et al. (1990). "High efficiency transformation of *Escherichia coli* with plasmids." *Gene* **96**(1): 23-8.
77. Ishiguro, K. and Y. Watanabe (2007). "Chromosome cohesion in mitosis and meiosis." *J Cell Sci* **120**(Pt 3): 367-9.
78. Ishihara, K., M. Oshimura, et al. (2006). "CTCF-dependent chromatin insulator is linked to epigenetic remodeling." *Mol Cell* **23**(5): 733-42.
79. Kagey, M. H., J. J. Newman, et al. (2010). "Mediator and cohesin connect gene expression and chromatin architecture." *Nature* **467**(7314): 430-5.
80. Kanduri, C., C. Holmgren, et al. (2000). "The 5' flank of mouse H19 in an unusual chromatin conformation unidirectionally blocks enhancer-promoter communication." *Curr Biol* **10**(8): 449-57.
81. Kanduri, C., V. Pant, et al. (2000). "Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive." *Curr Biol* **10**(14): 853-6.
82. Karolchik, D., A. S. Hinrichs, et al. (2004). "The UCSC Table Browser data retrieval tool." *Nucleic Acids Res* **32**(Database issue): D493-6.
83. Kehayova, P., K. Monahan, et al. (2011). "Regulatory elements required for the activation and repression of the protocadherin-alpha gene cluster." *Proc Natl Acad Sci U S A* **108**(41): 17195-200.
84. Kent, W. J. (2002). "BLAT--the BLAST-like alignment tool." *Genome Res* **12**(4): 656-64.
85. Kent, W. J., C. W. Sugnet, et al. (2002). "The human genome browser at UCSC." *Genome Res* **12**(6): 996-1006.
86. Kim, T. H., Z. K. Abdullaev, et al. (2007). "Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome." *Cell* **128**(6): 1231-45.
87. Kitchen, N. S. and C. J. Schoenherr (2010). "Sumoylation modulates a domain in CTCF that activates transcription and decondenses chromatin." *J Cell Biochem* **111**(3): 665-75.

88. Klenova, E. and R. Ohlsson (2005). "Poly(ADP-ribosyl)ation and epigenetics. Is CTCF PART of the plot?" *Cell Cycle* **4**(1): 96-101.
89. Klenova, E. M., I. V. Chernukhin, et al. (2001). "Functional phosphorylation sites in the C-terminal region of the multivalent multifunctional transcriptional factor CTCF." *Mol Cell Biol* **21**(6): 2221-34.
90. Klenova, E. M., R. H. Nicolas, et al. (1993). "CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms." *Mol Cell Biol* **13**(12): 7612-24.
91. Klenova, E. M., R. H. Nicolas, et al. (1997). "Molecular weight abnormalities of the CTCF transcription factor: CTCF migrates aberrantly in SDS-PAGE and the size of the expressed protein is affected by the UTRs and sequences within the coding region of the CTCF gene." *Nucleic Acids Res* **25**(3): 466-74.
92. Klochkov, D., H. Rincon-Arango, et al. (2006). "A CTCF-dependent silencer located in the differentially methylated area may regulate expression of a housekeeping gene overlapping a tissue-specific gene domain." *Mol Cell Biol* **26**(5): 1589-97.
93. Klug, A. (2010). "The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation." *Q Rev Biophys* **43**(1): 1-21.
94. Kohne, A. C., A. Baniahmad, et al. (1993). "NeP1. A ubiquitous transcription factor synergizes with v-ERBA in transcriptional silencing." *J Mol Biol* **232**(3): 747-55.
95. Kolpakov, F. (2004). "BioUML—open source extensible workbench for systems biology. ." *Proceedings of The Fourth International Conference on Bioinformatics of Genome Regulation and Structure, Novosibirsk, Russia.*: 77–80.
96. Kolpakov, F., M. Puzanov, et al. (2006). "BioUML: visual modeling, automated code generation and simulation of biological systems." *Proceedings of The Fifth International Conference on Bioinformatics of Genome Regulation and Structure, Novosibirsk, Russia*: 281–285.
97. Kornblihtt, A. R. (2012). "CTCF: from insulators to alternative splicing regulation." *Cell Res* **22**(3): 450-2.
98. Kowalski, A. and J. Palyga "Chromatin compaction in terminally differentiated avian blood cells: the role of linker histone H5 and non-histone protein MENT." *Chromosome Res* **19**(5): 579-90.
99. Krajewski, W. A. and P. B. Becker (1998). "Reconstitution of hyperacetylated, DNase I-sensitive chromatin characterized by high conformational flexibility of nucleosomal DNA." *Proc Natl Acad Sci U S A* **95**(4): 1540-5.
100. Kristie, T. M. and B. Roizman (1986). "Alpha 4, the major regulatory protein of herpes simplex virus type 1, is stably and specifically associated with promoter-regulatory domains of alpha genes and of selected other viral genes." *Proc Natl Acad Sci U S A* **83**(10): 3218-22.
101. Krivega, I. and A. Dean (2012). "Enhancer and promoter interactions-long distance calls." *Curr Opin Genet Dev* **22**(2): 79-85.

102. Kulaeva, O. I., E. V. Nizovtseva, et al. (2012). "Distant activation of transcription: mechanisms of enhancer action." *Mol Cell Biol* **32**(24): 4892-7.
103. Kurukuti, S., V. K. Tiwari, et al. (2006). "CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2." *Proc Natl Acad Sci U S A* **103**(28): 10684-9.
104. Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." *Nature* **227**(5259): 680-5.
105. Laemmli, U. K. and M. Favre (1973). "Maturation of the head of bacteriophage T4. I. DNA packaging events." *J Mol Biol* **80**(4): 575-99.
106. Lai, A. Y., M. Fatemi, et al. (2010). "DNA methylation prevents CTCF-mediated silencing of the oncogene BCL6 in B cell lymphomas." *J Exp Med* **207**(9): 1939-50.
107. Langmead, B. and S. L. Salzberg "Fast gapped-read alignment with Bowtie 2." *Nat Methods* **9**(4): 357-9.
108. Lee, J. T. (2009). "Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome." *Genes Dev* **23**(16): 1831-42.
109. Leighton, P. A., J. R. Saam, et al. (1995). "An enhancer deletion affects both H19 and Igf2 expression." *Genes Dev* **9**(17): 2079-89.
110. Lewis, A. and A. Murrell (2004). "Genomic imprinting: CTCF protects the boundaries." *Curr Biol* **14**(7): R284-6.
111. Li, G., X. Ruan, et al. (2012). "Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation." *Cell* **148**(1-2): 84-98.
112. Litt, M. D., M. Simpson, et al. (2001). "Transitions in histone acetylation reveal boundaries of three separately regulated neighboring loci." *EMBO J* **20**(9): 2224-35.
113. Lobanekov, V. V. and G. G. Gudvin (1989). "[CCCTC-binding protein: a new nuclear protein factor which interaction with 5'-flanking sequence of chicken c-myc oncogene correlates with repression of the gene]." *Dokl Akad Nauk SSSR* **309**(3): 741-5.
114. Lobanekov, V. V., R. H. Nicolas, et al. (1990). "A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene." *Oncogene* **5**(12): 1743-53.
115. Lutz, M., L. J. Burke, et al. (2000). "Transcriptional repression by the insulator protein CTCF involves histone deacetylases." *Nucleic Acids Res* **28**(8): 1707-13.
116. Lutz, M., L. J. Burke, et al. (2003). "Thyroid hormone-regulated enhancer blocking: cooperation of CTCF and thyroid hormone receptor." *EMBO J* **22**(7): 1579-87.
117. Machanick, P. and T. L. Bailey "MEME-ChIP: motif analysis of large DNA datasets." *Bioinformatics* **27**(12): 1696-7.
118. MacPherson, M. J., L. G. Beatty, et al. (2009). "The CTCF insulator protein is posttranslationally modified by SUMO." *Mol Cell Biol* **29**(3): 714-25.

119. MacPherson, M. J. and P. D. Sadowski (2010). "The CTCF insulator protein forms an unusual DNA structure." *BMC Mol Biol* **11**: 101.
120. Magdinier, F., T. M. Yusufzai, et al. (2004). "Both CTCF-dependent and -independent insulators are found between the mouse T cell receptor alpha and Dad1 genes." *J Biol Chem* **279**(24): 25381-9.
121. Mahy, N. L., P. E. Perry, et al. (2002). "Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH." *J Cell Biol* **159**(5): 753-63.
122. Martin, D., C. Pantoja, et al. "Genome-wide CTCF distribution in vertebrates defines equivalent sites that aid the identification of disease-associated genes." *Nat Struct Mol Biol* **18**(6): 708-14.
123. Mendez-Catala, C. F., S. Gretton, et al. (2013). "A novel mechanism for CTCF in the epigenetic regulation of Bax in breast cancer cells." *Neoplasia* **15**(8): 898-912.
124. Mitchell, J. A. and P. Fraser (2008). "Transcription factories are nuclear subcompartments that remain in the absence of transcription." *Genes Dev* **22**(1): 20-5.
125. Monahan, K., N. D. Rudnick, et al. (2012). "Role of CCCTC binding factor (CTCF) and cohesin in the generation of single-cell diversity of protocadherin-alpha gene expression." *Proc Natl Acad Sci U S A* **109**(23): 9125-30.
126. Nakahashi, H., K. R. Kwon, et al. (2013). "A genome-wide map of CTCF multivalency redefines the CTCF code." *Cell Rep* **3**(5): 1678-89.
127. Nicodemi, M., B. Panning, et al. (2008). "A thermodynamic switch for chromosome colocalization." *Genetics* **179**(1): 717-21.
128. Nicodemi, M. and A. Prisco (2007). "Self-assembly and DNA binding of the blocking factor in x chromosome inactivation." *PLoS Comput Biol* **3**(11): e210.
129. Nicodemi, M. and A. Prisco (2007). "Symmetry-breaking model for X-chromosome inactivation." *Phys Rev Lett* **98**(10): 108104.
130. Nicolas, R. H., G. Partington, et al. (1991). "Induction of differentiation of avian erythroblastosis virus-transformed erythroblasts by the protein kinase inhibitor H7: analysis of the transcription factor EF1." *Cell Growth Differ* **2**(3): 129-35.
131. Nikolaev, L. G., S. B. Akopov, et al. (2009). "Vertebrate Protein CTCF and its Multiple Roles in a Large-Scale Regulation of Genome Activity." *Curr Genomics* **10**(5): 294-302.
132. Ohlsson, R., R. Renkawitz, et al. (2001). "CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease." *Trends Genet* **17**(9): 520-7.
133. Orkin, S. H., F. I. Harosi, et al. (1975). "Differentiation in erythroleukemic cells and their somatic hybrids." *Proc Natl Acad Sci U S A* **72**(1): 98-102.
134. Orlando, V. (2000). "Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation." *Trends Biochem Sci* **25**(3): 99-104.
135. Osborne, C. S., L. Chakalova, et al. (2004). "Active genes dynamically colocalize to shared sites of ongoing transcription." *Nat Genet* **36**(10): 1065-71.
136. Pabo, C. O., E. Peisach, et al. (2001). "Design and selection of novel Cys2His2 zinc finger proteins." *Annu Rev Biochem* **70**: 313-40.

137. Palstra, R. J., B. Tolhuis, et al. (2003). "The beta-globin nuclear compartment in development and erythroid differentiation." *Nat Genet* **35**(2): 190-4.
138. Pant, V., P. Mariano, et al. (2003). "The nucleotides responsible for the direct physical contact between the chromatin insulator protein CTCF and the H19 imprinting control region manifest parent of origin-specific long-distance insulation and methylation-free domains." *Genes Dev* **17**(5): 586-90.
139. Parelho, V., S. Hadjur, et al. (2008). "Cohesins functionally associate with CTCF on mammalian chromosome arms." *Cell* **132**(3): 422-33.
140. Patel, B., Y. Kang, et al. (2014). "Aberrant TAL1 activation is mediated by an interchromosomal interaction in human T-cell acute lymphoblastic leukemia." *Leukemia* **28**(2): 349-61.
141. Peters, J. M., A. Tedeschi, et al. (2008). "The cohesin complex and its roles in chromosome biology." *Genes Dev* **22**(22): 3089-114.
142. Phillips, J. E. and V. G. Corces (2009). "CTCF: master weaver of the genome." *Cell* **137**(7): 1194-211.
143. Philonenko, E. S., D. B. Klochkov, et al. (2009). "TMEM8 - a non-globin gene entrapped in the globin web." *Nucleic Acids Res* **37**(22): 7394-406.
144. Pikaart, M. J., F. Recillas-Targa, et al. (1998). "Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators." *Genes Dev* **12**(18): 2852-62.
145. Pope, B. D., T. Ryba, et al. "Topologically associating domains are stable units of replication-timing regulation." *Nature* **515**(7527): 402-5.
146. Pugacheva, E. M., V. K. Tiwari, et al. (2005). "Familial cases of point mutations in the XIST promoter reveal a correlation between CTCF binding and pre-emptive choices of X chromosome inactivation." *Hum Mol Genet* **14**(7): 953-65.
147. Quinlan, A. R. and I. M. Hall "BEDTools: a flexible suite of utilities for comparing genomic features." *Bioinformatics* **26**(6): 841-2.
148. Rayess, H., M. B. Wang, et al. (2012). "Cellular senescence and tumor suppressor gene p16." *Int J Cancer* **130**(8): 1715-25.
149. Razin, S. V., C. M. Farrell, et al. (2003). "Genomic domains and regulatory elements operating at the domain level." *Int Rev Cytol* **226**: 63-125.
150. Razin, S. V., A. Rynditch, et al. (2004). "The 33 kb transcript of the chicken alpha-globin gene domain is part of the nuclear matrix." *J Cell Biochem* **92**(3): 445-57.
151. Recillas-Targa, F., A. C. Bell, et al. (1999). "Positional enhancer-blocking activity of the chicken beta-globin insulator in transiently transfected cells." *Proc Natl Acad Sci U S A* **96**(25): 14354-9.
152. Recillas-Targa, F. and S. V. Razin (2001). "Chromatin domains and regulation of gene expression: familiar and enigmatic clusters of chicken globin genes." *Crit Rev Eukaryot Gene Expr* **11**(1-3): 227-42.
153. Renaud, S., D. Loukinov, et al. (2005). "CTCF binds the proximal exonic region of hTERT and inhibits its transcription." *Nucleic Acids Res* **33**(21): 6850-60.
154. Renda, M., I. Baglivo, et al. (2007). "Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding,

- and a single finger-DNA interaction controls binding at imprinted loci." *J Biol Chem* **282**(46): 33336-45.
155. Rhee, H. S. and B. F. Pugh (2011). "Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution." *Cell* **147**(6): 1408-19.
  156. Ribich, S., B. Tasic, et al. (2006). "Identification of long-range regulatory elements in the protocadherin-alpha gene cluster." *Proc Natl Acad Sci U S A* **103**(52): 19719-24.
  157. Robinett, C. C., A. O'Connor, et al. (1997). "The repeat organizer, a specialized insulator element within the intergenic spacer of the *Xenopus* rRNA genes." *Mol Cell Biol* **17**(5): 2866-75.
  158. Rubio, E. D., D. J. Reiss, et al. (2008). "CTCF physically links cohesin to chromatin." *Proc Natl Acad Sci U S A* **105**(24): 8309-14.
  159. Sambrook, J., E. F. Fritsch, et al. (1989). "Molecular cloning: A Laboratory manual. 2nd edition. ." Cold Spring Harbor Laboratory Press, New York.
  160. Schoenherr, C. J., J. M. Levorse, et al. (2003). "CTCF maintains differential methylation at the *Igf2/H19* locus." *Nat Genet* **33**(1): 66-9.
  161. Schubeler, D., C. Francastel, et al. (2000). "Nuclear localization and histone acetylation: a pathway for chromatin opening and transcriptional activation of the human beta-globin locus." *Genes Dev* **14**(8): 940-50.
  162. Shao, Z., Y. Zhang, et al. "MANorm: a robust model for quantitative comparison of ChIP-Seq data sets." *Genome Biol* **13**(3): R16.
  163. Shukla, S., E. Kavak, et al. (2011). "CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing." *Nature* **479**(7371): 74-9.
  164. Simonis, M., J. Kooren, et al. (2007). "An evaluation of 3C-based methods to capture DNA interactions." *Nat Methods* **4**(11): 895-901.
  165. Sjakste, N. (2004). "Site-specific excision or protection of an alpha A globin gene genomic site in apoptotic transformed chicken erythroblasts." *Cell Mol Biol Lett* **9**(3): 429-37.
  166. Splinter, E., H. Heath, et al. (2006). "CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus." *Genes Dev* **20**(17): 2349-54.
  167. Sun, B. K., A. M. Deaton, et al. (2006). "A transient heterochromatic state in *Xist* preempts X inactivation choice without RNA stabilization." *Mol Cell* **21**(5): 617-28.
  168. Szabo, P. E., S. H. Tang, et al. (2004). "Role of CTCF binding sites in the *Igf2/H19* imprinting control region." *Mol Cell Biol* **24**(11): 4791-800.
  169. Targa, F. R., C. V. de Moura Gallo, et al. (1993). "Silencer and enhancer elements located at the 3'-side of the chicken and duck alpha-globin-encoding gene domains." *Gene* **129**(2): 229-37.
  170. Tolhuis, B., R. J. Palstra, et al. (2002). "Looping and interaction between hypersensitive sites in the active beta-globin locus." *Mol Cell* **10**(6): 1453-65.
  171. Tsytsykova, A. V., R. Rajsbaum, et al. (2007). "Activation-dependent intrachromosomal interactions formed by the *TNF* gene promoter and two distal enhancers." *Proc Natl Acad Sci U S A* **104**(43): 16850-5.

172. Ulianov, S. V., A. A. Gavrilov, et al. (2012). "Spatial organization of the chicken beta-globin gene domain in erythroid cells of embryonic and adult lineages." *Epigenetics Chromatin* **5**(1): 16.
173. Valadez-Graham, V., S. V. Razin, et al. (2004). "CTCF-dependent enhancer blockers at the upstream region of the chicken alpha-globin gene domain." *Nucleic Acids Res* **32**(4): 1354-62.
174. Vetchinova, A. S., S. B. Akopov, et al. (2006). "Two-dimensional electrophoretic mobility shift assay: identification and mapping of transcription factor CTCF target sequences within an FXYD5-COX7A1 region of human chromosome 19." *Anal Biochem* **354**(1): 85-93.
175. Vietri Rudan, M., C. Barrington, et al. "Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture." *Cell Rep* **10**(8): 1297-309.
176. Vostrov, A. A. and W. W. Quitschke (1997). "The zinc finger protein CTCF binds to the APBbeta domain of the amyloid beta-protein precursor promoter. Evidence for a role in transcriptional activation." *J Biol Chem* **272**(52): 33353-9.
177. Vostrov, A. A., M. J. Taheny, et al. (2002). "A region to the N-terminal side of the CTCF zinc finger domain is essential for activating transcription from the amyloid precursor protein promoter." *J Biol Chem* **277**(2): 1619-27.
178. Wada, Y., Y. Ohta, et al. (2009). "A wave of nascent transcription on activated human genes." *Proc Natl Acad Sci U S A* **106**(43): 18357-61.
179. Wallace, J. A. and G. Felsenfeld (2007). "We gather together: insulators and genome organization." *Curr Opin Genet Dev* **17**(5): 400-7.
180. Wang, H., M. T. Maurano, et al. (2012). "Widespread plasticity in CTCF occupancy linked to DNA methylation." *Genome Res* **22**(9): 1680-8.
181. Wang, J., Y. Wang, et al. (2012). "De-SUMOylation of CCCTC binding factor (CTCF) in hypoxic stress-induced human corneal epithelial cells." *J Biol Chem* **287**(15): 12469-79.
182. Wang, L., L. J. Di, et al. (2009). "Inter-MAR association contributes to transcriptionally active looping events in human beta-globin gene cluster." *PLoS One* **4**(2): e4629.
183. Wei, Z., D. Huang, et al. (2013). "Biological implications and regulatory mechanisms of long-range chromosomal interactions." *J Biol Chem* **288**(31): 22369-77.
184. Wendt, K. S., K. Yoshida, et al. (2008). "Cohesin mediates transcriptional insulation by CCCTC-binding factor." *Nature* **451**(7180): 796-801.
185. West, A. G., M. Gaszner, et al. (2002). "Insulators: many functions, many mechanisms." *Genes Dev* **16**(3): 271-88.
186. Whittington, T., A. C. Perkins, et al. (2009). "High-throughput chromatin information enables accurate tissue-specific prediction of transcription factor binding sites." *Nucleic Acids Res* **37**(1): 14-25.
187. Wutz, A., T. P. Rasmussen, et al. (2002). "Chromosomal silencing and localization are mediated by different domains of Xist RNA." *Nat Genet* **30**(2): 167-74.

188. Xiao, T., J. Wallace, et al. (2011). "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." *Mol Cell Biol* **31**(11): 2174-83.
189. Xie, X., T. S. Mikkelsen, et al. (2007). "Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites." *Proc Natl Acad Sci U S A* **104**(17): 7145-50.
190. Xu, N., M. E. Donohoe, et al. (2007). "Evidence that homologous X-chromosome pairing requires transcription and Ctfc protein." *Nat Genet* **39**(11): 1390-6.
191. Xu, N., C. L. Tsai, et al. (2006). "Transient homologous chromosome pairing marks the onset of X inactivation." *Science* **311**(5764): 1149-52.
192. Yao, H., K. Brick, et al. (2010). "Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA." *Genes Dev* **24**(22): 2543-55.
193. Yu, W., V. Ginjala, et al. (2004). "Poly(ADP-ribosyl)ation regulates CTCF-dependent chromatin insulation." *Nat Genet* **36**(10): 1105-10.
194. Yusufzai, T. M. and G. Felsenfeld (2004). "The 5'-HS4 chicken beta-globin insulator is a CTCF-dependent nuclear matrix-associated element." *Proc Natl Acad Sci U S A* **101**(23): 8620-4.
195. Yusufzai, T. M., H. Tagami, et al. (2004). "CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species." *Mol Cell* **13**(2): 291-8.
196. Zhang, Y., T. Liu, et al. (2008). "Model-based analysis of ChIP-Seq (MACS)." *Genome Biol* **9**(9): R137.
197. Zhang, Z., X. Yang, et al. (2009). "Enhanced amplification of GC-rich DNA with two organic reagents." *Biotechniques* **47**(3): 775-9.
198. Zhao, H. and A. Dean (2004). "An insulator blocks spreading of histone acetylation and interferes with RNA polymerase II transfer between an enhancer and gene." *Nucleic Acids Res* **32**(16): 4903-19.
199. Zhao, J., B. K. Sun, et al. (2008). "Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome." *Science* **322**(5902): 750-6.
200. Ziebarth, J. D., A. Bhattacharya, et al. (2012). "CTCFBDB 2.0: a database for CTCF-binding sites and genome organization." *Nucleic Acids Res* **41**(Database issue): D188-94.
201. Zlatanova, J. and P. Caiafa (2009). "CCCTC-binding factor: to loop or to bridge." *Cell Mol Life Sci* **66**(10): 1647-60.
202. Разин, С. В., С. В. Ульянов, et al. (2012). "Домены альфа- и бета-глобиновых генов в контексте структурно-функциональной организации эукариотического генома." *Успехи биологической химии* **52**: 3–36.

## ПРИЛОЖЕНИЯ

**Таблица 2. Фрагменты  $\alpha$ -глобинового локуса кур, связывающие CTCF *in vitro* (по данным двумерного EMSA). Координаты сборки генома *Gallus gallus* 2011 года (galGal4).**

Хромосома	Начальная координата	Конечная координата	Количество секвенированных фрагментов	Сайты рестрикции, фланкирующие фрагмент
14	11961575	11962315	3	Sau3aI
14	11962394	11963203	3	Sau3aI
14	11962630	11963444	1	Csp6I
14	11964000	11964471	1	Sau3aI
14	11965268	11966148	1	Sau3aI
14	11968549	11969437	1	Sau3aI
14	11969403	11970506	1	Csp6I
14	11969442	11970752	6	Sau3aI
14	11975757	11976295	1	Sau3aI
14	11976339	11977061	1	Sau3aI
14	11976357	11977370	2	Csp6I
14	11977066	11977829	1	Sau3aI
14	11978308	11979547	2	Csp6I
14	11978832	11979650	1	Sau3aI
14	11980405	11981227	4	Csp6I
14	11981486	11982469	9	Sau3aI
14	11981864	11983028	2	Csp6I
14	11984902	11986013	2	Sau3aI
14	11988716	11989696	3	Sau3aI
14	11989891	11990724	4	Csp6I

14	11991672	11992290	1	Sau3aI
14	11996126	11997199	2	Sau3aI
14	11997802	11999092	1	Csp6I
14	12011396	12012237	1	Csp6I
14	12013571	12014147	1	Sau3aI
14	12013859	12015342	1	Csp6I
14	12014430	12014975	5	Sau3aI
14	12015551	12016731	1	Csp6I
14	12018555	12019837	2	Sau3aI
14	12030464	12031061	1	Sau3aI
14	12031568	12033117	1	Sau3aI
14	12033802	12034833	1	Csp6I
14	12038414	12039289	2	Sau3aI
14	12041232	12042410	2	Sau3aI
14	12044312	12045118	2	Csp6I
14	12049876	12050928	1	Sau3aI
14	12050933	12051868	4	Sau3aI
14	12051267	12052289	2	Csp6I
14	12051873	12052661	4	Sau3aI
14	12054556	12055418	5	Sau3aI
14	12056791	12057736	1	Sau3aI
14	12057741	12058567	6	Sau3aI
14	12057926	12058524	1	Csp6I
14	12058783	12059601	1	Sau3aI
14	12059692	12060393	2	Sau3aI
14	12061554	12062554	5	Csp6I
14	12062211	12063388	1	Sau3aI
14	12064852	12065604	1	Csp6I
14	12070253	12071132	1	Sau3aI

14	12076747	12077629	2	Csp6I
14	12077027	12078194	1	Sau3aI
14	12078850	12079769	1	Sau3aI
14	12083246	12084150	1	Sau3aI
14	12085743	12086944	1	Csp6I
14	12087344	12088079	1	Sau3aI
14	12092082	12092664	1	Csp6I
14	12093503	12094354	3	Sau3aI
14	12095958	12096968	1	Csp6I
14	12099458	12100798	3	Csp6I
14	12099807	12100872	3	Sau3aI
14	12101752	12102528	1	Sau3aI
14	12102533	12102662	1	Sau3aI
14	12102989	12103636	2	Sau3aI
14	12106517	12107425	2	Sau3aI
14	12107430	12108690	2	Sau3aI
14	12110994	12111462	1	Sau3aI
14	12112259	12113297	2	Sau3aI
14	12115662	12117102	1	Csp6I
14	12118745	12119311	2	Sau3aI
14	12119459	12120474	2	Sau3aI
14	12122895	12123973	3	Csp6I
14	12125597	12126365	2	Sau3aI
14	12127375	12128705	3	Csp6I
14	12130333	12131148	1	Sau3aI
14	12133110	12133466	1	Sau3aI
14	12136898	12137661	3	Sau3aI
14	12139189	12140110	3	Csp6I
14	12139527	12140575	2	Sau3aI

14	12140115	12140972	2	Csp6I
14	12140664	12141432	2	Sau3aI
14	12140977	12141957	2	Csp6I
14	12144729	12145846	16	Csp6I
14	12147411	12148215	2	Sau3aI
14	12153908	12154764	1	Csp6I
14	12156827	12157632	4	Sau3aI
14	12160450	12161555	1	Sau3aI
14	12162174	12163455	2	Csp6I
14	12166334	12167256	1	Sau3aI
14	12170130	12170646	1	Sau3aI
14	12177512	12178292	1	Csp6I
14	12181739	12182818	1	Sau3aI
14	12183542	12184614	12	Sau3aI

**Таблица 3. Координаты участков  $\alpha$ -глобинового локуса кур, связывающих CTCF в клетках линии HD3 (по данным ChIP-seq). Координаты сборки генома *Gallus gallus* 2011 года (galGal4).**

Хромосома	Начальная координата	Конечная координата
14	12006777	12007459
14	12066678	12067808
14	12099755	12100556
14	12117090	12118000
14	12145795	12146484

**Таблица 4. Координаты участков  $\alpha$ -глобинового локуса кур, связывающих CTCF в клетках линии DT40 (по данным ChIP-seq). Координаты сборки генома *Gallus gallus* 2011 года (galGal4).**

Хромосома	Начальная координата	Конечная координата
14	12006720	12007580
14	12028465	12029400
14	12066667	12067797
14	12117134	12118013

**Таблица 5. Координаты участков  $\alpha$ -глобинового локуса кур, связывающих CTCF в эритроблестах пятидневных эмбрионов кур (по результатам обработки данных Martin D и соавторов [122]). Координаты сборки генома *Gallus gallus* 2011 года (galGal4).**

Хромосома	Начальная координата	Конечная координата
14	12094203	12094800
14	12098936	12099910
14	12104876	12105209
14	12117431	12117685

**Таблица 6. Координаты участков  $\alpha$ -глобинового локуса кур, связывающих CTCF в эритроблестах десятидневных эмбрионов кур (по результатам обработки данных Martin D и соавторов [122]). Координаты сборки генома *Gallus gallus* 2011 года (galGal4).**

Хромосома	Начальная координата	Конечная координата
14	12028880	12029186
14	12094292	12094707
14	12098982	12099646
14	12117360	12117830