

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

На правах рукописи

Котова Елена Сергеевна

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

специальность – 03.01.03 – молекулярная биология

АВТОРЕФЕРАТ

диссертации на соискание учёной степени
кандидата биологических наук

Москва 2016

Работа выполнена в лаборатории структуры и функций генов человека в Федеральном государственном бюджетном учреждении науки Институте биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова Российской академии наук (ИБХ РАН).

Научный руководитель:

Акопов Сергей Борисович, доктор биологических наук.

Официальные оппоненты:

Пасюкова Елена Генриховна, доктор биологических наук, профессор, лаборатория геномной изменчивости Федерального государственного бюджетного учреждения науки Института молекулярной генетики Российской академии наук, заведующий.

Величко Артем Константинович, кандидат биологических наук, лаборатория структурно-функциональной организации хромосом Федерального государственного бюджетного учреждения науки Института биологии гена Российской академии наук, старший научный сотрудник.

Ведущая организация:

Федеральное государственное бюджетное учреждение науки Институт молекулярной биологии им. В.А. Энгельгардта Российской академии наук (ИМБ РАН).

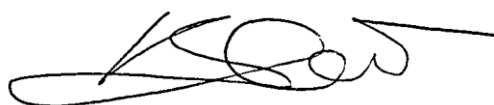
Защита состоится «14» декабря 2016 года в 10 часов на заседании диссертационного совета Д 002.019.01 при ИБХ РАН по адресу: 117997, Москва В-437, ул. Миклухо-Маклая, д. 16/10.

С диссертацией можно ознакомиться в библиотеке и на сайте www.ibch.ru ИБХ РАН.

Автореферат разослан « » 2016 года.

Ученый секретарь диссертационного совета, доктор физико-математических наук

В.А. Олейников



ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность исследования

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 расположении и функционировании этих регуляторных элементов. Таким образом, необходимы картирование и анализ активности областей геномов, ответственных за регуляцию экспрессию 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 петлевою 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. Изучение сайтов связывания CTCF, расположенных в одном локусе, позволяет в дальнейшем выявлять механизмы регуляции экспрессии генов локуса, в которых

одновременно задействовано несколько CTCF-связывающих участков локуса. В этом состоит 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 α -глобиновых генов кур. Множество работ посвящено выявлению и анализу активности различных функциональных элементов этого домена.

В α -глобиновом локусе кур гены α -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 α -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 границ функционального хроматинового домена α -глобиновых генов, данный локус, причисляют к доменам открытого типа. Однако, в эритроидных клетках происходят некоторые изменения структуры хроматина в области α -глобиновых генов. Например, повышение уровня ацетилирования гистонов по ряду специфических позиций. По-видимому, этот процесс контролируется путем низкоуровневой транскрипции всего домена. В области MRE в направлении генов α -глобинов начинается синтез длинного транскрипта, размером около 33 т.п.о. Возможно, домен α -глобиновых генов все же имеет определенные границы, 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 α -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 α -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.

В ходе данной работы был проведен поиск фрагментов α -глобинового локуса кур, способных связывать CTCF *in vitro*, с использованием двумерного EMSA (Electrophoretic Mobility Shift Assay). Методом иммунопреципитации хроматина с последующим анализом иммунопреципитированной ДНК было изучено связывание CTCF с этими фрагментами *in vivo* в геноме эритроидных и лимфоидных клеток кур. Также 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).

Степень разработанности темы исследования

Для поиска сайтов связывания CTCF использовались различные подходы. В ранних исследованиях сайты связывания CTCF были найдены случайным образом в составе некоторых регуляторных последовательностей геномов позвоночных (Banahmad et al., 1990; Kohne et al., 1993, Lobanekov 1989). В α -глобиновом локусе кур для некоторых сайтов гиперчувствительности к ДНКазе (Valadez-Graham V 2004) и GC-богатых участков ДНК (Klochkov 2006) 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 (Valadez-Graham V 2004, Klochkov 2006).

Также в результате полногеномного выявления фрагментов ДНК, связывающих CTCF, был обнаружен ряд участков α -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 α -глобиновом локусе кур, активных во взрослых эритроидных и неэритроидных клетках.

Цели и задачи

Целью диссертационной работы стало выявление в α -глобиновом локусе кур CTCF-связывающих участков, сравнение связывания CTCF в α -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 α -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 α -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.

Научная новизна и практическая значимость

В настоящей работе методом двумерного EMSA (Electrophoretic Mobility Shift Assay) впервые был выявлен ряд фрагментов α -глобинового локуса кур, способных связывать CTCF *in vitro*. Методом иммунопреципитации хроматина было изучено связывание CTCF с этими

фрагментами *in vivo* в геноме эритроидных и лимфоидных клеток кур, клеточные линии HD3 и DT40, соответственно. Впервые был проведен полногеномный поиск участков ДНК, связывающих CTCF *in vivo*, в геноме неэмбриональных эритроидных и лимфоидных клеток кур методом ChIP-seq (иммунопреципитация хроматина с последующим масштабным секвенированием). Вблизи α -глобиновых генов было выявлено два сайта связывания CTCF, общих для эритроидных и лимфоидных неэмбриональных куриных клеток. Был выявлен сайт связывания CTCF, имеющий высокую сродство к белку только в клетках взрослых кур эритроидного происхождения.

Картирование CTCF-связывающих последовательностей в геномах позвоночных необходимо для изучения роли, которую играет CTCF в ДНК-зависимых процессах. Проведенные в данной работе картирование и анализ активности CTCF-связывающих последовательностей, α -глобинового локуса кур важны также для дальнейшего изучения механизмов, регулирующих экспрессию генов этого локуса. Полученные данные могут быть использованы в дальнейшем для изучения механизмов регуляции экспрессии генов на уровне хроматиновых доменов не только в α -глобиновом локусе кур, но и в целом, в геномах позвоночных. Таким образом, полученная в ходе данной работы информация может быть использована в фундаментальных молекулярно-биологических исследованиях.

Известно, что изменения в экспрессии и связывании CTCF с ДНК наблюдаются при различных заболеваниях. На основе полученных данных могут быть выявлены закономерности, важные для развития медицины.

Апробация работы

Результаты работы представлены на российских и международных конференциях: научная конференция по биоорганической химии и биотехнологии “X чтения памяти академика Юрия Анатольевича Овчинникова” (Москва, 2011), 4-th international IMBG conference for young scientists "MOLECULAR BIOLOGY: ADVANCES AND PERSPECTIVES" (Kiev, Ukraine, 2011), VIII международная научная конференция студентов, аспирантов и молодых ученых “Ломоносов”.

Структура и объем работы

Диссертационная работа изложена на 119 страницах и состоит из введения, обзора литературы, экспериментальной части, результатов и их обсуждения, основных выводов, списка литературы из 202 наименований и приложений. Диссертация содержит 6 таблиц и 19 рисунков.

Публикации

По материалам диссертации опубликовано 4 статьи в рецензируемых журналах.

СОДЕРЖАНИЕ РАБОТЫ

Получение и функциональный анализ фракции, обогащенной куриным белком CTCF

Для выявления CTCF-связывающих последовательностей генома кур мы разработали простой и быстрый способ получения белковой фракции, обогащенной куриным белком CTCF, синтезированным в клетках линии COS-1. Система экспрессии гена куриного CTCF с использованием клеток позвоночных была выбрана нами на основании данных о

консервативности структуры белка CTCF и возможном существовании посттрансляционных модификаций, важных для его функциональной активности.

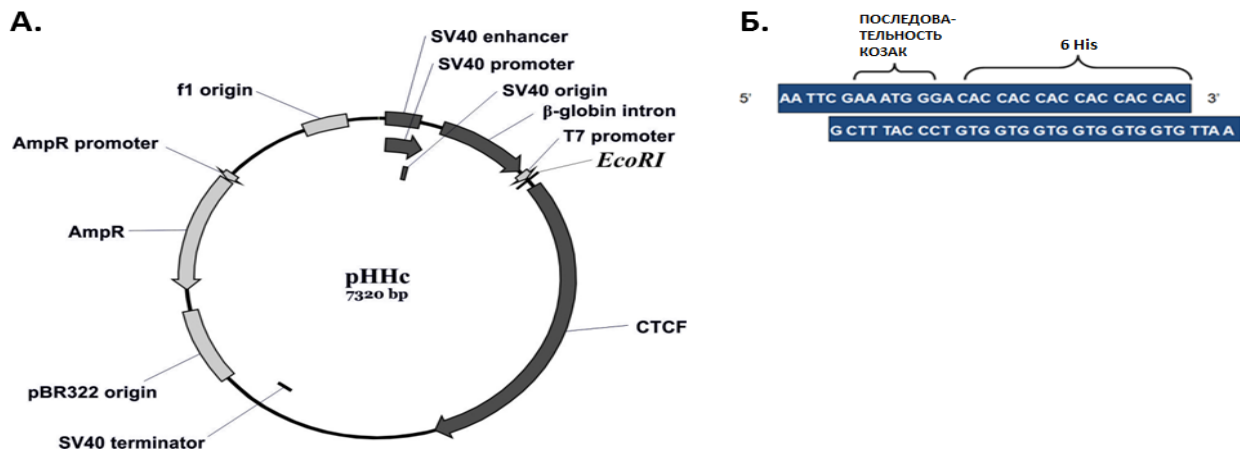


Рисунок 1. Получение плазмиды рNHс-his, предназначенной для экспрессии полноразмерного гена куриного белка CTCF с N-концевым гистидиновым тагом. А – Схема плазмиды рNHс, содержащей ген полноразмерного куриного белка CTCF. Плазида получена на основе вектора рSG5. SV40 enhancer – энхансер вируса SV40, SV40 promoter – промотор вируса SV40, SV40 origin – точка начала репликации вируса SV40, β -globin intron – интрон β -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 β -лактамазы, экспрессия которого обеспечивает устойчивость к ампициллину, AmpR promoter – промотор гена β -лактамазы, f1 origin – точка начала репликации бактериофага f1. Б – Олгонуклеотидный дуплекс, клонированный по сайту *EcoRI* в вектор рNHс. Дуплекс содержит последовательность Козак, а также участок, кодирующий 6 последовательных гистидиновых остатков (гистидиновый таг).

Для последующей очистки белка CTCF с помощью никель-аффинной хроматографии в начало открытой рамки считывания гена *CTCF* плазмиды рNHс, обеспечивающей повышенную экспрессию этого гена в клетках COS-1, была добавлена последовательность, кодирующая гистидиновый таг (аминокислотная последовательность из шести остатков гистидина (H6) (рисунок 1). Полученная конструкция была названа рNHс-his. С помощью вестерн-блот анализа было показано повышенное образование белка CTCF в клетках COS-1, трансфицированных рNHс-his. Методом ретардации в геле и дополнительной ретардации в геле был проведен функциональный анализ белка CTCF с N-концевым гистидиновым тагом, присутствующего в ядерном экстракте клеток COS-1, трансфицированных рNHс-his. Показано специфичное взаимодействие модифицированного белка CTCF с известными CTCF-связывающими последовательностями. Показана ядерная локализация белка CTCF, образующегося в результате трансфекции клеток COS-1 плазмидой рNHс-his. Была проведена оптимизация условий очистки белка CTCF с N-концевым гистидиновым тагом из лизата клеток COS-1 (рисунок 2). Проводилась двухстадийная очистка белка CTCF путем никель-аффинной хроматографии и последующей ионообменной хроматографии на Q-сефарозе. В результате была получена фракция, содержащая приблизительно 10% CTCF. Методом дополнительной ретардации в геле было показано специфичное узнавание белком CTCF с N-концевым гистидиновым тагом в составе этой фракции известных CTCF-связывающих последовательностей (рисунок 3).

Добавление антител к гистидиновому тагу приводило к дополнительному сдвигу полос, соответствующих ДНК-белковым комплексам. Следовательно, в состав ДНК-белковых комплексов входил белок CTCF с гистидиновым тагом.

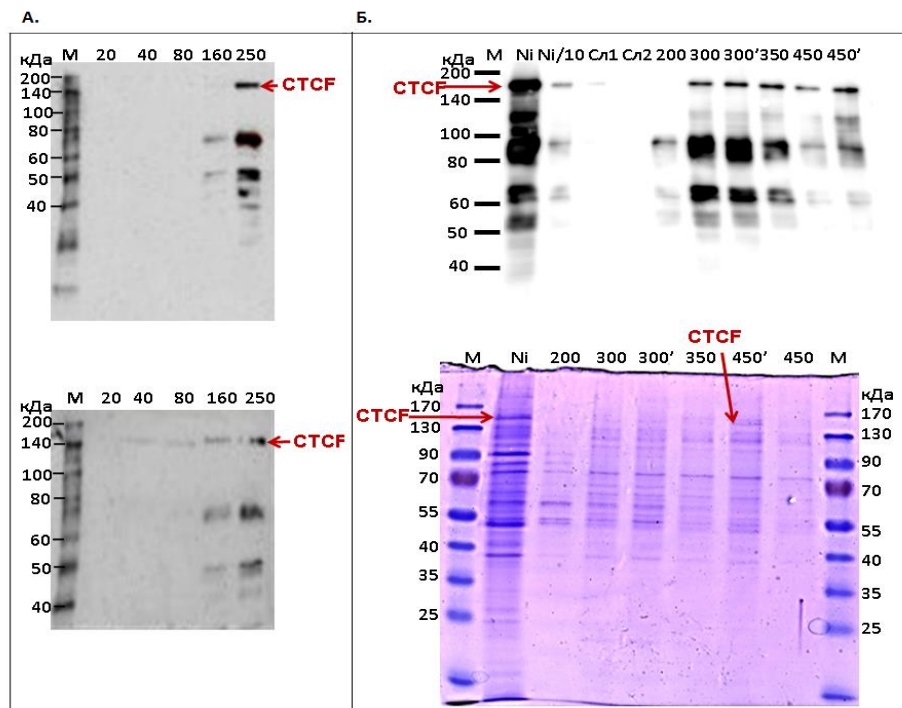


Рисунок 2. Оптимизация условий очистки белка CTCF с N-концевым гистидиновым тагом из лизата клеток COS-1. CTCF – полоса, соответствующая полноразмерному белку CTCF. М – маркер молекулярных масс. А. – Вестерн-блот анализ фракций, полученных при очистке CTCF с N-концевым гистидиновым тагом с помощью колонок 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 CTCF человека. Б. – Анализ фракций, полученных при дополнительной очистке CTCF 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.

Выявление фрагментов α -глобинового локуса кур, способных связывать белок CTCF, методом двумерного EMSA.

Получения библиотеки фрагментов ДНК, связывающих CTCF *in vitro*, проводилось методом двумерного EMSA (Electrophoretic Mobility Shift Assay) (Vetchinova 2006). Для приготовления библиотеки перекрывающихся фрагментов ДНК α -глобинового локуса искусственная бактериальная хромосома, содержащая вставку локуса α -глобиновых генов кур длиной 227366 п.н., была гидролизована эндонуклеазой рестрикции Sau3AI либо Csp6I. К концам полученных рестриктных фрагментов присоединяли синтетические адапторы. Полученные суббиблиотеки амплифицировали с помощью ПЦР, а затем смешивали в равных количествах. Итоговую библиотеку коротких фрагментов (около 1000 фрагментов со средней длиной ~500 п.н.) метили ^{32}P и смешивали с белковой фракцией, обогащенной полноразмерным CTCF, экспрессированным в клетках COS-1. Реакционную смесь далее разделяли электрофоретически в неденатурирующем полиакриламидном геле (электрофоретическое разделение в первом направлении). Полосу с образцом вырезали, инкубировали в содержащем SDS буфере для

разрушения ДНК-белковых комплексов, и фрагменты ДНК разделяли в содержащем SDS геле (электрофоретическое разделение во втором направлении). Область геля, содержащую большинство фрагментов, взаимодействующих с CTCF *in vitro* (обведена на рисунке 4) вырезали. Фрагменты ДНК элюировали и амплифицировали путем ПЦР. Полученная библиотека использовалась для второго раунда двумерного EMSA для повышения эффективности отбора CTCF-связывающих фрагментов.

Специфичность отбора проверяли при помощи ПЦР-амплификации полученной и исходной библиотек с праймерами к известным последовательностям α -глобинового локуса кур, связывающим CTCF: CDS (CTCF-зависимый сайленсер) (Klochkov 2006), M9 (Valadez-Graham 2004). В качестве отрицательного контроля использовался участок экзона гена HBAD, не взаимодействующий с CTCF.

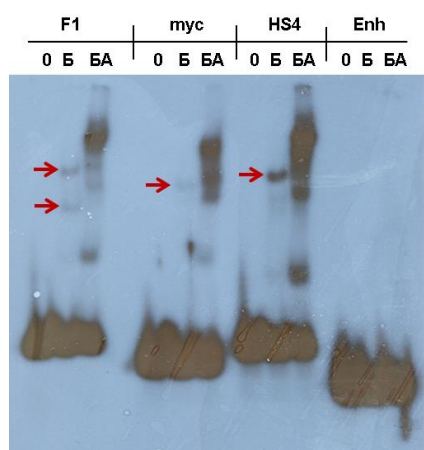


Рисунок 3. Функциональный анализ фракции, обогащенной куриным CTCF с N-концевым гистидиновым тагом методом дополнительного сдвига электрофоретической подвижности. Фрагменты ДНК, связывающие CTCF (Kohne, 1993, Chung, 1997; Bell, 1999, Fillipova 1996) обозначены: F1 – F1-блок лизоцимного сайленсера кур, тус – фрагмент промотора гена *c-тус* человека, HS4 – FII-фрагмент HS4-инсулятора из β -глобинового локуса кур. Enh – энхансер из β -глобинового локуса кур, не связывает CTCF (Klochkov 2006). 0 – фрагмент без белка, Б – фрагмент, инкубированный с белком, БА – фрагмент, инкубированный с белком и антителами к гистидиновому тагу. Стрелками

отмечены полосы, соответствующие ДНК-белковым комплексам. ДНК-белковые комплексы выявляются в результате добавления белковой фракции, обогащенной CTCF с N-концевым гистидиновым тагом, ко всем фрагментам ДНК, за исключением не связывающего CTCF энхансера из β -глобинового локуса кур.

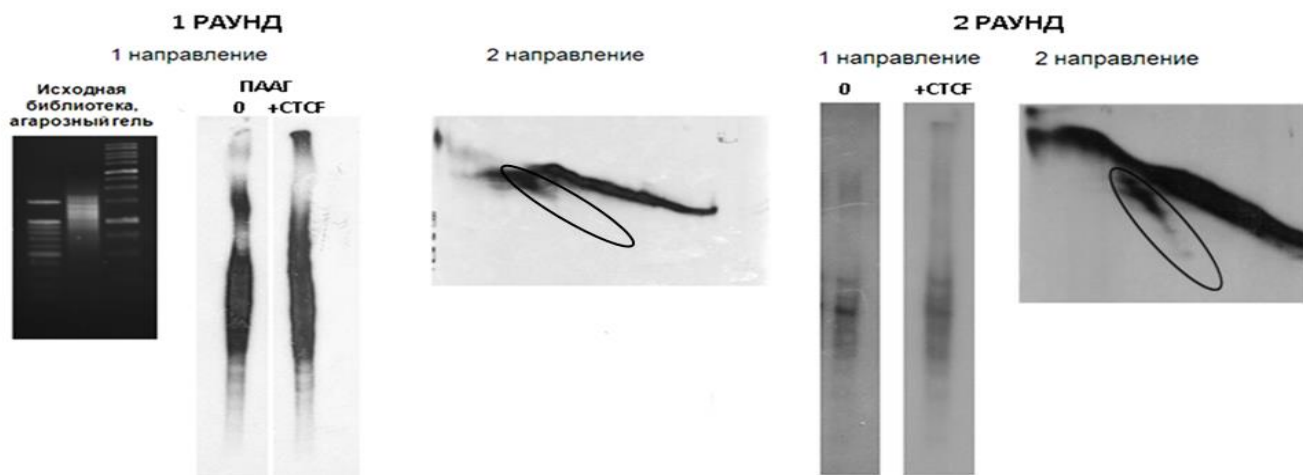


Рисунок 4. Получение библиотеки CTCF-связывающих фрагментов α -глобинового локуса кур методом двумерного EMSA. Слева направо. Первый раунд двумерного EMSA: электрофоретическое разделение в агарозном геле библиотеки фрагментов α -глобинового локуса кур; электрофоретическое разделение в ПААГ этой же библиотеки без белка (0) и после инкубации с CTCF-содержащей фракцией (+CTCF); электрофоретическое разделение инкубированной с белком библиотеки во втором направлении. 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.

Результаты амплификации представлены на рисунке 5. Согласно данным результатам можно примерно оценить, что обогащение конечной библиотеки CTCF-связывающими фрагментами составляет ~64–512 раз.

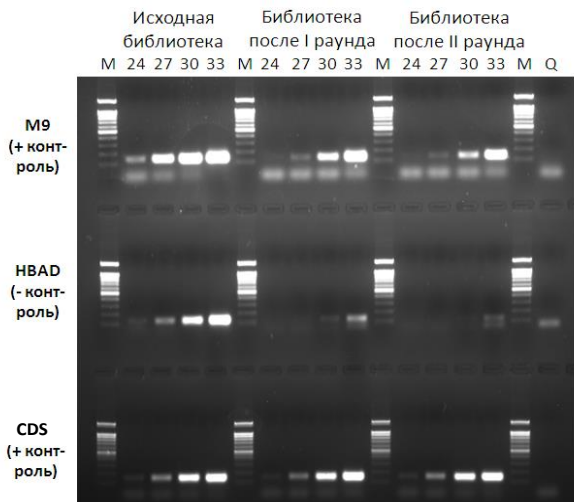


Рисунок 5. Анализ обогащения библиотеки, полученной в результате двумерного EMSA, последовательностями, связывающими CTCF *in vitro*. ПЦР-амплификация с отбором через три цикла. Электрофореграмма отобранных аликуот реакционной смеси. М9 – амплификация CTCF-связывающего фрагмента М9, HBAD – амплификация 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 аликуота. М-маркер масс, Q – контроль на контаминацию – в реакционной смеси отсутствует матрица.

Фрагменты ДНК, полученные после второго раунда отбора, клонировали в вектор pGEM-T. 230 клонов ранжировали в 96-луночных планшетах и секвенировали выделенные из них плазмиды в области вставки. Из выявленных в результате секвенирования последовательностей 22 соответствовали фрагментам ВАС, геномной ДНК *Escherichia coli* или химерным фрагментам, 208 соответствовали нехимерным фрагментам α -глобинового локуса. Среди них выявлено 79 уникальных последовательностей (рисунок 7). Построенная кривая разрежения (рисунок 6А) указывает, что секвенирование произведено с глубиной, достаточной для идентификации большинства CTCF-связывающих фрагментов локуса.

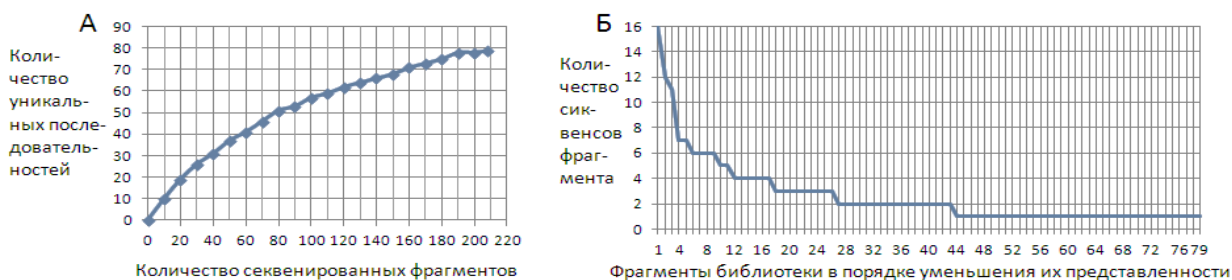


Рисунок 6. Графики, характеризующие глубину секвенирования библиотеки, полученной после двумерного EMSA, и представленность выявленных фрагментов ДНК. А. Кривая разрежения. Б. Кривая ранг-представленности.

Среди выявленных последовательностей содержались 10 из 13 последовательностей α -глобинового локуса, для которых ранее было показано связывание с CTCF. Это последовательности С10 – С14 (Valadez-Graham, 2004), последовательность CDS, 4 из 6 сайтов

связывания CTCF в хроматине эмбриональных эритробластов. Обогащение библиотеки еще одним фрагментом, M9 было показано путем ПЦР-амплификации (рисунок 5). По-видимому, он не попал в число секвенированных.

Мы проверили способность фрагментов, отобранных в результате двумерного EMSA, связывать CTCF методом ретардации в геле и дополнительной ретардации в геле (рисунок 8). В качестве зондов использовались 10 фрагментов, выявленных поблизости от α -глобиновых генов. Они обозначены на рисунке 7 номерами 1–4, 6–10, 13. Была продемонстрирована способность всех 10 фрагментов ДНК образовывать комплексы с белком CTCF, что говорит о достаточно высокой (около 100%) специфичности отбора, проведенного методом двумерного EMSA.

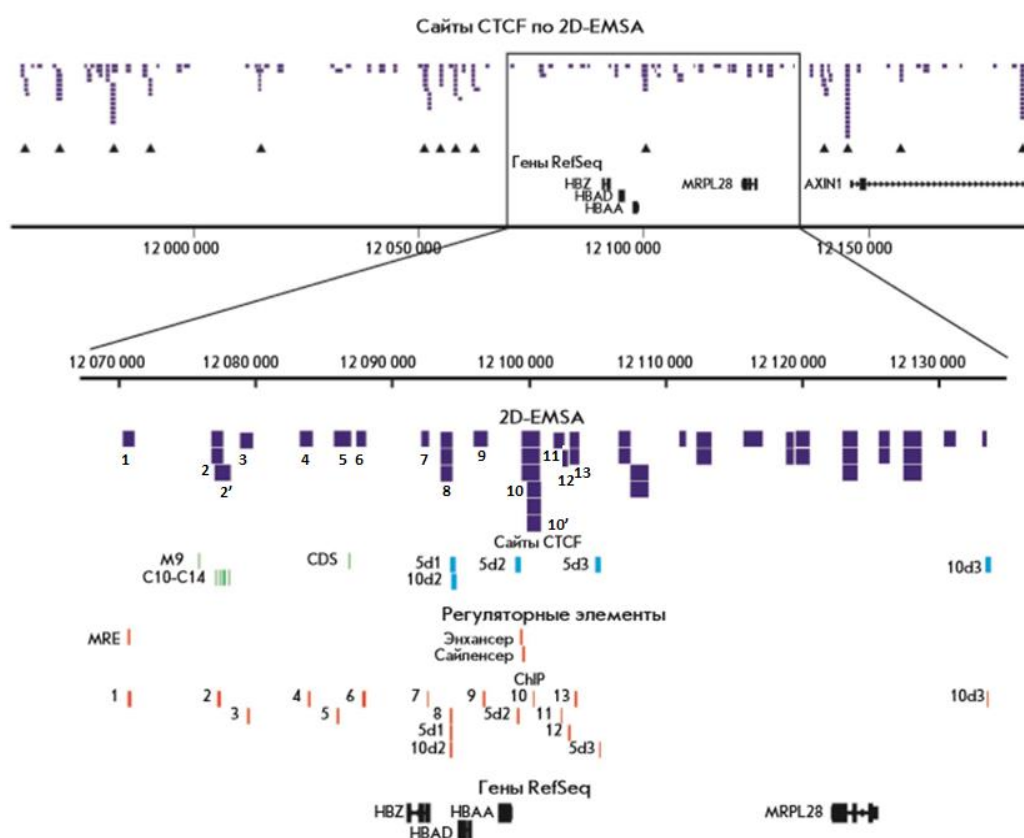


Рисунок 7. Расположение CTCF-связывающих участков и других регуляторных элементов в α -глобиновом локусе генома кур. На верхней карте показано расположение всех выявленных в результате двумерного EMSA фрагментов ДНК. Стрелки указывают участки ДНК, наиболее представленные в библиотеке. Ниже

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

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 куриных клеток эритроидного происхождения – линия HD3 и лимфоидного происхождения – DT40. Для иммунопреципитации хроматина использовались интактные клетки линии HD3 и индуцированные к дифференцировке (HD3-ind). Для каждого типа клеток иммунопреципитация хроматина была проведена в трех либо двух повторностях (в случае клеток HD3-ind). Анализ полученных иммунопреципитатов хроматина осуществлялся путем ПЦР в реальном времени.

Из тринадцати фрагментов ДНК, выявленных в результате двумерного EMSA, только один (фрагмент 10) связывает CTCF *in vivo* в используемых типах клеток. Связывание данного фрагмента осуществляется тканеспецифично. Его занятость белком CTCF в клетках HD3 до и после индукции выше, чем в клетках DT40 (рисунок 9).

Также путем иммунопреципитации хроматина проводился анализ занятости CTCF выявленных Martin и соавторами участков α -глобинового локуса, связывающих CTCF в эмбриональных эритроблестах кур. В клетках HD3 до и после индукции дифференцировки и в клетках DT40 связывание CTCF с этими участками локуса выявлено не было.

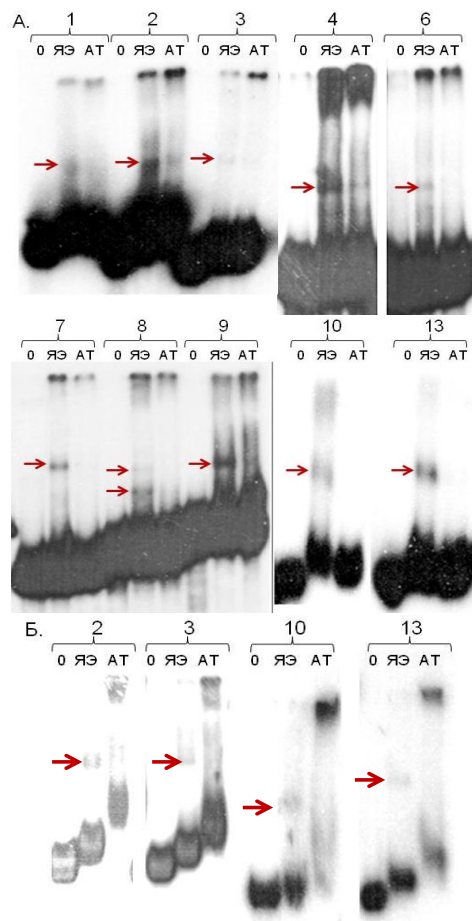


Рисунок 8. Анализ CTCF-связывающей активности десяти выявленных фрагментов ДНК, методом дополнительного сдвига электрофоретической подвижности. Числа соответствуют номерам фрагментов на рисунке 7. 0 – фрагмент ДНК не инкубировался с белковыми фракциями, ЯЭ – фрагмент ДНК инкубировался с ядерным экстрактом клеток COS-1, содержащих 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.

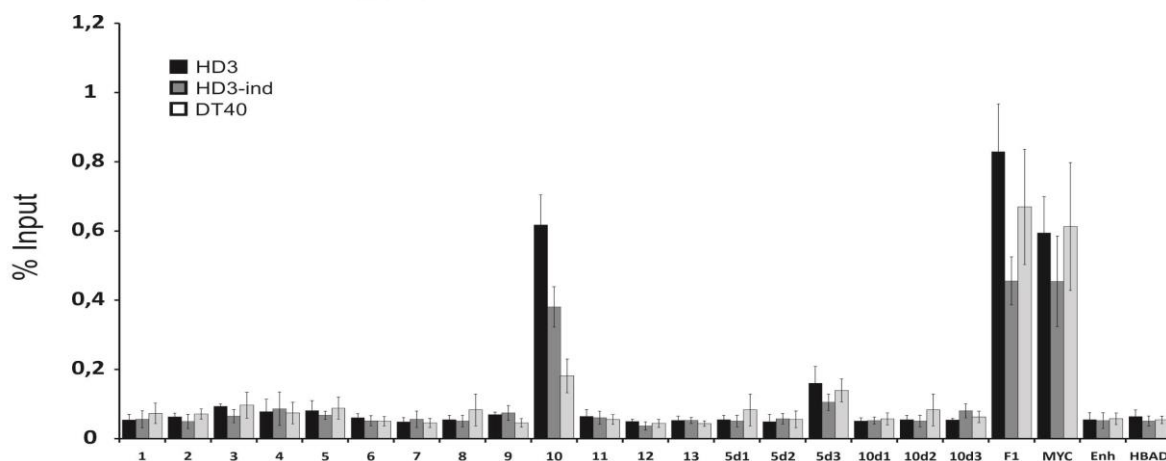


Рисунок 9. Результаты иммунопреципитаций хроматина с антителами к белку CTCF клеток HD3, интактных и после индукции дифференцировки, и DT40. Анализ иммунопреципитатов проводился путем ПЦР в реальном времени. По вертикальной оси отложен процент содержания данного фрагмента в иммунопреципитате от содержания данного фрагмента в образце Input (тотальная ДНК, выделенная из клеточного лизата после ультразвуковой обработки) - % Input.

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 (Martin 2011). 10d1 – 10d3 - сайты связывания CTCF, в хроматине эритробластов пятидневных эмбрионов кур (Martin 2011). F1 – F1-блок лизоцимного сайленсера, связывающий CTCF (Kohne 1993), MYC – CTCF-связывающий фрагмент промотора гена *c-myc* кур (Filipova 1996), Enh – энхансер из β -глобинового локуса кур, не связывает CTCF (Klochkov 2006), HBAD – последовательность экзона гена HBAD, не связывает CTCF. Высота столбиков – среднее арифметическое, планки погрешностей – стандартная ошибка среднего по трем либо двум (в случае HD3-ind) экспериментам по иммунопреципитации хроматина.

Полногеномное выявление связывающих белок CTCF нуклеотидных последовательностей методом ChIP-seq

Методом ChIP-seq был проведен полногеномный поиск областей, связывающих CTCF *in vivo*, в интактных клетках линии HD3, клетках линии HD3, индуцированных к дифференцировке (HD3-ind) и клетках линии DT40.

Массированному секвенированию были подвергнуты четыре образца ДНК: иммунопреципитаты хроматина клеток HD3, HD3-ind и DT40, полученные с антителами к CTCF, 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 исследуемых клетках: F1 (Baniahmad, 1990, Kohne, 1993), и MYC (Filipova 1996). Для сравнения использовались не связывающие CTCF фрагменты ДНК: Enh (Klochkov 2006) и HBAD (рисунок 10Б). 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.

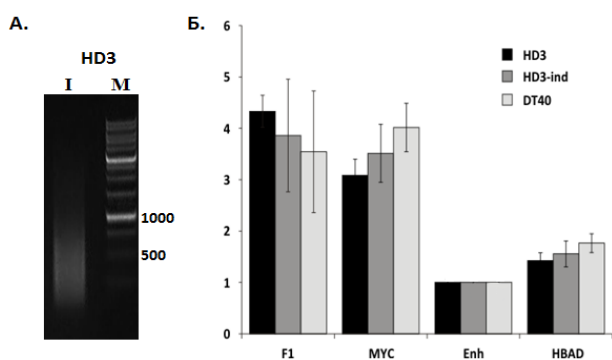


Рисунок 10. Оценка качества

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 β -глобинового локуса кур, не связывающего белок CTCF (фрагмент Enh). F1 – F1-блок лизоцимного сайленсера кур (Baniahmad, 1990, Kohne, 1993), MYC - CTCF-связывающий фрагмент промотора гена *c-myc* кур (Filipova 1996), HBAD – не связывающий CTCF фрагмент экзона гена *HBAD*. 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 4.0 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. 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.

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 антителами к белку CTCF (рисунок 10), обогащение иммунопреципитата CTCF-связывающими фрагментами MYC, F1 и 10, в случае интактных клеток HD3 было также примерно в 1,5 раза больше, чем в случае индуцированных. По-видимому, в клетках HD3 в результате индукции была снижена средняя по популяции клеток занятость белком сайтов связывания CTCF. Это может быть связано с процессами компактизации хроматина, сопровождающими заключительные стадии эритроидной дифференцировки у птиц (Kowalsky 2011). Снижение занятости белком CTCF его сайтов связывания также может быть следствием апоптотических изменений хроматина при дифференцировке клеток HD3 (Sjakste 2003, Iarovaia 2001).

Расположение сайтов связывания CTCF в α -глобиновом локусе кур

Расположение пиков, выявленных программой MACS 1.4.2 в α -глобиновом локусе кур (геномные координаты chr14:11957588-12184953 в сборке куриного генома 2011) приведено на рисунке 11.

В клетках линии DT40 было выявлено 4 пика, HD3 — 5 пиков. В индуцированных клетках HD3 программа не выявила в локусе α -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. “Снижение занятости белком CTCF-связывающих областей при индукции дифференцировки клеток HD3”), 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 11), общий для клеток HD3 и DT40, расположен ближе других к α -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 α -глобиновых генов в клетках взрослых кур (возможно, только в клетках крови и гематопозитических клетках). Сайт связывания CTCF в составе этого пика может препятствовать перекрестному влиянию цис-регуляторных элементов α -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 11), расположен с 3' стороны от α -глобиновых генов. Он перекрывается с CTCF-связывающими фрагментами 10 и 10' (рисунок 7), 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. “Выявление среди отобранных в результате двумерного EMSA фрагментов ДНК, последовательностей, способных связывать CTCF *in vivo*”).

С 3'-стороны от α -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 α -глобинов пик, расположенный при этом ниже промотора гена TMEM8, находится в интроне этого гена (пик 8, рисунок 11). Он выявляется в клетках HD3 и DT40, а также эмбриональных эритроблестах (см. “Сравнение результатов ChIP-seq анализа для взрослых и эмбриональных куриных клеток”). Ему соответствует область повышенной плотности прочтений в случае клеток HD3-ind. Он является ближайшим к α -глобиновым генам пиком, лежащим ниже промотора гена TMEM8 и, таким образом, может соответствовать второй границе функционального хроматинового домена α -глобиновых генов.

Оценка тканеспецифичности связывания CTCF

С помощью программы MAAnnot нами были оценены различия в занятости белком CTCF выявленных в разных типах клеток CTCF-связывающих областей (пиков). При сравнении клеток HD3 и DT40 в α -глобиновом локусе, разница в занятости CTCF была показана ($p < 10^{-3}$) только для пика, обозначенного на рисунке 11 номером 6. Он перекрывался с фрагментами 10 и 10',

выявленными в результате двумерного EMSA (рисунок 7). 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 MAopm, 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 "Выявление среди отобранных в результате двумерного EMSA фрагментов ДНК, последовательностей, способных связывать CTCF *in vivo*"). 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.

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 соавторами были опубликовали результаты ChIP-seq с антителами к CTCF для эритробластов пятидневных и десятидневных эмбрионов кур (Martin 2011). 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.

К публикации Martin и соавторов также были приложены файлы с координатами пиков, выявленных программой SISRrs в геноме кур сборки 2006 г. Геномные координаты пиков, выявленных в α -глобиновом локусе, были конвертированы в координаты генома сборки 2011 г. (сайты связывания CTCF в эмбриональных эритроблестах 5d1-5d3, 10d1-10d3, см. рисунок 7 и рисунок 11).

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 (там же).

С помощью программы MACS 1.4.2 в обоих типах эмбриональных эритробластов в α -глобиновом локусе было выявлено четыре пика, 3 из которых перекрывались. Таким образом, вблизи α -глобиновых генов расположены три участка генома, связывающие CTCF в эритроблестах эмбрионов разных возрастов (пики 4, 5, 8, рисунок 11).

Расположение пиков, выявленных в α -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 SISRrs, не было обнаружено 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, наблюдался и в эмбриональных эритроблестах (пик 8, рисунок 11). 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 11 и 12 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 12 видно, что середины пиков 5 и 6, 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 иммунопреципитации хроматина 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 12. По-видимому, пики 5 и 6 соответствуют разным сайтам связывания CTCF.

Различия в расположении ближайших к генам α -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 α -глобинового локуса в ходе индивидуального развития гематопозитической системы кур.

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 α -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 (Zhizhou Zhang 2009), а часть сайтов связывания CTCF входит в состав GC-богатых участков генома.

Chr12:11957588-12184953 (galGal4)

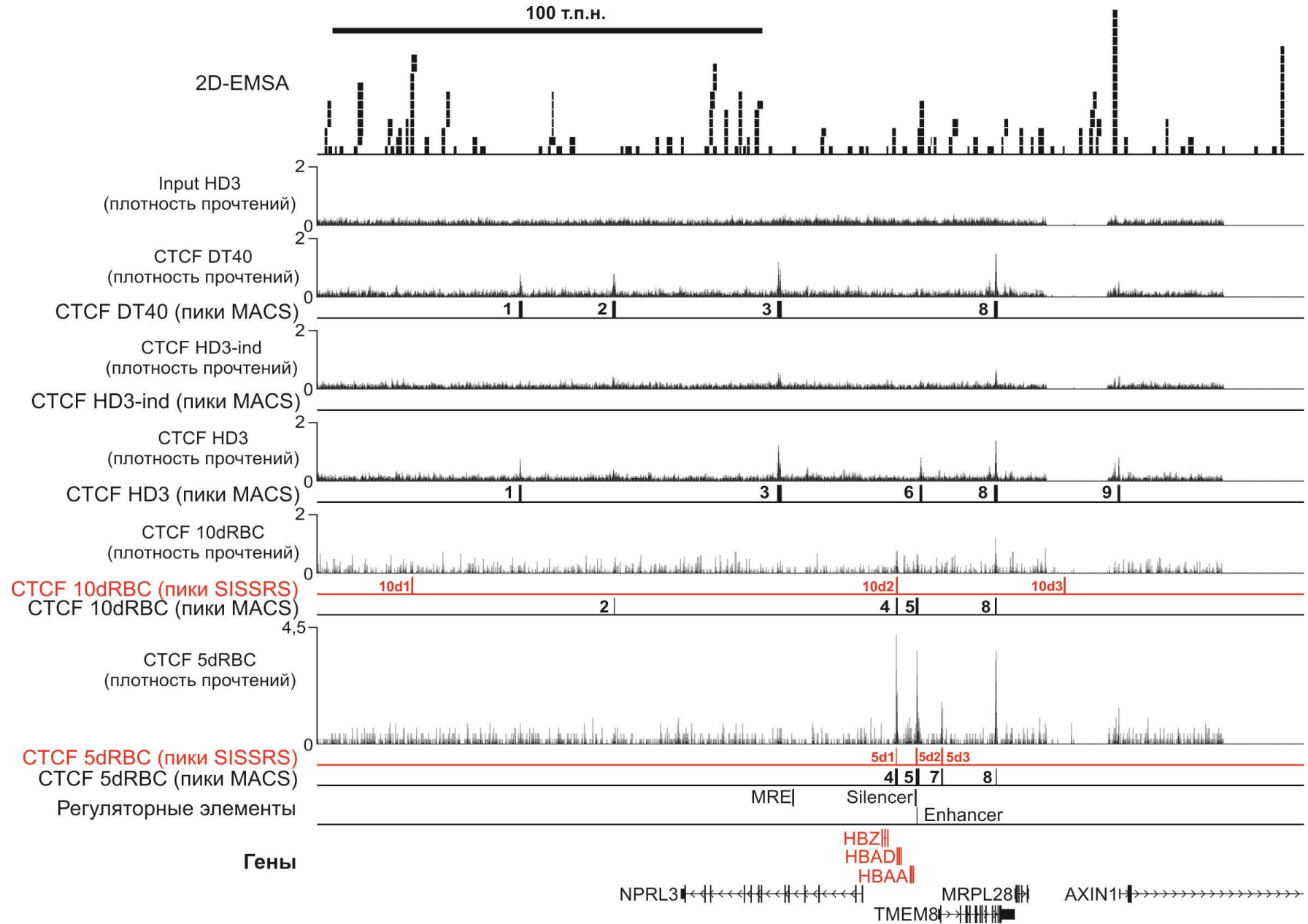


Рисунок 11. Расположение CTCF-связывающих последовательностей в α -глобиновом локусе кур. 2D-EMSA – CTCF-связывающие последовательности, выявленные путем двумерного EMSA. Input – вся выделенная после ультразвуковой дезинтеграции ДНК. CTCF – ДНК, преципитированная с антителами к CTCF. DT40, HD3-ind, HD3, 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 SISRrs – расположение пиков, выявленных программой SISRrs (Martin et al, 2011). Пики, выявленные в α -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. раздел “Сравнение результатов ChIP-seq анализа для взрослых и эмбриональных куриных клеток”). Регуляторные элементы – регуляторные элементы локуса: MRE, энхансер и сайленсер. Гены – гены локуса.

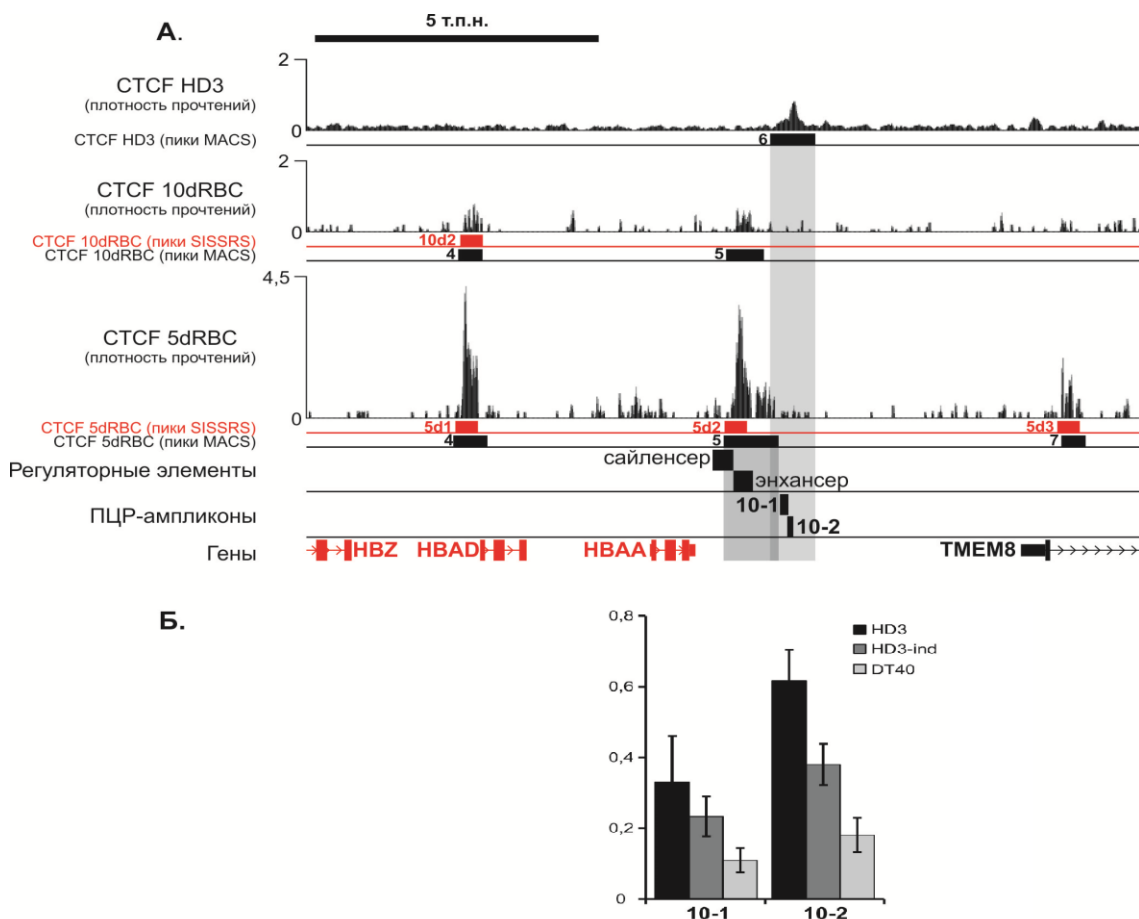


Рисунок 12. Связывание белка CTCF в области сайленсера-энхансера α -глобинового локуса в клетках линии HD3 и эмбриональных эритробластах кур. А. Карта участка α -глобинового локуса, содержащего регуляторный элемент, состоящий из сайленсера и энхансера. Показано

расположение генов (дорожка «Гены»), регуляторных элементов (дорожка «Регуляторные элементы»), распределение плотности прочтений (плотность прочтений) и пиков сигнала (пики MACS и пики SISRrs) по результатам ChIP-seq с антителами к CTCF. Обозначения дорожек, содержащих данные ChIP-seq и пики – см. рисунок 11. 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

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 α -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 α -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 α -глобинового локуса кур, способных связывать белок CTCF *in vitro*. При помощи иммунопреципитации хроматина показано, что из тринадцати выявленных участков ДНК, расположенных по близости от генов α -глобинов, 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 α -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 РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ:

Статьи по теме диссертации

1. Котова Е.С., Сорокина И.В., Акопов С.Б., Николаев Л.Г., Свердлов Е.Д. Экспрессия гена CTCF кур в клетках COS-1 и частичная очистка белка CTCF. (2013) *Биохимия* 78(8), 1122 – 1127.
2. Gushchanskaya E.S., Artemov A.V., Ulyanov S.V., Logacheva M.D., Penin A.A., Kotova E.S., Akopov S.B., Nikolaev L.G., Iarovaia O.V., Sverdlov E.D, Gavrilo A.A., Razin S.V. The clustering of CpG islands may constitute an important determinant of the 3D organization of interphase chromosomes. (2014) *Epigenetics* 9(7), 951–963.
3. Kotova E.S., Akopov S.B., Sverdlov E.D., Nikolaev L.G. Transcription factor CTCF and mammalian genome organization. (2014) *Biopolymers and Cell* 30(4), 260–272.
4. Котова Е.С., Акопов С.Б., Дидыч Д.А., Петрова Н.В., Яровая О.В., Разин С.В., Николаев Л.Г. (2016) Связывание белкового фактора CTCF в α -глобиновом локусе генома кур. *Acta Naturae* 8(1), 90–97.

Тезисы научных конференций по теме диссертации

1. Kotova E.S., Sorokina I.V., Akopov S.B., Nikolaev L.G. Chicken CTCF expression in mammalian cells. The 4-th International IMBG Conference For Young Scientists "Molecular Biology: Advances and Perspectives" 2011, Kyiv, Ukraine.
2. Котова Е.С., Сорокина И.В., Дидыч Д.А., Акопов С.Б., Николаев Л.Г., Свердлов Е.Д. Поиск CTCF-связывающих последовательностей в α -глобиновом локусе кур. Научная конференция по биоорганической химии и биотехнологии “X чтения памяти академика Юрия Анатольевича Овчинникова”. 2011. Москва, Россия.
3. Сорокина И.В., Котова Е.С., Акопов С.Б., Николаев Л.Г. Экспрессия гена *CTCF* кур в клетках млекопитающих. VIII международная научная конференция студентов, аспирантов и молодых ученых “Ломоносов”. 2011. Москва, Россия.