

**Федеральное агентство научных организаций (ФАНО России)  
Федеральное государственное бюджетное учреждение науки**

**ИНСТИТУТ БИООРГАНИЧЕСКОЙ ХИМИИ**

**им. академиков М.М. Шемякина и Ю.А. Овчинникова  
Российской академии наук**

*На правах рукописи*

**Акопов Сергей Борисович**

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

03.01.03 - молекулярная биология

Автореферат

диссертации на соискание ученой степени

доктора биологических наук

МОСКВА – 2015

Работа выполнена в лаборатории структуры и функции генов человека Института биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова РАН.

**Научный консультант:**

академик РАН Свердлов Евгений Давидович

**Официальные оппоненты:**

чл.-корр. РАН, д.б.н., проф. **Янковский Николай Казимирович**, заведующий лабораторией анализа генома Института общей генетики Российской академии наук

чл.-корр. РАН, д.х.н., проф. **Кочетков Сергей Николаевич**, заведующий лабораторией молекулярных основ действия физиологически активных соединений Института молекулярной биологии Российской академии наук

чл.-корр. РАН, д.б.н., проф. **Рысков Алексей Петрович**, заведующий лабораторией организации генома Института биологии гена Российской академии наук

**Ведущая организация:**

Федеральное государственное бюджетное учреждение науки Институт химической биологии и фундаментальной медицины Сибирского отделения Российской академии наук

Защита диссертации состоится \_\_\_\_\_ 2015 г. в \_\_\_\_\_ часов на заседании диссертационного совета Д 002.019.01 при Институте биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова Российской академии наук по адресу: 117997, Москва, ул. Миклухо-Маклая 16/10.

С диссертацией можно ознакомиться в библиотеке Института биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова РАН и на сайте Института [www.ibch.ru](http://www.ibch.ru).

Автореферат разослан \_\_\_\_\_ 2015 г.

Ученый секретарь  
диссертационного совета,  
доктор физ.-мат. наук

В.А. Олейников

## ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность проблемы.** Геном человека содержит множество функционально значимых элементов, таких как промоторы, энхансеры, сайленсеры, инсуляторы, сайты связывания гормонов и транскрипционных факторов. Однако определение одной только нуклеотидной последовательности генома человека не даёт исчерпывающей информации о расположении генов, их функциональном назначении и регуляции их экспрессии. Недавние исследования генома человека позволили идентифицировать большое число транскриптов и потенциальных регуляторных элементов, при этом регуляторных элементов в геноме существенно больше, чем генов, которые они контролируют. Кроме того регуляторные элементы нередко располагаются на больших расстояниях от генов, которые они регулируют. Регуляторные элементы часто подвергаются более быстрым эволюционным превращениям, поэтому их предсказание и поиск невозможен с использованием одних только методов биоинформатики, поскольку в первичной структуре ДНК большинства из них не выявлено высокой степени идентичности с уже известными регуляторами. Одно из актуальных направлений современной функциональной геномики - создание экспериментальных методов идентификации регуляторных элементов, а также разработка подходов к их функциональному анализу. В геномах млекопитающих существует примерно 20-25 тысяч кодирующих белки генов, что составляет один процент геномной ДНК (Lander et al., 2001), но транскрибируется гораздо большая часть генома (Carninci, 2006; Rosenbloom et al., 2009; Djebali et al., 2012). По теоретическим прогнозам в геномах эукариот функционируют сотни тысяч *цис*-регуляторных элементов. Консорциумом ENCODE в геноме человека идентифицированы ~400000 потенциальных энхансеров, и, возможно, это число может увеличиться до миллиона (Bernstein et al., 2012). Таким образом, значительная часть нашего генома, 25% и более, занята регуляторной информацией, что позволяет предположить, что типичный человеческий ген может регулироваться десятками энхансеров. Недавно появившиеся работы позволяют, таким образом, количественно охарактеризовать наиболее значимые регуляторные элементы генома человека (de Laat et al., 2013).

В среднем регуляторный «арсенал» в геномах млекопитающих содержит:

около 20000 генов;

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

545 “генных пустынь”;

до четырех энхансеров связываются с активным геном в клетке одного типа;

средний размер петли, образующейся при взаимодействии энхансера и промотора

составляет 120 т.п.о.;

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

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

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

В свете приведенных выше данных очевидна необходимость дальнейшего изучения обнаруженных, а также поиска новых последовательностей, выполняющих регуляторные функции, таких, как промоторы, энхансеры, инсуляторы, или участвующих в структурной организации ядра, таких как фрагменты, связывающиеся с ядерным матриксом (S/MAR-элементы).

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

эндогенных ретровирусов с их LTR позволило получить ряд важных новых данных о механизмах регуляции на уровне полного генома человека.

Методы масштабного поиска регуляторных элементов, которые используют в настоящее время, можно разделить на две группы: структурные и функциональные. Во многих из этих подходов используется один и тот же ключевой этап – получение высокообогащенных клонотек фрагментов ДНК, содержащих соответствующие регуляторные элементы. Структурные методы, например иммунопреципитация хроматина, основаны на структурных взаимодействиях ДНК с регуляторными белками или субклеточными структурами, а также на структурных особенностях и модификациях ДНК и хроматина (метилование ДНК, метилирование и ацетилирование гистонов). Применение этих методов ограничено нашими знаниями о транскрипционных факторах и их сайтах связывания с ДНК. Поэтому, определенное преимущество получают функциональные подходы, использующие «репортерные гены», т.е. методы, основанные на анализе функциональной активности последовательностей ДНК. Эти методы позволяют идентифицировать фрагменты генома, обладающие специфической активностью.

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

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

В представленной работе, будут описаны подходы, основанные на использовании именно таких методов экспериментального поиска и картирования регуляторных элементов в протяженных полигенных участках генома на основе их функциональной активности. В качестве объекта исследования использовали хромосому 19 человека и глобиновый локус кур.

**Цели и задачи исследования.** Целью настоящей работы являлось развитие подходов к выявлению, функциональному картированию и функциональной характеристике *цис*-регуляторных элементов внутри протяженных областей геномов и их применение для идентификации таких элементов в геномах человека и кур.

В ходе работы были поставлены и решены следующие задачи:

1. Разработать метод, позволяющий проводить поиск энхансеров в протяженных участках ДНК; с его помощью выявить и картировать в локусе *FXYD5-COX7A1* хромосомы 19 человека последовательности, проявляющие свойства энхансеров; провести анализ обнаруженных фрагментов методом сдвига электрофоретической подвижности в полиакриламидном геле (EMSA), а также определить активность обнаруженных фрагментов с помощью системы двойной люциферазной детекции;
2. С помощью разработанного нами ранее метода выявить инсуляторы и построить карту их распределения в локусе хромосомы 19 человека между генами *FXYD5* и *COX7A1*;
3. Функционально охарактеризовать потенциально инсуляторные CTCF-связывающие последовательности, расположенные в локусе *FXYD5-COX7A1* хромосомы 19 человека, в частности определить их энхансер-блокирующую активность;
4. С помощью разработанного нами ранее метода 2-мерного сдвига электрофоретической подвижности в полиакриламидном геле (EMSA) выявить сайты связывания транскрипционного фактора CTCF и построить карту их распределения в глобиновом локусе кур.
5. Функционально охарактеризовать регуляторный потенциал одиночных длинных концевых повторов (LTR) эндогенных ретровирусов человека, включая тканеспецифичность их промоторных, энхансерных и негативных регуляторных элементов.
6. Выявить регуляторные участки связывания последовательности LTR эндогенных ретровирусов человека семейства К с белковыми факторами; идентифицировать их и функционально охарактеризовать, а также выделить и охарактеризовать белки, связывающиеся с этими участками.

**Научная новизна. Теоретическая и практическая значимость работы.** В настоящей работе применены новые подходы к функциональному картированию протяженных областей генома, в частности, включающие в себя построение подробных карт *цис*-регуляторных элементов. Для этого был разработан ряд оригинальных функциональных тестов, позволяющих экспериментально идентифицировать *цис*-регуляторные элементы и определить их положение относительно генов. В работе применен разработанный нами метод функционального картирования энхансер-подобных элементов в протяженных участках генома, основанный на отборе геномных фрагментов по их способности усиливать транскрипцию репортерного гена. С помощью предложенного подхода впервые идентифицированы в клетках линии HeLa 15 потенциальных энхансеров в локусе *FXVD5-COX7A1* хромосомы 19 человека; построена карта распределения обнаруженных в работе потенциальных энхансеров. С помощью системы позитивно-негативной селекции был проведен анализ энхансер-блокирующей активности десяти CTCF-связывающих последовательностей, картированных ранее в локусе *FXVD5-COX7A1* хромосомы 19 человека. Использование метода 2-мерного сдвига электрофоретической подвижности в полиакриламидном геле, который является оригинальной модификацией известного метода торможения в геле ДНК-белковых комплексов (EMSA), позволило выявить новые сайты связывания транскрипционного фактора CTCF и построить карту их распределения в глобиновом локусе кур. С помощью разработанного нами ранее метода поиска инсуляторов, основанного на защите транскрипции репортерного гена от влияния энхансера (Акоров et al. 2006) выявлены и картированы 10 новых последовательностей в локусе *FXVD5-COX7A1* хромосомы 19 человека. Был проведен подробный анализ библиотеки потенциальных инсуляторов, позволивший выявить большую часть последовательностей, проявляющих свойства инсуляторов, в протяженном участке генома. Предложенный нами метод идентификации последовательностей, обладающих инсуляторной активностью, может с успехом использоваться при создании функциональных карт геномов, поскольку в настоящее время не существует иных путей обнаружения инсуляторов, кроме прямого анализа специфической активности.

Изучение свойств LTR представляется важной задачей, поскольку большая часть LTR не утратила регуляторный потенциал в процессе эволюции и участвует в регуляции экспрессии клеточных генов. Нами было проведено исследование промоторной, энхансерной и сайленсерной активностей одиночного LTR семейства HERV-K(HML-2), в котором показана зависимость транскрипционной активности LTR от типа клеток. Выявлена способность специфически связываться с клеточными белковыми факторами центральной и 3'-концевой области LTR HERV-K, в частности области негативного регуляторного элемента. Показано, что

в клетках, где LTR HERV-K обладает энхансерной активностью, с центральной областью LTR связывается дополнительный клеточный фактор. Была проанализирована тканеспецифичность белковых факторов, связывающихся с определённым участком длинного концевой повтора человеческого эндогенного ретровируса семейства K. Был разработан метод аффинной элюции специфических ДНК-связывающих белков с гепарин-агарозной колонки, и с его помощью выделены белки, связывающиеся с 5'-концом области U3 LTR HERV-K. Один из белков был идентифицирован и охарактеризован методами масс-спектрометрии и иммунохимии. Используемые в настоящей работе методы открывают возможность создания исчерпывающих функциональных карт протяженных локусов и полных геномов.

**Апробация результатов работы.** Основные результаты работы были представлены на научных конференциях и симпозиумах, в том числе: 2-м рабочем совещании «Retrotransposons and genome evolution» (Сочи, 2003), XV Всероссийском симпозиуме «Структура и функции клеточного ядра» (Санкт-Петербург, 2005), IV съезде Российского общества биохимиков и молекулярных биологов (Новосибирск, 2008), симпозиуме 29<sup>th</sup> meeting on Retroviruses (Cold Spring Harbor, 2004), чтениях, посвященных памяти академика Ю.А. Овчинникова, конгрессах HUPO Human Proteome Organization (Монреаль, 2003г.), FEBS Federation-of-European-Biochemical-Societies (Будапешт 2005), HUGO Human Genome Meeting (Хельсинки, 2006 г.), 3<sup>rd</sup> ESF Functional Genomics Conference “Functional genomics and Disease” (Инсбрук, 2008), конференциях «Постгеномные методы анализа в биологии, лабораторной и клинической медицине: геномика, протеомика» (Новосибирск, 2011), «Molecular biology: advances and perspectives» (Киев, 2011).

**Структура и объем диссертации.** Диссертационная работа состоит из введения, обзора литературы, описания материалов и методов исследования, изложения результатов работы и их обсуждения, выводов и списка литературы. Работа изложена на 203 страницах, содержит 50 рисунков и 13 таблиц. Список литературы включает 448 источников.

### **Публикации**

Диссертация обобщает данные 25 основных статей.

## **РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ**

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



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

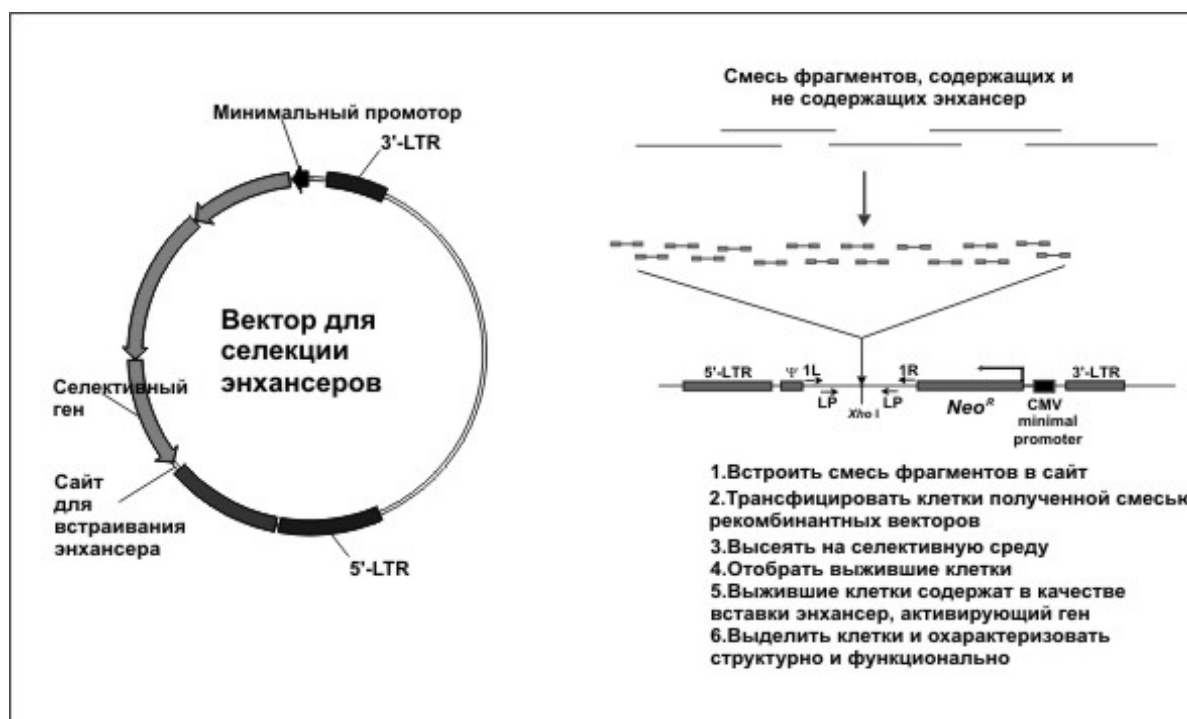


Рис. 1. Стратегия выявления энхансер-подобных элементов в протяженных участках генома.

Эту общую стратегию, разработанную совместно с Л.Г. Николаевым и И.П. Черновым, для практического использования следовало детализировать. Для селекции использовали ретровирусные векторы, преимущественно встраивающиеся в активно транскрибирующиеся участки генома, в результате чего формируются стабильные клеточные линии, обладающие высоким уровнем экспрессии исследуемых генов и не содержащие внутренних перестроек.

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

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

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

### 2.1. Создание векторных конструкций для селекции энхансер-подобных элементов

Для создания конструкций, предназначенных для поиска энхансеров, мы использовали коммерческий экспрессионный самоинактивирующийся ретровирусный вектор pQCXIX (Clontech, США) и сконструированную Е.В. Снежковым плазмиду pCMV-SGTN, содержащую ген составного белка *GTN*, под контролем минимального промотора и энхансера цитомегаловируса (CMV). Ген *GTN* собран из фрагментов, кодирующих тимидинкиназу вируса простого герпеса (*HSV-tk*), неомицинофосфотрансферазу II (*NPTII*), придающую клеткам устойчивость к генетицину (G-418), и ген зеленого флуоресцентного белка EGFP.

Из вектора pQCXIX обработкой эндонуклеазами рестрикции *EcoRV* и *XbaI* были удалены промотор и энхансер цитомегаловируса, внутренний сайт посадки рибосомы (IRES) и полилинкер. Одновременно плазмиду pCMV-SGTN обработали эндонуклеазой *NotI*, выступающие рестриктные концы достроили фрагментом Кленова ДНК-полимеразы I *E. coli*, и провели гидролиз эндонуклеазой *AsuNI*. Часть плазмиды, содержащая минимальный промотор цитомегаловируса и ген *GTN*, была лигирована с фрагментом pQCXIX *EcoRV-XbaI*. В полученную кассету по сайту *ClaI* был лигирован синтетический полилинкер (CGATCTAGACTCGAGAATTC), содержащий сайты рестрикции эндонуклеаз *ClaI*, *XbaI*, *XhoI* и *EcoRI*. Полученную ретровирусную конструкцию назвали pQCXIX-ENH. Схема вектора pQCXIX-ENH представлена на рис. 2. Для предотвращения интерференции ретровирусных промоторов ген *NPTII* транскрибируется в направлении, противоположном тому, в котором транскрибируются ретровирусные гены. Структуру полученного вектора проверили секвенированием.

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

В полученную конструкцию клонировали фрагменты библиотеки локуса *FXVD5-COX7A1*. Кроме того, были приготовлены две контрольные плазмиды. Положительный контроль pQCXIX-Enh(+) содержал кассету *GTN* под контролем энхансера и промотора цитомегаловируса (CMV).

Для этого плазмиду pCMV-SGTN гидролизовали *ClaI* и *AsuNI* и выступающие концы достраивали фрагментом Кленова. Фрагмент, содержащий ген *GTN*, минимальный промотор и энхансер цитомегаловируса, лигировали с фрагментом pQCXIX *EcoRV-XbaI*. Конструкция содержала в регуляторной области перед геном *GTN* минимальный промотор и энхансер цитомегаловируса. Отрицательный контроль pQCXIX-Enh(-) содержал кассету *GTN* без промотора и энхансера. Для этого плазмиду pCMV-SGTN обрабатывали *BamHI* и *AsuNI* и

выступающие концы достраивали фрагментом Кленова. Фрагмент, содержащий только ген *GTN*, лигировали с фрагментом pQCXIX *EcoRV-XbaI*.

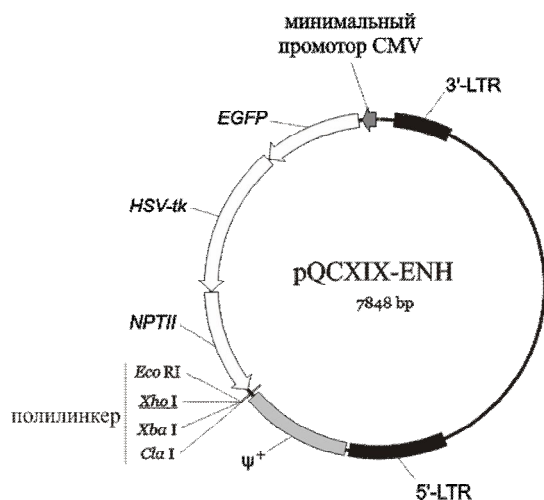


Рис. 2. Схема самоинактивирующегося ретровирусного вектора pQCXIX-ENH, предназначенного для поиска энхансеров в длинных геномных последовательностях. Вектор содержит 5'- и 3'- длинные концевые повторы (LTR) и сигнал упаковки ( $\psi^+$ ). Лигирование в вектор фрагментов библиотеки осуществлялось по сайту рестрикции *XhoI*. *HSV-tk* – ген тимидинкиназы вируса простого герпеса; *NPTII* – ген неомифосфотрансферазы II, придающий клеткам устойчивость к генетицину (G-418); *EGFP* - ген зеленого флуоресцентного белка.

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

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

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

содержащих вставку ( $\sim 1,3 \times 10^4$ ), более чем в четыре раза превышает число уникальных последовательностей, что указывает на высокую репрезентативность библиотеки.

Из выросших клонов выделяли плазмидную ДНК, представляющую собой ретровирусные конструкции рQСХІХ-Enh(L), содержащие фрагменты библиотеки локуса *FXYD5-COX7A1* хромосомы 19 человека.

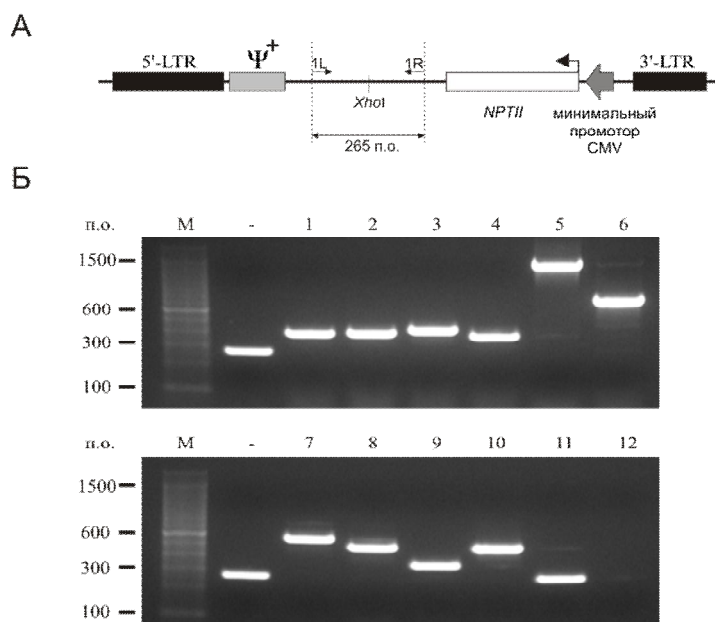


Рис. 3. ПЦР-скрининг клонов, трансформированных рQСХІХ-ENH(L), на наличие вставки. А) Схема плазмиды рQСХІХ-ENH в линейаризованном виде. Плаزمида содержит уникальный сайт узнавания эндонуклеазы рестрикции *XhoI*, по которому была лигирована библиотека локуса *FXYD5-COX7A1* хромосомы 19 человека. 1L и 1R – плазмидные праймеры, фланкирующие сайт *XhoI*. Б) Электрофореграмма продуктов ПЦР-амплификации клонов, трансформированных рQСХІХ-ENH(L). ПЦР проводили с

использованием праймеров 1L и 1R. М – маркер, (-) – ПЦР-амплификация вектора рQСХІХ-ENH (контрольный вектор без вставки). Цифры указывают номер клона. Клоны 11 и 12 не содержат вставки.

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

Всего было получено четыре популяции вирусных частиц pQCXIX-Enh, pQCXIX-Enh(L), pQCXIX-Enh(+) и pQCXIX-Enh(-). Для определения концентрации частиц, способных инфицировать клетки использовали, экспресс-метод *Byun et al.* (Byun et al., 1996). Метод позволяет определить нижний предел значений титра вируса путем инфицирования клеток NIH3T3 с их последующей селекцией на среде, содержащей генетицин G418. Высокий титр вирусных частиц pQCXIX-Enh(L), содержащих фрагменты библиотеки, важен для сохранения репрезентативности библиотеки. В наших экспериментах для конструкций pQCXIX-Enh(L) титр составил не менее  $5 \times 10^3$  частиц/мл. Фактически титр был значительно выше, так как не все вирусные частицы популяции pQCXIX-Enh(L) способны обеспечить зараженные клетки необходимым уровнем неомидинфосфотрансферазы II (GTN) из-за отсутствия энхансерной активности у части фрагментов библиотеки.

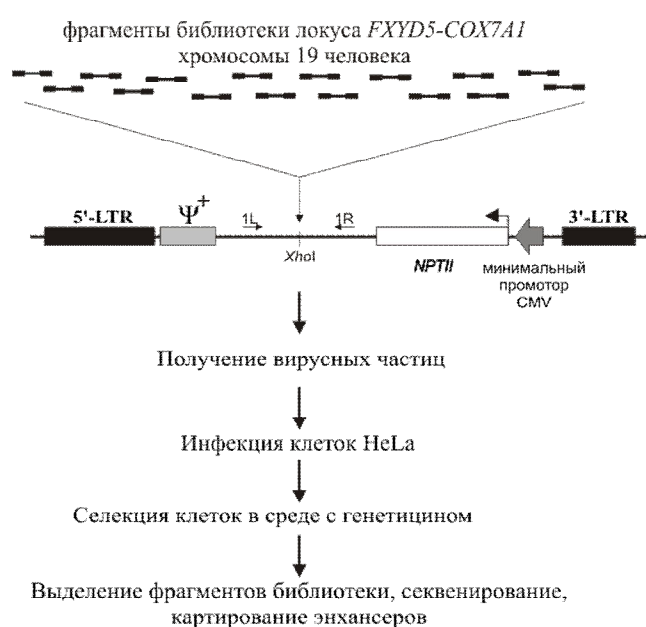


Рис. 4. Основные этапы селекции последовательностей, обладающих энхансерной активностью. Фрагменты библиотеки локуса *FXVD5-COX7A1*, полученной путем расщепления 30 космид, содержащих перекрывающиеся последовательности области хромосомы 19 человека длиной 1 млн. п.о., были клонированы в вектор pQCXIX-Enh(+) по сайту XhoI. Полученным набором конструктов (pQCXIX-Enh(L)) трансфицировали клетки пакующей линии Phoenix-AMPHO для получения вирусных частиц. Вирусными частицами инфицировали клетки HeLa и проводили селекцию в среде с генетицином. Из

выживших селектированных клеток выделяли геномную ДНК с фрагментами библиотеки, обладающими энхансерной активностью. Выделенные фрагменты секвенировали и определяли их местоположение в геноме человека.

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

При инфицировании вирусными частицами происходит интеграция кДНК-копии вирусной мРНК в геном хозяйской клетки. В случае, если интегрированный провирус содержал фрагмент библиотеки, проявляющий энхансерную активность, он увеличивал активность гена

селективного маркера (*NPTII* в составе белка GTN), и клетки, содержащие такую конструкцию, выживали в среде с генетицином. После раунда селекции из клеток выделяли геномную ДНК и анализировали при помощи двустадийной ПЦР (nested-PCR) с использованием а) плазмидных праймеров (1L и 1R) и б) с помощью библиотечного праймера 27R (рис.5А, раунд 1).

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

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

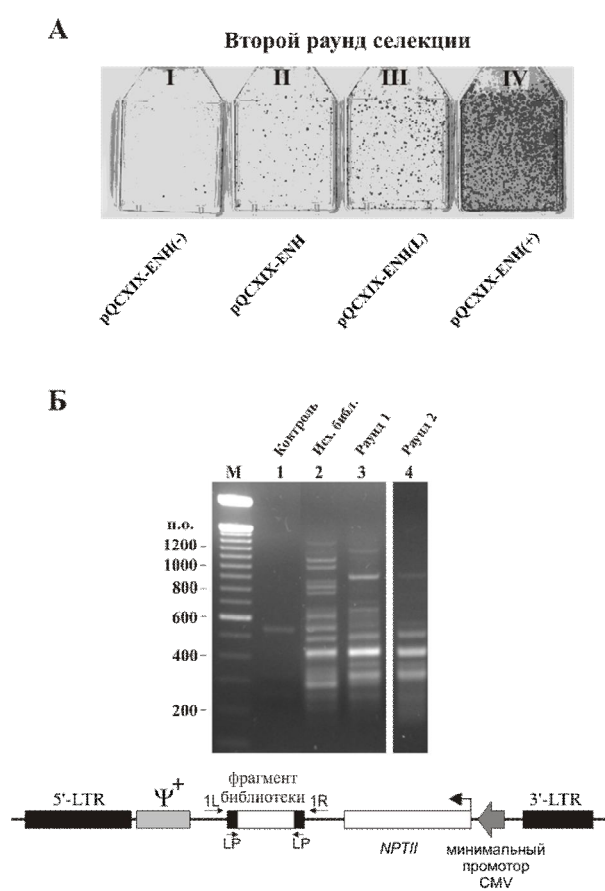


Рис. 5. Отбор потенциальных энхансеров из фрагментов библиотеки локуса *FXVD5-COX7A1*, расположенного на хромосоме 19 человека. А) четыре популяции инфицированных клеток HeLa после второго раунда селекции. Окраска Coomassie Blue (объяснение в тексте). Б) электрофорез в агарозном геле продуктов двустадийной ПЦР-амплификации фрагментов исходной библиотеки (дор. 2), после первого (дор. 3) и второго (дор. 4) раундов селекции. Первую стадию амплификации проводили с использованием внешних праймеров 1L и 1R. На второй стадии использовали библиотечный праймер LP, фланкирующий фрагменты библиотеки.

После инфицирования клеток HeLa каждую популяцию при пассировании высевали на два флакона и проводили селекцию в среде с генетицином на протяжении примерно двух недель. После селекции по одному из двух флаконов с популяциями клеток, инфицированных контрольными и ретровирусными конструкциями, содержащими фрагменты библиотеки, использовали для подсчета селектированных колоний. Для этого клетки во флаконах фиксировали при помощи метанола и уксусной кислоты и окрашивали Coomassie Blue (рис. 5). Клетки из оставшихся флаконов собирали и выделяли геномную ДНК для дальнейшего анализа. Из рис.5 видно, что наибольшее количество выживших после селекции и образующих колонии клеток содержит популяция, инфицированная позитивным контролем рQCXIX-Enh(+), содержащим энхансер и промотор цитомегаловируса в регуляторной области гена *GTN*. Популяция клеток, инфицированная конструктором рQCXIX-Enh(-), содержащим только минимальный промотор (негативный контроль), содержит гораздо меньше выживших клеток. Флакон с клетками, в которых проходил отбор фрагментов библиотеки рQCXIX-Enh(L), содержит также меньше клеток по сравнению с позитивным контролем, но больше чем в популяции, инфицированной конструкцией с негативным контролем рQCXIX-Enh(-), что свидетельствует о присутствии среди фрагментов библиотеки последовательностей, обладающих энхансерной активностью. Среди клеток, инфицированных конструкцией рQCXIX-Enh(-), выживают единичные клетки, несмотря на то, что сама конструкция рQCXIX-Enh(-) содержит ген *GTN* без энхансера и промотора. Вероятно, интеграция в этих клетках происходила вблизи активного промотора. Подобные события весьма редки, что и демонстрирует малое количество выросших колоний.

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

### 2.3. Анализ отобранных потенциальных энхансерных фрагментов

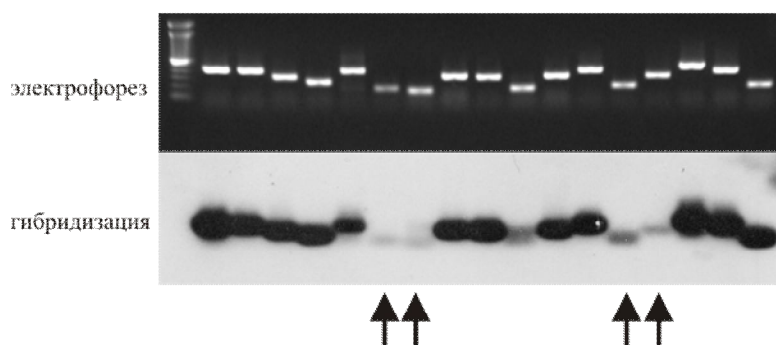
Фрагменты библиотеки, амплифицированные после второго раунда селекции, лигировали в плазмидный вектор рGEM-T Easy (Promega) и трансформировали ими клетки *E.coli*. Полученные клоны ранжировали в 96-луночные планшеты. Полученную клонотекку потенциальных энхансеров использовали для дальнейшего анализа.

Среди выделенных фрагментов, невзирая на два раунда селекции, часть могла быть представлена последовательностями, не являющимися энхансерами. Для отсеивания этих

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

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

А



Б

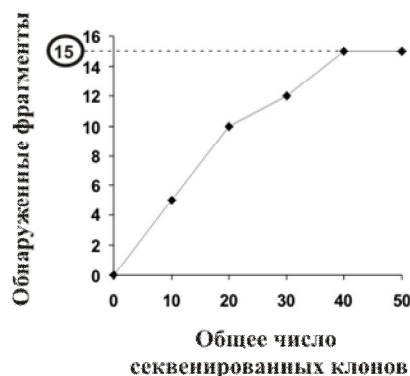


Рис. 6. Анализ клонов библиотеки потенциальных энхансеров. А) Гибридизация индивидуальных клонов с тотальной библиотекой, полученной после двух раундов селекции (показана часть проанализированных клонов). Верхняя панель – электрофорез в 1% агарозном геле продуктов ПЦР-амплификации фрагментов библиотеки, выделенных из индивидуальных клонов. Нижняя панель – результат гибридизации амплифицированных фрагментов с тотальной библиотекой, полученной после двух раундов селекции. Клоны, давшие слабый гибридизационный сигнал (отмечены стрелками), были исключены из дальнейшего



анализа. Б) Размер полученной библиотеки потенциальных энхансеров. Показана зависимость числа идентифицированных уникальных последовательностей от общего числа секвенированных клонов.

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

Далее была определена функциональная активность полученных фрагментов. Отличительной особенностью функциональных элементов, в том числе энхансеров, является их способность формировать ДНК-белковые комплексы. Большинство моделей функционирования энхансеров предполагает образование комплекса с белками, взаимодействие с промоторным комплексом, и выпетливание ДНК между ними. Мы проверили способность нескольких выявленных энхансер-подобных элементов взаимодействовать с ядерными белками при помощи метода сдвига электрофоретической подвижности (EMSA), позволяющего выявлять ДНК-белковые взаимодействия. EMSA-анализ проводили с использованием ядерного экстракта из клеток HeLa, поскольку отбор энхансеров проводили с использованием этой клеточной линии.

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

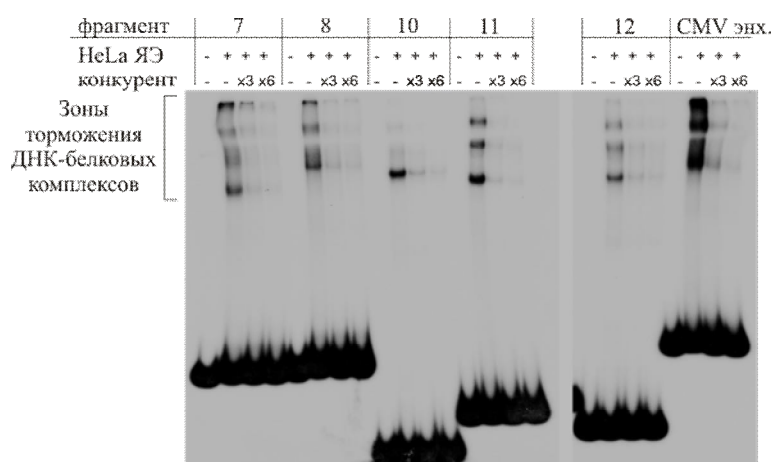


Рис. 7. Проверка связывания обнаруженных потенциальных энхансеров с белками ядерного экстракта (ЯЭ) из клеток линии HeLa методом сдвига электрофоретической подвижности (EMSA).

Из рис. 7 видно, что все исследованные фрагменты образуют

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

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

Энхансерную активность 15 выявленных последовательностей измеряли с использованием двойной люциферазной системы (Promega). Полученными конструкциями проводили транзientную трансфекцию клеток HeLa с использованием реагента Lipofectamine 2000. В качестве контролей использовали pGL3pv без вставки и pGL3 control vector (pGL3cv), содержащий промотор и энхансер SV40.

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

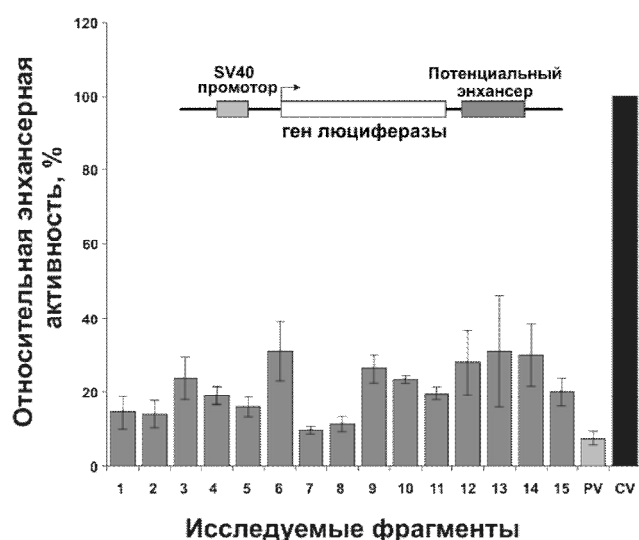


Рис. 8. Способность обнаруженных энхансеров активировать промотор SV40 в клетках HeLa. Приведены результаты трех независимых трансфекций. Активность потенциальных энхансеров представлена в процентах от активности энхансера SV40, принятой за 100%. PV – в трансфекции использовали контрольный вектор pGL3pv,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

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

Характеристика обнаруженных потенциальных энхансеров представлена в таблице 1. Положение энхансер-подобных элементов в исследуемом локусе изображено на рис. 9. Шесть фрагментов были локализованы внутри генов (в интронах), пять фрагментов располагаются с 5'-стороны от генов, один предполагаемый энхансер расположен с 3'-стороны от гена и три

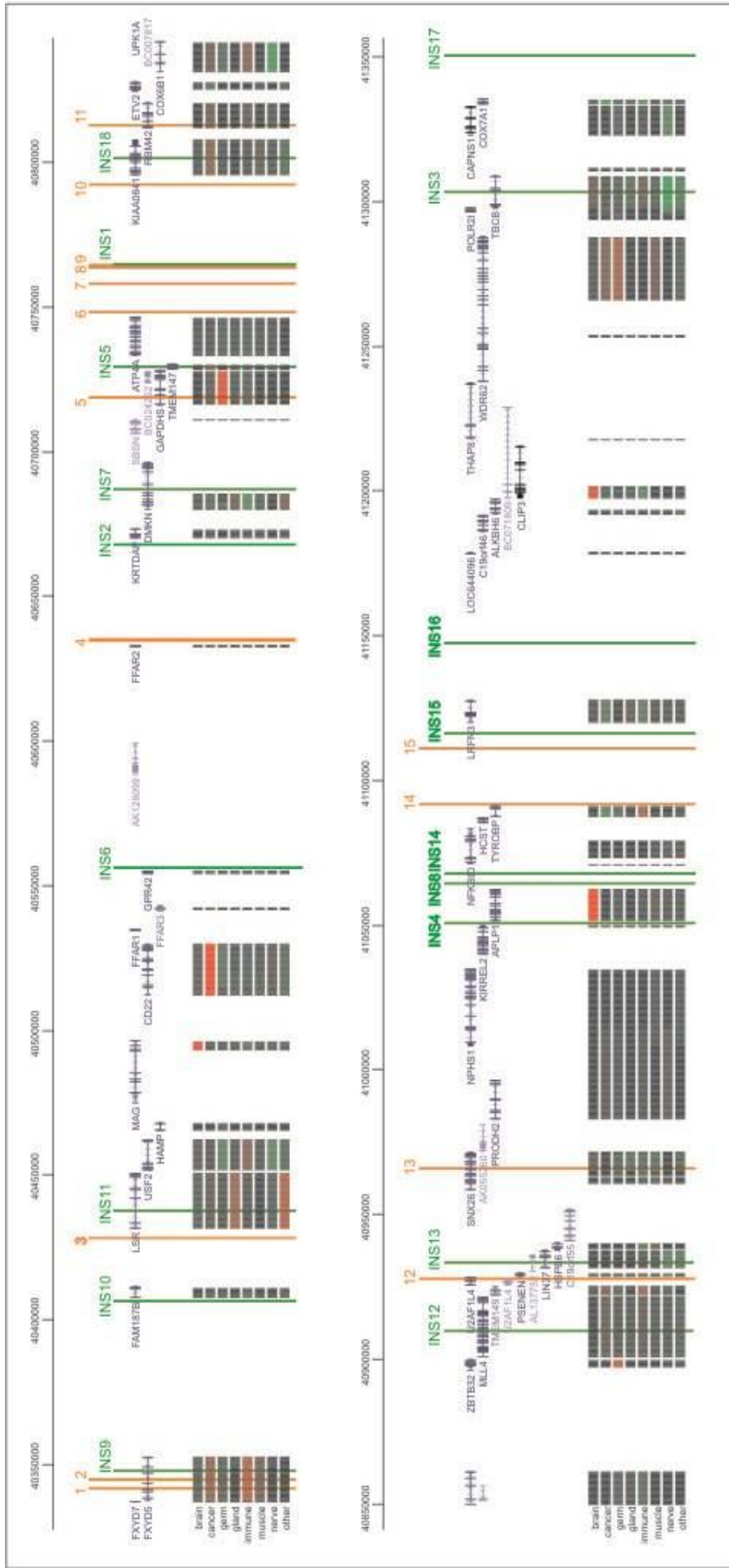
фрагмента в межгенных областях (ближайшие гены находятся на расстоянии более 10 т.п.о.). Представляет интерес то, что два из трех потенциальных энхансеров, выявленных в межгенных областях, перекрываются с повторяющимися элементами, Alu-элементом и длинным концевым повтором семейства HERV-K эндогенных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

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

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

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

Инсуляторы – регуляторные элементы генома, способные блокировать взаимодействие между энхансером и промотором, если находятся между ними. Инсуляторы также защищают фланкированный ими трансген от эффекта положения. Основываясь на способности большинства инсуляторов блокировать гетерологичные энхансеры, нами предложена стратегия для выявления инсулятор-подобных элементов в протяженных участках генома. Мы разработали функциональный подход, который позволяет отбирать инсуляторы из смеси геномных фрагментов. В нём применяется процедура позитивно-негативной селекции с использованием гена тимидинкиназы вируса простого герпеса (рис. 10). Анализируемый фрагмент, проявляющий инсуляторную активность, в конструкции стабильно интегрированной в геном, блокирует взаимодействие энхансера с промотором, что приводит к подавлению экспрессии тимидинкиназы вируса простого герпеса (HSV-tk). Тимидинкиназа сначала



**Рис. 9.** Физическая карта локуса *FXYD5-COX7A1* хромосомы 19 человека длиной 1млн п.о. Зеленые вертикальные линии отображают положение потенциальных инсуляторных последовательностей, идентифицированных нами ранее (№ 1-8) (Акоров et al., 2006) и позднее (№ 9-18). Оранжевые вертикальные линии отображают положение выявленных нами потенциальных энхансеров. Синим отмечены позиции генов, столбцы внизу воспроизводят данные по экспрессии генов локуса в разных тканях человека (красный - высокий уровень экспрессии, зеленый - низкий) (Su et al., 2004). Карта приготовлена с помощью Human Genome Browser (сборка от марта 2006 г.)

Таблица 1. Характеристика потенциальных энхансеров, выявленных в области *FXVD5-COX7A1* хромосомы 19 человека.

№ Энхансера	Координаты на хромосоме 19**	Длина, п.о.	Положение относительно генов локуса	Активность*	Перекрытие с повторяющимися элементами
1	40341821–40342118	298	Интрон 3 гена <i>FXVD5</i>	2.2	
2	40344506–40344897	392	Интрон 4 гена <i>FXVD5</i>	2.1	
3	40427994–40428292	299	~2 т.п.о. от 5'-конца гена <i>LSR</i>	3.5	1-232 Alu-Y
4	40634421–40635308	888	1 т.п.о. от 3'-конца гена <i>FFAR2</i>	2.8	
5	40718720–40719018	299	Интрон 1 гена <i>GAPDHS</i>	2.3	
6	40748237–40748490	254	2 т.п.о. от 5'-конца гена <i>ATP4A</i>	4.7	
7	40757795–40758182	388	Межгенное***	1.4	1-388 HERV-K LTR
8	40763469–40763799	331	Межгенное***	1.6	
9	40764184–40764486	303	Межгенное***	3.9	1-240 Alu-Y
10	40792389–40792543	156	3 т.п.о. от 5'-конца гена <i>KIAA0841</i>	2.8	
11	40812850–40813109	260	Интрон 2 гена <i>RBM42</i>	2.8	1-220 Alu-Sc
12	40927410–40927666	257	Интрон 2 гена <i>U2AF1L4</i> ; 0.7 т.п.о. от 5'-конца гена <i>PSENNEN</i>	4.3	
13	40965824–40966019	196	Интрон 15 гена <i>SNX26</i>	4.9	
14	41091517–41091714	198	0.5 т.п.о. от 5'-конца гена <i>TYROBP</i>	4.3	
15	41110960–41111236	277	8 т.п.о. от 5'-конца гена <i>LRFN3</i>	2.9	70-277 Alu-Sx

\*Данные по транзиентным трансфекциям в двойной люциферазной системе детекции (Рис. 7).

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

\*\*\*На расстоянии  $> 10^3$  п.о. от ближайших генов.

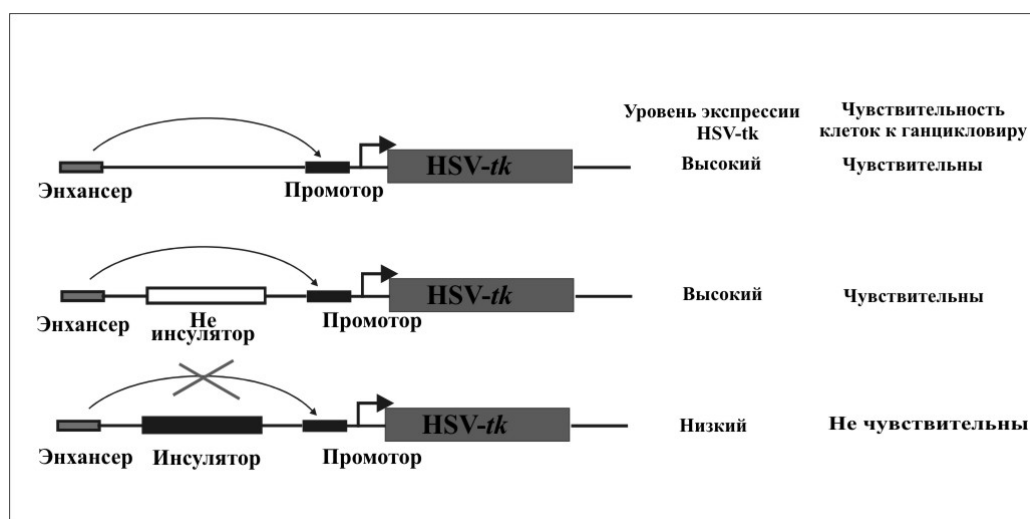


Рис. 10. Механизм селекции инсулятор-подобных последовательностей.

фосфорилирует ганцикловир, который является синтетическим аналогом 2-дезоксигуанозина, затем фосфорилированный ганцикловир фосфорилируется клеточными киназами до трифосфата и встраивается в растущую при репликации цепь ДНК, блокируя ее, что приводит к гибели клеток. При селекции клеток ганцикловиром выживут клетки, в которых подавлен синтез *HSV-tk* вследствие блокирования энхансера инсулятором.

Стратегия идентификации инсулятор-подобных последовательностей в протяженных фрагментах геномов (рис. 11) предполагает:

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

### 3.1. Система для экспериментального поиска инсуляторов

Общая схема метода представлена на рис. 11А. Для селекции потенциальных инсуляторов была приготовлена конструкция, названная рPNT/EmP, схема которой представлена на рис. 11Б. За основу была взята плаزمида рPNT (Tybulewicz et al., 1991), содержащая ген устойчивости к неомифину ( $Neo^R$ ) и ген тимидинкиназы вируса простого герпеса, придающий экспрессирующим его клеткам чувствительность к ганцикловиру (Herpes Simplex Virus, *HSV-tk*,  $GANC^S$ ), позволяющие проводить позитивно-негативную селекцию. В полученной конструкции происходит эффективная экспрессия гена *HSV-tk*, однако, если между промотором и энхансером этого гена клонировать последовательность ДНК, способную блокировать действие энхансера (инсулятор), экспрессия гена *HSV-tk* в стабильно трансфицированных такой плазмидой клетках прекращается или значительно снижается, и содержащие эту последовательность клетки становятся устойчивыми к ганцикловиру.

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

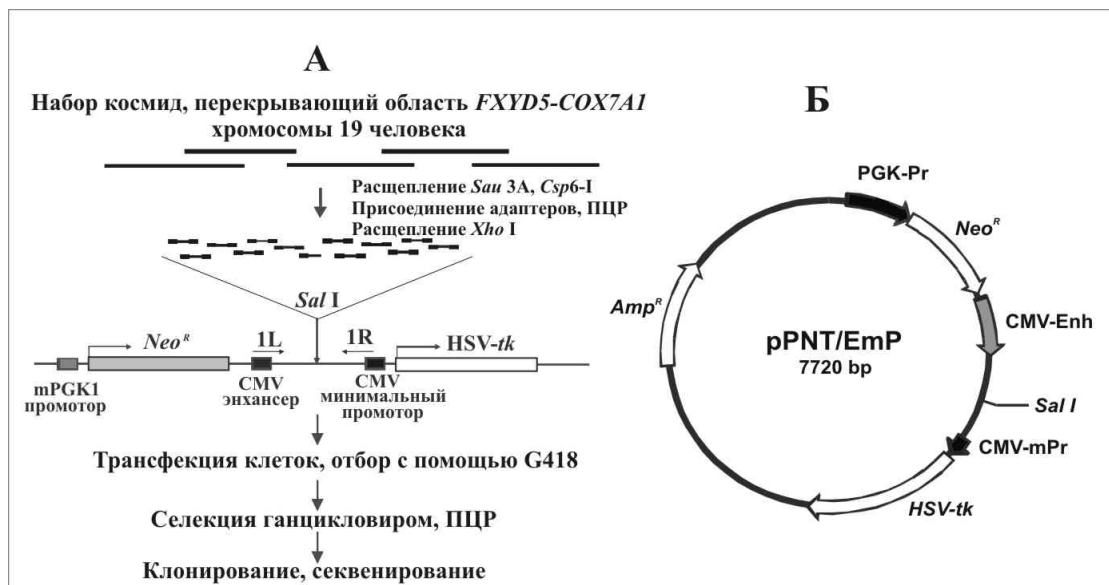


Рис. 11. (А) Общая схема отбора инсуляторов. mPGK1 - промотор фосфоглицераткиназы человека; *Neo<sup>R</sup>* - ген неомифосфотрансферазы; *HSV-tk* - ген тимидинкиназы вируса простого герпеса. (Б) Схема плазмиды pPNT/EmP для позитивно-негативной селекции инсуляторов.

Поиск потенциальных инсуляторов проводили в локусе хромосомы 19 человека длиной  $10^6$  пар оснований, расположенного между генами *FXYD5* и *COX7A1*. Для получения библиотеки фрагментов, предназначенных для отбора потенциальных инсуляторов, ДНК 30 космид, перекрывающих область *FXYD5-COX7A1*, расщепляли по-отдельности рестриктазами *Sau3A* и *Csp6I*. К полученным фрагментам (длиной 200–700 п.о.) при помощи синтезированных адаптеров был лигирован библиотечный праймер LP, обеспечивающий ПЦР-амплификацию фрагментов библиотеки и содержащий сайт расщепления *XhoI*. Затем обе библиотеки смешивали. Полученную библиотеку амплифицировали при помощи ПЦР с использованием праймера LP и обрабатывали рестриктазой *XhoI*. Продукты, полученные после рестриктной обработки, клонировали в плазмиду pPNT/EmP по сайту *SalI*. Для репрезентативной библиотеки было получено 12000 колоний, что соответствует 4-кратному перекрытию области *FXYD5-COX7A1*. Выросшие клоны смывали с чашек Петри и использовали для выделения плазмидной ДНК.

Клетки линии CHO трансфицировали электропорацией полученными на предыдущем этапе плазмидами, которые предварительно линейаризовали при помощи эндонуклеазы рестрикции *SspI*. Условия электропорации (20 мс, 360 В, 800 мкФ) для клеток культуры CHO были отработаны автором в лаборатории проф. Боде в Национальном Центре Биотехнологии (Германия), и обеспечивают интеграцию одной копии линейаризованной плазмиды в геном. В тех же условиях проводили трансфекции клеток контрольными плазмидами pPNT/mP и pPNT/EmP. На этапе позитивной селекции клетки, в геном которых интегрировал вектор,

отбирали по устойчивости к генетицину (G418). После этого проводили негативную селекцию генетицин-устойчивых клеток ганцикловиrom. 100%-ная гибель клеток, содержащих контрольную плазмиду pPNT/EmP, определяющую чувствительность клеток к ганцикловиру, служила подтверждением успешно проведенной селекции. При этом популяция клеток, трансфицированных контрольной плазмидой pPNT/mP, была устойчива к ганцикловиру.

Чувствительность клеток к ганцикловиру может наблюдаться в ситуации, если клонированный фрагмент ДНК обладает свойствами сайленсера. Впрочем, последовательность со свойствами сайленсера не обладает направленностью, и будет подавлять транскрипцию гена устойчивости к неомицину, что приведет к гибели таких клеток на этапе позитивной селекции. Для экспериментального подтверждения этого предположения мы клонировали три из идентифицированных последовательностей (№ 2, 5 и 8,) в плазмиду pGL3-promoter vector (Promega) с 5'-конца от промотора SV40 и регистрировали люциферазную активность этих конструкций в транзientной трансфекции клеток HeLa. Сайленсерной активности не наблюдалось во всех трех конструкциях (не показано).

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

На рис. 12А представлены результаты ПЦР с праймерами 2L (GGATTTCCAAGTCTCCAGGGGAT) и 2R (ACCTCCCACCGTACACGCCT) пула плазмид перед трансфекцией (дор. 1) и результаты ПЦР с теми же праймерами на матрице геномной ДНК из клеток, трансфицированных библиотекой в плазмиде pPNT/EmP и прошедших негативную и позитивную селекции (дор. 2). Видно, что в результате селекции исходная библиотека, содержащая набор большого числа фрагментов разной длины, превращается в «лестницу» из ограниченного числа фрагментов.

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



выявил 10 новых последовательностей. Эти последовательности нанесли на карту исследуемой хромосомы 19 между генами *FXYD5* и *COX7A1* (рис. 9, инсуляторы 9-18). Их свойства представлены в таблице 3.

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

### 3.2. Анализ расположения инсуляторов в геноме

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

**Таблица 2.** Расположение потенциальных инсуляторов в области *FXYD5-COX7A1* хромосомы 19 человека.

№	Координаты на хромосоме 19 <sup>1</sup>	Длина (п.о.)	Расположение относительно генов
9	40347492-40347685	194	6-й интрон гена <i>FXYD5</i>
10	40406076-40406590	515	~1 т.п.о. с 3'-стороны от гена <i>TMEM162</i>
11	40437627-40437699	73	2-й интрон гена <i>LSR</i>
12	40909389-40909561	173	14-й интрон гена <i>MLL4</i>
13	40932935-40933049	114	1-й интрон гена <i>LIN37</i>
14	41067647-41067745	99	~4 т.п.о. с 3'-стороны от гена <i>NFKBID</i>
15	41116085-41116309	225	~4 т.п.о. с 5'-стороны от гена <i>LRFN3</i>
16	41147417-41147594	178	Межгенное <sup>2</sup>
17	41350295-41350437	143	Межгенное <sup>2</sup>
18	40801090-40801286	197	10-й и 11-й интроны, 11-й экзон гена <i>KIAA0841</i>

<sup>1</sup>По данным Human Genome Browser [Kuhn, 2007], сборка от марта 2006 г.

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

Как отмечалось выше, термином «инсулятор» обозначают как энхансер-блокирующие элементы, так и пограничные элементы генома, отделяющие участки хроматина с различной структурой или активностью. Поскольку мы идентифицировали инсуляторы только по энхансер-блокирующей активности, то, в принципе, они могут не проявлять «барьерной» активности, хотя она и располагается, скорее всего, вблизи этих элементов. Тем не менее, в отдельных случаях можно выделить гипотетические домены, заключенные между инсуляторами.

В локусе *FXYD5-COX7A1* содержится кластер генов, кодирующих рецепторы, ассоциированные с G-белком, – *FFAR1*, *FFAR2*, *FFAR3* и *GPR42P*. Ранее было показано, что гены *FFAR3* и *FFAR2* имеют различные профили экспрессии (Brown et al., 2003). Позиции

инсулятора 6, находящегося между генами *FFAR3* и *FFAR2*, и инсулятора 2, отделяющего *FFAR2* от гена *KRTDAP*, экспрессирующегося преимущественно в эпидермальных кератиноцитах (Bazzi et al., 2007), соотносятся с независимой экспрессией этих генов. Гены *ATP4A* и *KIAA0841* располагаются каждый в своем гипотетическом домене (между инсуляторами 5 и 1, 1 и 18 соответственно). Ген *ATP4A* транскрибируется преимущественно в тканях желудка (Chalaya et al., 2006; Herrmann et al., 2007) и явно отличается в этом отношении от окружающих генов.

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

Из 10 идентифицированных последовательностей восемь содержат фрагменты повторяющихся элементов генома (табл.3). Инсулятор 9 представляет собой фрагмент LINE (Long Interspersed Element), а инсуляторы 10 и 17 – LTR эндогенных ретровирусов человека MaLR и ERVL, соответственно. Другие потенциальные инсуляторы перекрываются с Alu-повторами (инсуляторы 12, 15–17). Последовательность 14 представляет собой повторяющийся элемент из семейства транспозонов Tc2.

**Таблица 3.** Повторяющиеся элементы генома, входящие в состав потенциальных инсуляторов.

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

Имеются указания на то, что некоторые Alu-повторы могут обладать инсуляторными свойствами (Ruiz-Narvaez et al., 2008; Willoughby et al., 2000). Область *FXYD5-COX7A1*, однако,

содержит более 700 Alu-элементов, и обнаружение только двух из них среди предполагаемых инсуляторов указывает на то, что лишь отдельные Alu-повторы обладают такой активностью.

Ретротранспозиция элементов, подобных Alu и LINE, таким образом, может быть одним из механизмов перестройки доменной структуры хроматина (Arnold et al., 2000). Можно сделать предположение, что одной из регуляторных функций ретроэлементов является их участие в образовании функциональных доменов.

Следует отметить, что исследованная нами область может содержать инсуляторы, которые не были идентифицированы из-за ограниченных возможностей использованного подхода. Идентифицируемые при данном подходе инсуляторы должны быть активны в клетках СНО и по отношению к паре «промотор-энхансер» цитомегаловируса. При этом некоторые ткане- или энхансер-специфичные инсуляторы могли быть пропущены. Подобные недостатки, однако, могут быть преодолены путем использования других клеточных линий и промоторно-энхансерных конструкций.

В целом, учитывая перечисленные слабые места, предложенный метод может использоваться как важный подход функциональной геномики.

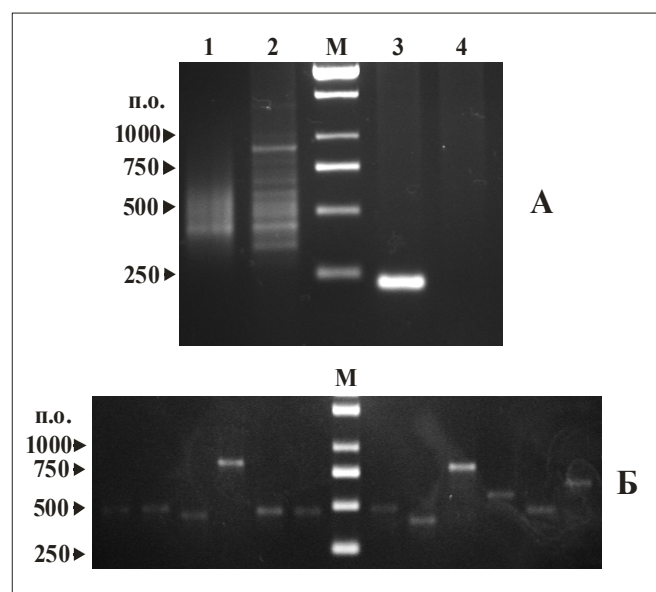


Рис. 12. (А) Результаты процедуры селекции. ПЦР амплификация с праймерами 1L и 1R (Б). Матрицами для ПЦР служили: дор. 1 - пул плазмид до трансфекции, дор. 2 - геномная ДНК клеток, трансфицированных библиотекой коротких фрагментов, клонированных в вектор рPNT/EmP после позитивной и негативной селекций, дор. 3 - геномная ДНК из клеток, трансфицированных исходной плазмидой рPNT/EmP (без селекции GANC), дор. 4 - геномная ДНК клеток, трансфицированных плазмидой рPNT/mP. (Б) ПЦР амплификация с праймерами 1L и 1R 12 случайных клонов из библиотеки инсуляторов. М - 1 т.п.о. ДНК маркер (Сибэнзим).

#### 4. Энхансер-блокирующая активность CTCF-связывающих последовательностей

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

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

Десять фрагментов, выявленных по связыванию с транскрипционным фактором CTCF были клонированы в обеих ориентациях в плазмиду pPNT/EmP (Акопов et al., 2006) между энхансером и минимальным промотором цитомегаловируса.

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

В качестве контроля функциональной активности мы приготовили плазмиду pPNT/E-sns-mP, содержащую между энхансером и минимальным промотором CMV инсулятор *sns* морского ежа *Paracentrotus lividus* (Di Simone et al., 2001; Melfi et al., 2000). В качестве негативного контроля мы приготовили плазмиду pPNT/E-λ-mP, содержащую между энхансером и минимальным промотором CMV фрагмент ДНК фага λ длиной ~200 п.о.

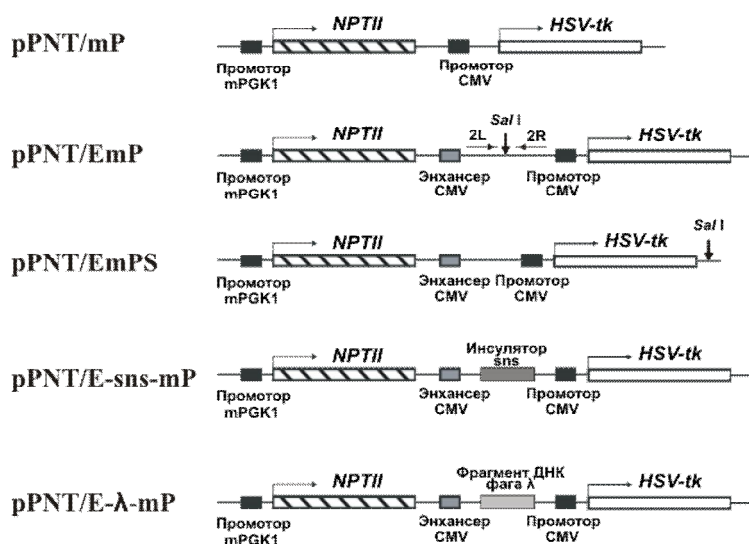


Рис. 13. Схемы плазмидных конструкций. mPGK-1 - промотор фосфоглицераткиназы-1 мыши; *NPTII* - ген неомин-фосфотрансферазы *E. coli*; *HSV-tk* - ген тимидинкиназы вируса простого герпеса.

Для проверки возможного наличия сайленсерной активности CTCF-связывающих фрагментов из плазмиды pPNT/EmP был удален сайт рестрикции *Sal I*, расположенный между энхансером и минимальным промотором CMV. Приготовленная плазмида была обозначена как pPNT/EmPS (рис. 13). В нее по сайту *Sal I* были клонированы три CTCF-связывающих фрагмента (3, 7 и 8). Схемы приготовленных конструкций приведены на рис. 13.

Для анализа десяти CTCF-связывающих фрагментов на способность блокировать энхансер в системе позитивно-негативной селекции, их клонировали в плазмиду pPNT/EmP между энхансером и минимальным промотором цитомегаловируса. Таким образом приготовили 20 плазмид, с одним CTCF-связывающим фрагментом в каждой, клонированным в прямой и обратной ориентациях. Анализ энхансер-блокирующей активности фрагментов проводили в обеих ориентациях, поскольку для отдельных CTCF-зависимых инсуляторов была показана зависимость их активности от ориентации (Hark et al., 2000).

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

Ранее отмечалось, что 5'HS4 элемент из  $\beta$ -глобинового локуса кур проявлял инсуляторные свойства в клетках линии эритролейкоза человека K562 (Chung et al., 1993), а инсулятор *sns* морского ежа *Paracentrotus lividus* блокировал действие энхансера на промотор в клетках остеосаркомы человека U2-OS и в клетках человеческой аденокарциномы легкого H1299 (Di Simone et al., 2001). Кроме того, показано, что человеческий CTCF связывается с соответствующими участками в геноме мыши (Filipponi et al., 1996). Поскольку инсуляторы организмов одного вида могут сохранять свою активность в клетках организмов других видов, вполне корректно использование клеток линии CHO для поиска инсуляторов среди фрагментов ДНК из локуса хромосомы 19 человека. После позитивно-негативной селекции, которую проводили как описано ранее, из прошедших отбор клеток выделили геномную ДНК.

Образцы геномной ДНК анализировали посредством двухстадийной ПЦР. На первой стадии на геномной ДНК в качестве матрицы с праймерами 2L и 2R амплифицировали фрагменты, находящиеся между энхансером и промотором CMV (рис. 14). После ПЦР получили смесь клонированных ранее CTCF-связывающих фрагментов, фланкированных небольшими фрагментами плазмиды pPNT/EmP. Эту смесь использовали в качестве матрицы для следующей стадии ПЦР.

К последовательностям каждого CTCF-связывающего фрагмента из локуса *FXYD5-COX7A1*, а также инсулятора морского ежа *sns* конструировали внутренние праймеры (табл. 4), которые в комбинации с плазмидными праймерами (2L или 2R) позволяли определить наличие каждой из вставок в смеси и их ориентацию (рис. 14).

**Таблица 4.** Пары праймеров для идентификации CTCF-связывающих последовательностей в прямой и обратной ориентации.

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

Из рис. 14 (верх) видно, что все 10 CTCF-связывающих фрагментов, а так же инсулятор *sns*, обнаруживаются в геномной ДНК клеток после селекции с генетицином G418, то есть содержащие их конструкции интегрировали в геном клеток. Те же фрагменты обнаруживаются и после селекции клеток в присутствии 10 мкМ ганцикловира в культуральной среде.

Аналогичные результаты получили при селекции с концентрацией селективного агента 4 мкМ (не показано). Таким образом, можно заключить, что все 10 исследованные нами CTCF-связывающих фрагментов, как и инсулятор *sns*, будучи клонированными между промотором и энхансером, придают клеткам устойчивость к ганцикловиру.

Устойчивость к ганцикловиру трансфицированным клеткам, помимо инсуляторов, могут придавать сайленсеры, т.е. последовательности, которые, в отличие от инсуляторов, способны подавлять активность промотора независимо от их положения относительно последнего (Ogbourne and Antalis, 1998). В таком случае, сайленсер способен инактивировать и промотор гена неомицинофосфотрансферазы, и клетки, содержащие такую конструкцию должны погибать на стадии позитивной селекции (Акоров et al., 2006). Однако для тестирования наличия потенциальной сайленсерной активности у CTCF-связывающих фрагментов три из них (3, 7 и 8) мы клонировали в плазмиду рPNT/EmPS с 3'-стороны от гена *HSV-tk* (Рис. 13). Клетки, трансфицированные этими плазмидами, погибали на стадии негативной селекции с

ганцикловиrom, т.е. CTCF-связывающие последовательности 3, 7 и 8 не оказывали ослабляющего действия на активность энхансера CMV и не являлись сайленсерами.

Таким образом, анализ в системе позитивно-негативной селекции фрагментов, картированных в локусе *FXVD5-COX7A1* и связывающихся с транскрипционным фактором CTCF показал, что все эти фрагменты обладали энхансер-блокирующими свойствами, причем эта их активность проявлялась в обеих ориентациях относительно промотора. Энхансер-блокирующая активность инсулятора морского ежа *sns* в системе позитивно-негативной селекции также не зависела от ориентации.

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

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

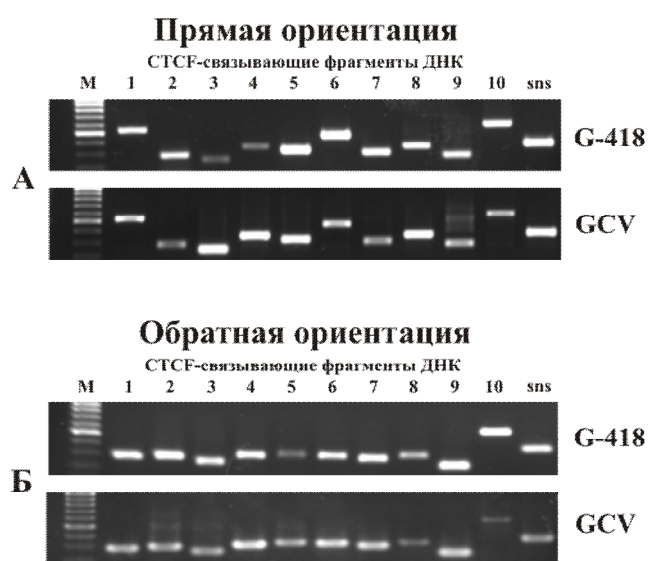


Рис. 14. ПЦР анализ на матрице геномной ДНК из трансфицированных клеток СНО после позитивной (G418) и негативной (GCV) селекции с праймерами, специфичными к 10 CTCF-связывающим последовательностям ДНК. (А) "Прямая" и (Б) "обратная" ориентации относительно промотора. М - маркер длин фрагментов ДНК.

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

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

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

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

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

Плазмидой рSG5/HisCTCF трансфицировали клетки линии COS-1, культивировали в течение 48 час, снимали раствором Версена, обрабатывали ультразвуком, и белок CTCF очищали методом никель-аффинной хроматографии. Качество очищенного белка проверяли при помощи вестерн-блоттинга с антителами к CTCF, а также при помощи EMSA, конкурентного EMSA и супершифта с известными фрагментами ДНК, способными связывать CTCF.

ДНК искусственной бактериальной хромосомы (BAC), содержащей исследуемую область генома кур (любезно предоставлена О.В. Яровой), расщепляли в двух параллельных образцах



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

11 случайно отобранных клонов библиотеки проверили на их способность связывать куриный CTCF (рис. 15). Оказалось, что все они связывают белок CTCF сравнимо с положительным контролем (CTCF-связывающим фрагментом ДНК из лизоцимного сайленсера кур).

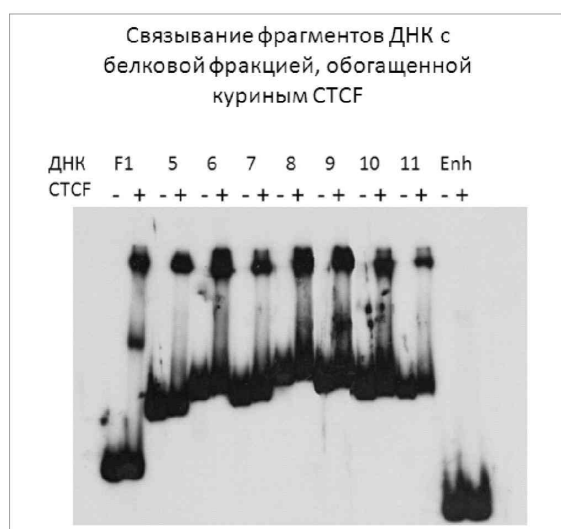


Рис. 15. Проверка эффективности связывания с белком CTCF фрагментов ДНК из библиотеки, полученной после селекции. F1 - CTCF-связывающий фрагмент ДНК из лизоцимного сайленсера кур (положительный контроль); Enh - фрагмент ДНК, содержащий энхансер бета-глобиновых генов кур (отрицательный контроль).

Далее было секвенировано 230 клонов библиотеки, среди которых найдено 77 уникальных последовательностей (рис. 16). Поскольку кривая рис. 16 практически достигает насыщения, полученная библиотека вполне представительна.

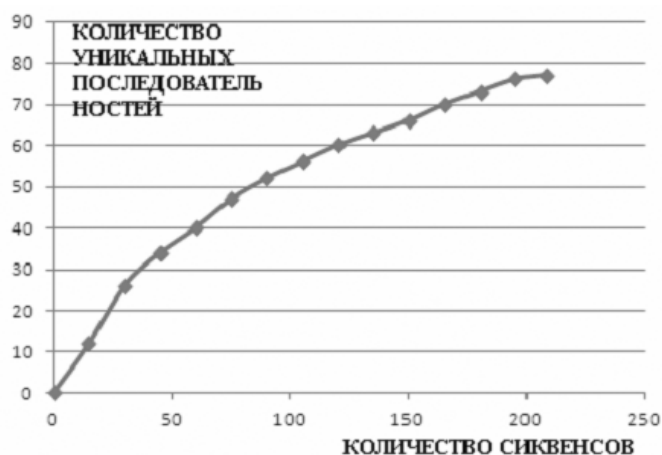


Рис. 16. Оценка числа независимых клонов в библиотеке. Показана зависимость числа уникальных последовательностей от полного количества секвенированных клонов.

Выявленные последовательности были нанесены на физическую карту локуса генома кур, содержащего альфа-глобиновые гены (рис. 17). На карту также были нанесены сайты связывания CTCF, идентифицированные различными методами в других работах (Martin et al., 2011; Klochkov et al., 2006; Valadez-Graham et al., 2004). Анализ распределения последовательностей показывает, что число сайтов генома, способных связывать CTCF *in vitro* по данным двумерного EMSA с применением очищенной ДНК значительно превышает число сайтов, связывающихся с CTCF в составе хроматина и идентифицируемых методом иммунопреципитации хроматина. Этот вывод следует из сравнения полученных нами данных с результатами работы (Martin et al., 2011), в которой для нашего участка генома было идентифицировано лишь 6 участков связывания CTCF в одном типе клеток, т.е. на порядок меньше, чем нами. Четыре из 6 участков перекрываются с идентифицированными в нашей работе. Причиной такого расхождения может быть недостаточно высокая чувствительность метода иммунопреципитации хроматина. При этом участки, относительно слабо связывающиеся с CTCF *in vivo*, не регистрируются, тогда как связывание со свободной ДНК выявляет все потенциально возможные участки связывания CTCF. Кроме того, степень сродства CTCF к одному и тому же участку связывания в составе свободной ДНК и в хроматине могут не совпадать, в частности, из-за положительной или отрицательной кооперативности связывания в присутствии других белков хроматина.

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

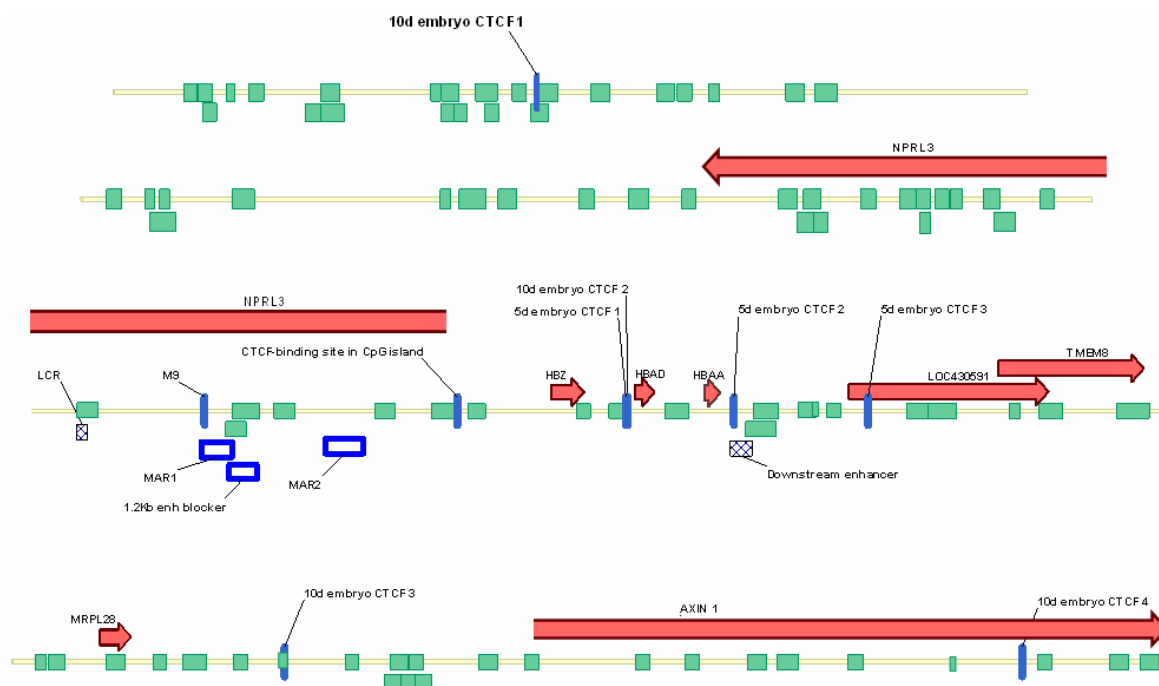


Рис. 17. Карта расположения потенциальных CTCF-связывающих последовательностей в области альфа-глобиновых генов кур. Красными стрелками обозначены расположение и направление транскрипции генов, зелеными прямоугольниками - расположение CTCF-связывающих последовательностей, идентифицированных методом двумерного EMSA. Синими вертикальными полосами обозначены участки, связывающие CTCF в клетках эмбрионов кур по данным иммунопреципитации хроматина (Martin et al., 2011; Klochkov et al., 2006; Valadez-Graham et al., 2004).

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

На рис. 18 представлены сравнительные результаты экспериментов по иммунопреципитации хроматина клеток кур HD3, индуцированных к эритроидной дифференцировке клеток HD3 и клеток DT40. Как видно из рисунка, идентифицированный нами фрагмент 12 связывает CTCF *in vivo* в клетках HD3 и индуцированных к дифференцировке клетках HD3 на уровне положительных контролей, зафиксировано также связывание CTCF фрагментами 4, 5 и 9 в индуцированных клетках HD3. Из идентифицированных ранее (Martin et al., 2011) CTCF-связывающих фрагментов лишь один (5d3) показал невысокий уровень связывания CTCF.

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

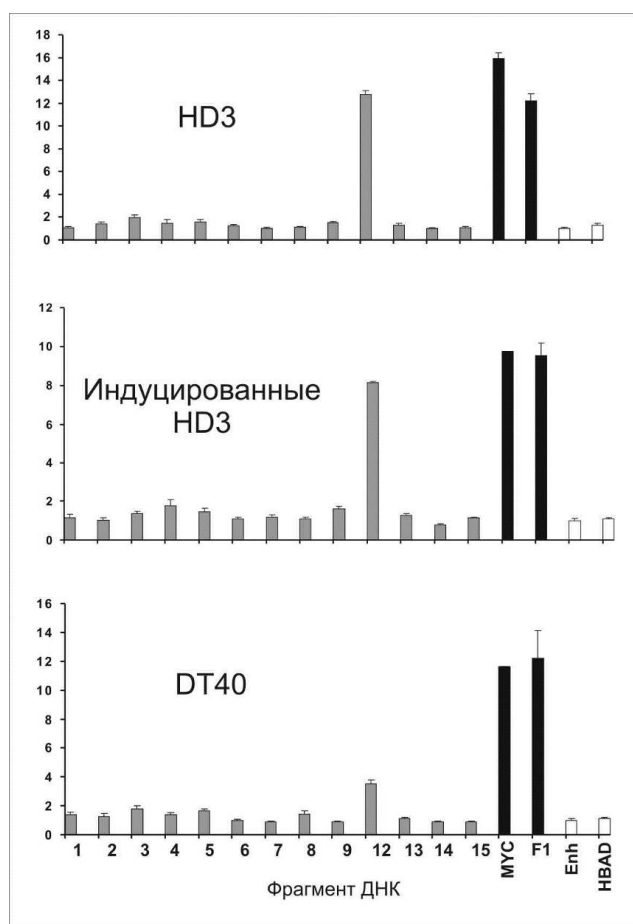


Рис. 18. Результаты экспериментов по иммунопреципитации хроматина с антителами к куриному CTCF для ряда фрагментов ДНК, идентифицированных нами в области генов альфа-глобинов кур (серые столбцы). Номера фрагментов соответствуют их положению на карте рис. 17. Черными столбцами обозначены положительные контроли (фрагменты промотора гена *тус* и инсулятора F1 лизоцимного гена кур), белыми – отрицательные контроли (фрагменты области энхансера бета-глобинового локуса кур и гена альфа-глобина D).

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

Одиночный LTR HERV-K(HML-2), расположенный на 19 хромосоме человека (номер по GenBank L47334; далее LTR L47334) использовали для определения тканеспецифичности его промоторной и энхансерной активностей. LTR L47334 отсутствует в соответствующем локусе генома шимпанзе и является представителем подгруппы HS-b “молодого” подсемейства семейства HERV-K(HML-2). Эта подгруппа предположительно образовалась около 10 млн. лет назад, а LTR L47334 – менее 5,6 млн. лет назад (Khil et al., 1997; Lebedev et al., 2000; Buzdin et al., 2003).

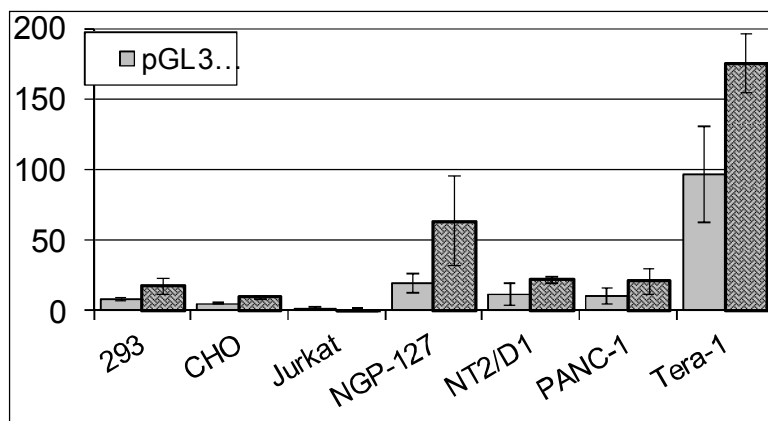
Ранее было показано, что этот LTR обладает промоторной и энхансерной активностью, а также в его U5-области расположен негативный регуляторный элемент (НРЭ) (Domansky, et al., 2000; Доманский и др., 2002).

Существенным казалось выснить, зависят ли указанные активности LTR от типа клеток, или, проявляют ли содержащиеся в нем регуляторные элементы тканеспецифичность. С этой целью мы провели серии транзientных трансфекций на десяти клеточных линиях различного происхождения. С этой целью использовали конструкции, приготовленные для исследования промоторной (pGL3BV) и энхансерной (pGL3PV) активности, в которые клонировали в

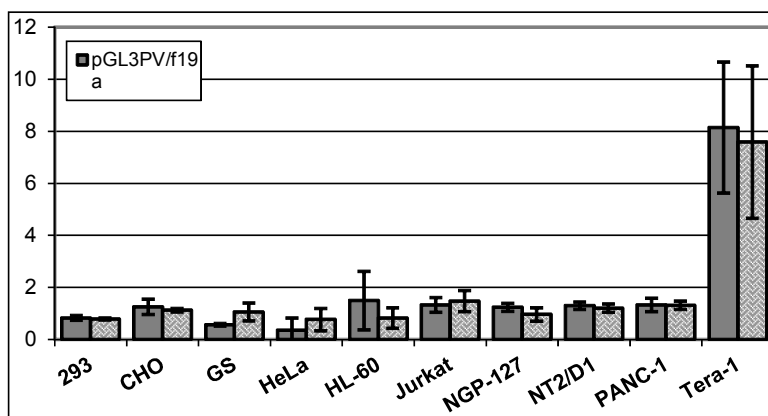
прямой ориентации полноразмерный LTR L47334 (f19a) или его фрагмент, не содержащий предполагаемого НРЭ (d19a).

Промоторную активность и влияние на нее НРЭ оценивали, сравнивая экспрессию репортерного гена (люциферазы) в клетках, трансфицированных плазмидой pGL3BV/f19a с экспрессией люциферазы в клетках, трансфицированных плазмидой pGL3BV/d19a или плазмидой pGL3PV, содержащей промотор вируса SV40 (рис.19). Полученные результаты трансфекции в исследованных линиях клеток: HEK-293, CHO, NGP-127, NT2/D1, PANC-1 и Tera-1 показали, что LTR L47334 проявляет относительно слабую промоторную активность во всех протестированных линиях клеток за исключением Tera-1, по сравнению с промотором SV40. Это можно объяснить действием НРЭ или тканеспецифичностью энхансера в составе LTR.

В отличие от регуляторных элементов LTR других ретровирусов, активность которых зависела от типа ткани и вида организма (Zhu et al., 2000; Wilson et al., 2003), исследуемый нами НРЭ не проявлял выраженной тканеспецифичности. Нами было показано, что (1) фрагмент LTR (d19a), не содержащий НРЭ, проявлял промоторную активность во всех линиях, где был активен полноразмерный L47334, (2) во всех исследованных линиях клеток относительное увеличение промоторной активности по сравнению с полноразмерным LTR практически одинаково (рис. 19). Поскольку большая часть одиночных LTR HERV-K(HML-2) в геноме человека представлена 5' и 3' делетированными фрагментами, эти результаты заслуживают особенного внимания (Lebedev et al., 1995; Vinogradova et al., 1997).



**Рис. 19.** Относительная промоторная активность HERV-K(HML-2) LTR L47334 (f19a) и его фрагмента (d19a) в различных клеточных линиях. Активность промотора вируса SV40 принята за 100.



**Рис. 20.** Относительная энхансерная активность HERV-K(HML-2) LTR L47334 (f19a) и его фрагмента (d19a) в различных клеточных линиях. Активность промотора вируса SV40 принята за единицу.

Таким образом, выявленные вариации в промоторной активности LTR L47334, очевидно, связаны не с влиянием НРЭ, а с активностью энхансера в составе LTR. На рис. 20 приведены результаты, свидетельствующие о наличии выраженной энхансерной активности у полноразмерного и частично делетированного LTR в клеточной линии Tera-1 (уровень экспрессии репортерного гена повышен в восемь раз).

Подобная тканеспецифичность может быть обусловлена взаимодействием LTR с белками, обнаруженными нами в ядерном экстракте клеток линии Tera-1. Ни в одной другой из использованных линий клеток полноразмерный или лишенный НРЭ LTR не проявляли энхансерной активности. Поскольку ранее мы показали, что присутствие НРЭ не снижает энхансерной активности LTR в Tera-1 и вместе с тем оказывает влияние на экспрессию репортерного гена, когда не входит в состав транскрипта (Domanskii et al., 2002), можно выдвинуть предположение, что он влияет на способность промотора собирать транскрипционный комплекс и/или на элонгацию транскриптов.

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

Фрагменты LTR для оценки их регуляторного потенциала одного из представителей семейства HERV-K были проверены нами на присутствие сайтов связывания ядерных факторов клетки-хозяина с использованием сдвига электрофоретической подвижности и ультрафиолетовой сшивки (Bogomolova et al., 1998; Nikolaev et al., 1996). Комплексы ДНК-белок были разделены методом торможения в геле (как на рис. 21Б), соответствующую полосу вырезали из геля, облучали УФ-светом и сшитые комплексы белка с мечеными олигонуклеотидами разделяли во втором направлении в соответствии с их молекулярными массами. Неспецифический комплекс b (рис. 21Б), содержал одиночный белок (b на рис. 21С) с кажущейся молекулярной массой 56 кДа, в то время как главная полоса торможения содержала три белка с молекулярными массами 98, 91 и 81 кДа.

Все белковые пятна были расположены строго на одной вертикальной линии, что указывает на то, что они были частью одного комплекса ДНК-белок. Мы обозначили белки как ERLBF1, ERLBF2 и ERLBF3 (Endogenous Retrovirus LTR Binding Factors) в порядке убывания их молекулярных масс. Для дальнейшей характеристики сайта связывания белка был проведен эксперимент по защите от ДНК-азы I или футпринт-анализ (рис. 22). Защищенная последовательность длиной 28 п.о. расположена в 5'-части U3 области, фланкирована сайтами гиперчувствительности к ДНК-азе I и участками, ранее идентифицированными как потенциальный энхансер и сайт связывания глюкокортикоидного рецептора (Ono et al., 1986; Leib-Mosch et al., 1993). Наиболее эффективно защищена последовательность, расположенная непосредственно с 5'-стороны за этими двумя сайтами.

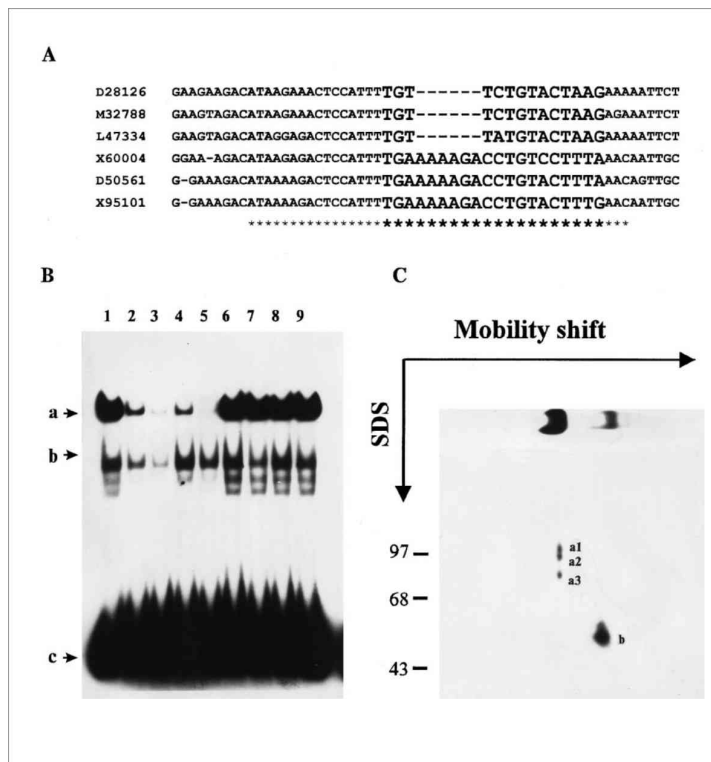


Рис. 21. Идентификация белков связывающих LTR. А. Выравнивание последовательностей LTR в области связывания с белками (отмечены звездочками). GRE и коровые энхансерные элементы, ранее определенные по гомологии последовательностей увеличены. Левая колонка - регистрационные номера из GenBank соответствующих последовательностей. Полноразмерные LTR HERV-K были определены в GenBank на основе гомологии с консенсусными последовательностями. Б. Анализ сдвига электрофоретической подвижности и конкуренции с использованием меченого двухцепочечного олигонуклеотида LTR1/LTR1-C в качестве матрицы и

ядерного экстракта из клеток линии Jurkat. а, б - основные комплексы белок-ДНК, с - свободный меченый олигонуклеотидный зонд. Дорожка 1, без добавления конкурента. Дорожки 2-9, сдвиг электрофоретической подвижности с добавлением 10 - и 20-кратного молярного избытка различных немеченых двухцепочечных олигонуклеотидов-конкурентов: дорожки 2,3 - LTR1/LTR1-C; дорожки 4,5 - LTR2/LTR2-C; полосы 6,7 - HIV/HIV-C; полосы 8,9 - HIVM/HIVM-C. С. Двумерный электрофорез ДНК-белковых комплексов LTR HERV-K: первое направление, неденатурирующий 10% полиакриламидный гель; второе направление, 12% SDS полиакриламидный гель после УФ-сшивки комплексов. а1 (ERLBF1), а2 (ERLBF2), а3 (ERLBF3) и б, белки, которые связываются с LTR.

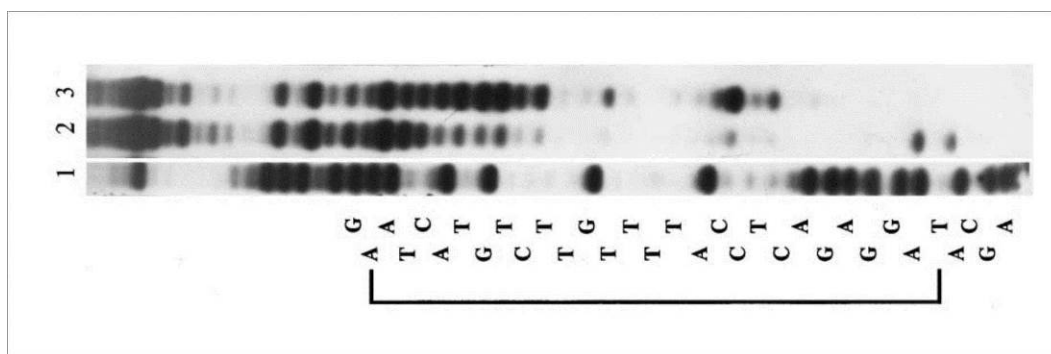


Рис. 22. Защита фрагмента LTR-F/590-R HERV-K LTR от действия ДНКазы I . Полоса 1, A + G последовательность (маркер). Дорожка 2, добавлено 50 мкг очищенного белка ядерного экстракта. Дорожка 3, белок не добавлялся.

Для оценки общей консервативности и стабильности этого региона в эволюции было сделано выравнивание полноразмерных последовательностей HERV-K LTR из GenBank. Область связывания белков показана на рис. 21А. Видно, что участок длиной 6 п.о. делетирован у трех LTR молодого подсемейства, но сохраняется у трех LTR старого подсемейства. Как

видно из рис. 22, эта делеция лежит в самом центре участка, защищенного белком (белками). Для изучения роли этой делеции в связывании белка были приготовлены два двунитевых олигонуклеотида, представляющие консенсусную последовательность каждой из двух групп и использованы для анализа методом сдвига электрофоретической подвижности (рис. 21Б). Дорожка 1 показывает результат сдвига электрофоретической подвижности при использовании в качестве зонда двухцепочечного олигонуклеотида LTR1/LTR1-С и ядерного экстракта из клеток Jurkat. Этот олигонуклеотид имеет консенсусную последовательность молодого подсемейства LTR, содержащего AAAAGA (рис. 21А). Помимо двух основных видимых полос торможения а и б (рис. 21Б), детектируются также несколько слабых полос. На других дорожках показаны результаты подобного эксперимента в присутствии 10- и 20-кратного молярного избытка различных немеченых конкурентов, которые либо содержали (олигонуклеотид LTR1/LTR1-С, идентичный зонду) или не содержали (олигонуклеотид LTR2/LTR2-С) сайта, делетированного в LTR молодого подсемейства (полосы 2, 3 и 3, 4, соответственно). Дорожки 6, 7, и 8, 9 содержат в качестве конкурентов олигонуклеотиды, представляющие соответственно нативную и мутированную энхансерные области вируса иммунодефицита человека типа 1 (Nikolaev et al., 1996). Видно, что белок (белки) главной полосы торможения (а) легко вытесняется из комплекса с меченым олигонуклеотидом относительно небольшим молярным избытком обоих олигонуклеотидных конкурентов независимо от наличия б п.о. делеции. В то же время последовательность лентивирусного энхансера не конкурирует за связывание белковых факторов, связывающихся с последовательностью LTR. Интенсивность полосы (b) изменялась незначительно при добавлении как олигонуклеотидов HERV-K, так и ВИЧ-1, указывая низкую специфичность связывания белка в этом комплексе. Таким образом, только белок (белки), отвечающие за появление полосы торможения (а) можно считать специфически связывающимися с LTR HERV-K. Несмотря на существенные различия между первичными структурами участка связывания, LTR обоих подсемейств HERV-K способны к образованию специфического комплекса с тем же самым белком (белками) ядерного экстракта клеток Jurkat. Таким образом, белки, которые образуют комплексы а и б с последовательностью HERV-K LTR, присутствуют в различных типах клеток человека и млекопитающих, сшиваются с ДНК УФ-светом, т.е. они находятся в тесном контакте с ней. Это объясняет довольно большую протяженность (около 30 п.о.) области, защищенной этими белками от расщепления ДНК-азой I.

Далее мы локализовали участок связывания этих факторов с большей точностью. С этой целью был синтезирован набор двунитевых олигонуклеотидов, перекрывающих различные участки внутри последовательности LTR, и их способность связывать комплекс ERLBF была проверена методом EMSA с ядерными экстрактами из различных линий клеток (рис. 23). Как



видно из рисунка, короткий олигонуклеотид длиной 21 п.о. (рис. 23В) способен связываться специфически с белками, но вместо одной полосы торможения ДНК-белкового комплекса наблюдаются две.

Для определения белкового состава комплексов с коротким (21 п.о.) олигонуклеотидом, мы проводили двумерный анализ EMSA с промежуточным сшиванием ДНК с белками при помощи ультрафиолетового излучения. Из рисунка (рис. 24Б) видно, что один из комплексов с олигонуклеотидом длиной 21 п.о. включает белковые факторы ERLBF1, ERLBF2 и ERLBF3, а другой (с большей электрофоретической подвижностью) – содержит только факторы ERLBF1 и ERLBF2. Вместе с тем, общий белковый состав комплекса с олигонуклеотидом длиной 21 п.о. не отличается от такового для комплекса, образованного с более длинным олигонуклеотидом (рис. 24Б).

Следует отметить, что состав комплексов области U3 LTR HERV-K сильно зависит от состояния клеток. Известно, что тепловой шок (Moggs and Orphanides, 2003) и митогены, такие как фитогемагглютинин и форболовые эфиры (Gurskaya et al., 1996), оказывают большое влияние на профиль экспрессии генов в клетке. Митогены, например, индуцируют продукцию эндогенных ретровирусов (Tacke et al., 2003). Мы исследовали влияние теплового шока и индукции клеток Jurkat митогенами на состав белкового комплекса, связывающегося с 5'-частью области U3 LTR HERV-K (рис. 24В, Г). Из рисунка видно, что в условиях теплового шока и при действии митогенов ДНК-белковый комплекс утрачивает фактор ERLBF3, тогда как факторы ERLBF1 и ERLBF2 остаются связанными с ДНК.

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

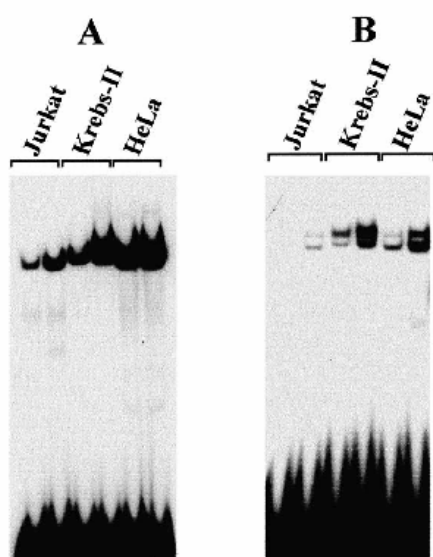


Рис. 23. Электрофореграмма сдвига подвижности (EMSA) ДНК-белковых комплексов, сформированных при добавлении 5 и 10 мкг белка ядерных экстрактов из клеточных линий Jurkat, Krebs-II и HeLa. В качестве зонда использовали двунитевые олигонуклеотиды длиной (А) 30 п.о. и (В) 21 п.о., содержащие участок связывания ERLBF.

### 6.2. Связывание белков с негативным регуляторным элементом (НРЭ)

Выше было показано, что частичная делеция области U5 приводила к повышению промоторной активности LTR независимо от его ориентации. Подобные негативные регуляторные элементы были описаны для пенистого вируса (foamy virus) человека (Seiki et al., 1990), и вируса Т-клеточного лейкоза 1 (HTLV-1) (Seiki et al., 1990, Okumura et al., 1996).

Для проверки способности содержащего НРЭ фрагмента специфически связывать клеточные белки, фрагмент LTR L47334 длиной 136 п.о. радиоактивно метили ПЦР-амплификацией и использовали для EMSA с ядерными экстрактами из нескольких клеточных линий (рис. 25). Видно, что белки, способные связываться с НРЭ-содержащим фрагментом, обнаруживаются во всех проверенных клеточных линиях. В экспериментах с конкурентным связыванием немеченого фрагмента, содержащего НРЭ, была подтверждена специфичность взаимодействия (не показано).

Количество связавшихся белков было различным для ядерных экстрактов из различных клеток, о чем свидетельствует различная интенсивность соответствующих ДНК-белковым комплексам полос. При этом интенсивность полос связывания, соответствующих

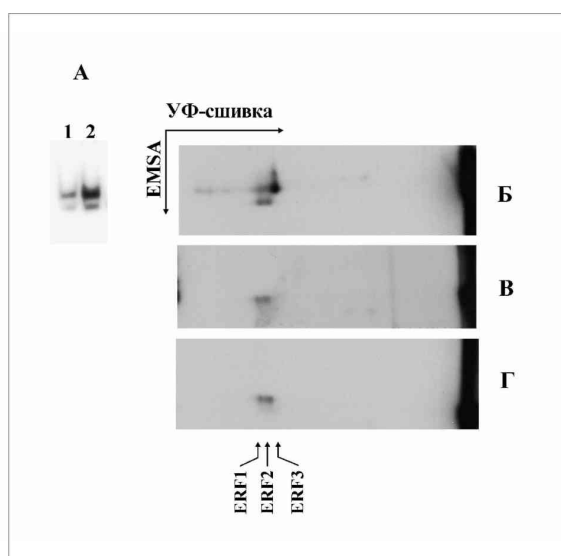


Рис. 24. Результаты эксперимента по EMSA с использованием олигонуклеотида длиной 21 п.о. (А), и белковый состав комплексов (Б, В, Г), полученный по результатам двумерного EMSA с ультрафиолетовой сшивкой.

комплексам белковых факторов ERLBF1, 2 и 3 с 5'-участком области U3 LTR, была практически одинаковой при использовании разных ядерных экстрактов (рис. 25).

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

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

Ранее группу факторов ERLBF мы выявляли в клетках китайского хомячка, и, так как системы регуляции транскрипции у человека и грызунов весьма похожи,

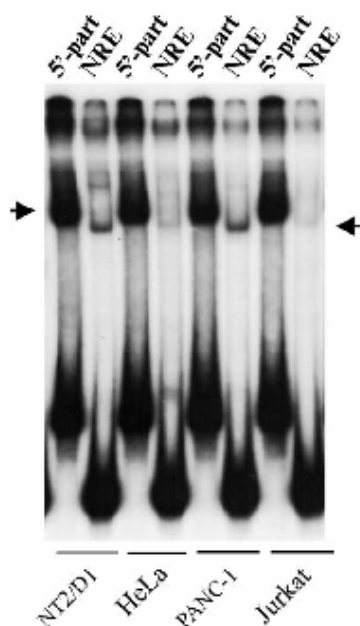


Рис. 25. EMSA с использованием в качестве зонда 5'-участка области U3 LTR L47334 и участка области U5, содержащего негативный регуляторный элемент. Под соответствующими дорожками приведены клеточные линии, использованные для приготовления ядерных экстрактов. Стрелками обозначены два основных ДНК-белковых комплекса.

мы предположили, что искомые белки могут содержаться и в опухолевых клетках грызунов. Поэтому в качестве источника ядерных белков далее использовали клетки перевиваемой асцитной карциномы мышей Krebs-II, из которых выделяли ядерный экстракт. Из рис. 26 видно, что в ядерном экстракте из асцитной опухоли мышей присутствует интересующий нас специфический белковый комплекс в количествах, сравнимых с содержанием его в больших объемах суспензионных культур клеток (например, Jurkat).

Основным методом для выделения и очистки ДНК-связывающих белков из ядерных экстрактов клеток является метод ДНК-аффинной хроматографии [Kadonaga and Tjian , 1986].

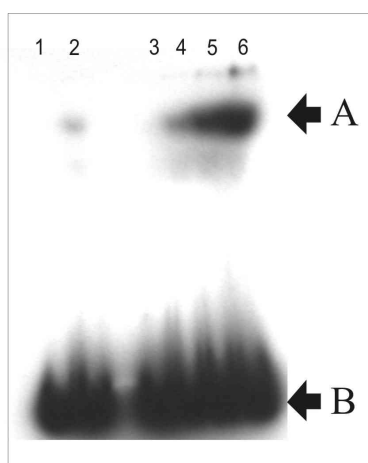


Рис. 26. Сравнение содержания комплексов олигонуклеотида с белками ERLBF ядерных экстрактов, выделенных из клеток Jurkat и асцитной опухоли мышей Krebs-II. Дорожки 1,2 - ядерный экстракт из клеток Jurkat. Дорожки 3-6 - ядерный экстракт из клеток асцитной эпителиальной карциномы мышей. А - ДНК-белковый комплекс, В - свободный олигонуклеотид.

Используется аффинная колонка, в которой к сефарозе пришиты предварительно лигированные олигонуклеотиды, содержащие сайт связывания определенного белка (Arndt-Jovin et al., 1975). В процессе элюции слабо связывающихся белков происходят большие (до 80%) потери специфических белков (таблица 5), их окончательный выход снижается. В целом, метод достаточно трудоёмок, так как требует приготовления специального аффинного носителя для конкретного выделяемого белка (комплекса белков). К тому же, большие потери специфических белков в процессе их выделения требуют значительного увеличения объема исходного материала и возрастания расхода сопутствующих материалов на всех этапах получения белков в количествах, необходимых для их надежной идентификации.

Для оптимизации процедуры выделения и выхода факторов ERLBF мы разработали альтернативный метод очистки ДНК-связывающих белков с использованием иммобилизованного на агарозе гепарина. Повторяющееся олигосахаридное звено гепарина можно рассматривать как аналог фосфатного остова ДНК, поэтому большинство белков, специфически связывающихся с ДНК, имеют сродство к гепарину (Genersch et al., 1995; Cheng et al., 1997; Gadgil et al., 1999). В то же время, сродство этих белков к ДНК, содержащей последовательность их специфических сайтов связывания, гораздо выше, чем к гепарину, что позволяет быстро и эффективно выделять их при помощи элюции низкосолевым буфером, содержащим такую ДНК (рис. 27). На первом этапе белки ядерного экстракта из клеток асцитной карциномы мышей Krebs-II наносили на колонку с гепарин-агарозным носителем, как для первого этапа ДНК-аффинной хроматографии. Затем колонку промывали буфером, содержащим 0,2 М КСl, т.е., в концентрации, при которой специфические белки с колонки не элюируются. Далее белковые факторы ERLBF элюировали тем же буфером с добавлением меченого <sup>32</sup>P двунитевого синтетического олигонуклеотида, содержащего последовательность, соответствующую участку связывания ERLBF. Элюированные с колонки фракции анализировали методом EMSA. Олигонуклеотид сорбирует белки из комплекса с гепарином, в результате чего на выходе с колонки выявляются высокоаффинные комплексы белков с олигонуклеотидом, совпадающие по электрофоретической подвижности с комплексами, получаемыми при использовании исходного ядерного экстракта. Разработанный нами метод очистки ДНК-связывающих белков дает возможность получать эти белки в виде комплексов с олигонуклеотидом.

**Таблица 5. Очистка ДНК-связывающих белков на гепарин-агарозе с помощью аффинной элюции олигонуклеотидом.**

Фракция	Белок (мкг)	Белок (%)	Связывающая активность (%)	Очистка (кратность)
Исходный ядерный	3200	100	100	1

экстракт				
Белки, элюированные с гепарин-агарозы	270	8,5	92	11
Белки с ДНК-афинной колонки	0,16	0,005	10	2000
Белки, элюированные с гепарин-агарозы двуцепочечным олигонуклеотидом	1,8	0,06	80	1400

Для оптимизации процесса очистки факторов ERLBF были проведены следующие эксперименты:

1. Определение ёмкости гепарин-агарозной колонки. В этой серии экспериментов колонку с гепарин-агарозой (0,5 мл) последовательно нагружали возрастающими количествами ядерного экстракта. Насыщение гепарин-агарозы факторами ERLBF происходило при нанесении на колонку 6,4 мг белка ядерного экстракта. Во всех экспериментах этой серии на колонку наносили не более 4 мг исходного белка.
2. Анализ полноты элюции факторов ERLBF с колонки двунитевым специфическим олигонуклеотидом. После элюции белков посредством олигонуклеотида, колонку промывали 0,5 М KCl. Количество белков, элюируемых дополнительно высокосолевым буфером, незначительно.

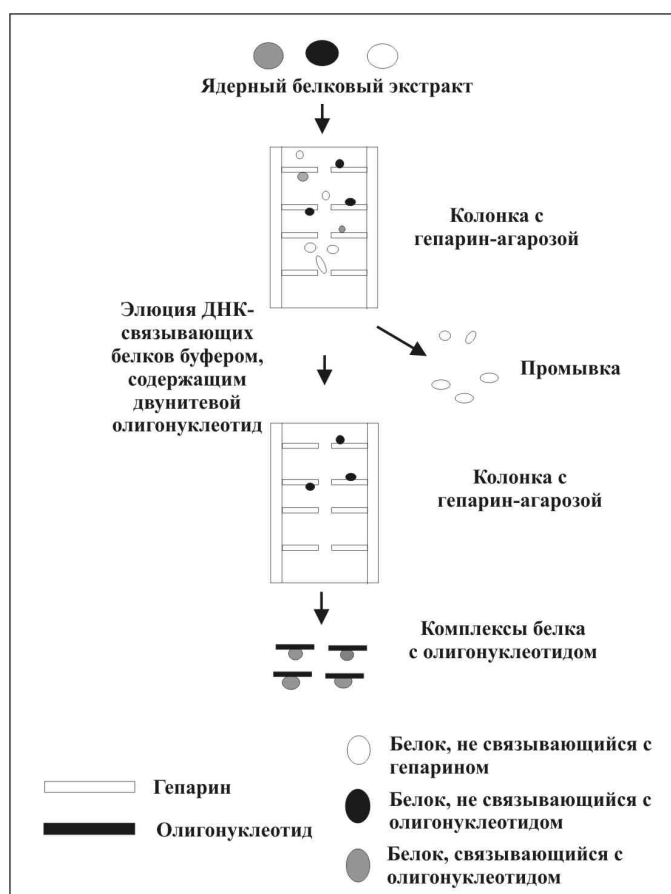


Рис.27. Схема очистки ДНК-связывающих белков методом аффинной элюции.

Выделение факторов ERLBF из 1 мл ядерного экстракта (концентрация тотального белка 2,5 мг/мл) проводили после оптимизации условий с использованием не содержащего радиоактивной метки олигонуклеотида. Элюат концентрировали, переосажда трихлоруксусной кислотой, и анализировали электрофорезом в денатурирующих условиях. На рис. 28 видны три мажорных белка с молекулярными массами от 60 до 70 кДа, не отличающиеся от полученных методом аффинной хроматографии. В таблице 5 приведены результаты очистки факторов ERLBF методом аффинной элюции в сравнении с методом ДНК-аффинной хроматографии. Анализ представленных в таблице результатов позволяет сделать вывод, что использование примененного нами метода позволяет получить из ядерного экстракта с выходом примерно на порядок большим, чем при ДНК-аффинной хроматографии, в одну стадию высокоочищенные ДНК-связывающие белки при кратности их очистки в 1400 раз.

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

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

Белки ERLBF1, 2 и 3, выделенные после аффинной элюции с гепарин-агарозы двунитевым олигонуклеотидом и затем разделенные с помощью ЭФ в ПААГ (рис. 28), анализировали методом MALDI-TOF-MS (Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry) (Shevchenko et al., 1996). Каждый из белков ERLBF(1-3) гидролизвали

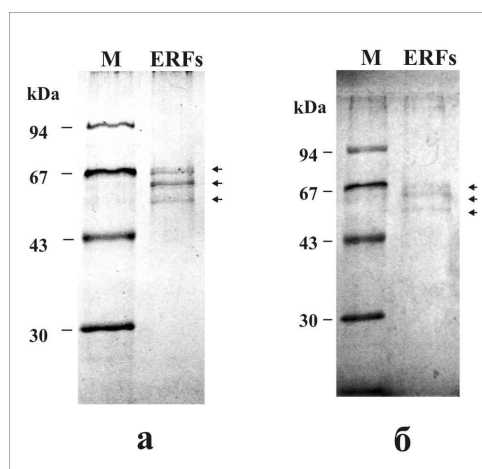


Рис. 28. Электрофорез в полиакриламидном геле белков ERLBF, выделенных двумя способами. (а) белковые факторы ERLBF, полученные при помощи аффинной элюции. Окраска Brilliant Blue R. (б) белковые факторы ERLBF, полученные методом ДНК-аффинной хроматографии. Окраска серебряным реагентом. М - маркер молекулярных масс.

трипсином и осуществляли масс-спектрометрический анализ полученных пептидов. Молекулярные массы выявленных пептидов сравнивали с молекулярными массами пептидов, входящих в базу данных MASCOT с последующей идентификацией наиболее вероятных белков, содержащих выявленные пептиды. По результатам анализа ERLBF-1 (Score 115) соответствовал конститутивному белку теплового шока Hsc70 (Heat shock 70 kDa protein 8, isoform 1, Homo sapiens) (рис. 29, таблица 6). MALDI-TOF-MS-анализ белков ERLBF(1-3) проводили 6 раз, при этом для выделения белков использовали ядерный экстракт, полученный в разное время. В итоге определили значительную часть аминокислотной последовательности ERLBF-1, которая на 66% совпадает со структурой белка теплового шока Hsc70 (рис. 30).

MASCOT-анализ пептидов, полученных в результате триптического гидролиза ERLBF-2 и ERLBF-3, не выявил известных белков, включая гипотетические. Вероятно, ERLBF2 и ERLBF3 могут быть новыми белками, поскольку в настоящее время отсутствуют данные о непосредственном взаимодействии белков теплового шока, в частности Hsc70, с нуклеиновыми кислотами, возможно, что эту функцию в составе комплекса ERLBF(1-3)-ДНК выполняет ERLBF-2.

Дальнейшую идентификацию ERLBF проводили при помощи иммунологического анализа. Поликлональные антитела (IgG фракция) к рекомбинантному Hsc70 (любезно предоставлен д-ром M. Ladjimi, Universite Pierre & Marie Curie, France) были получены Е.В. Свирщевской (ИБХ РАН). EMSA анализ проводили по стандартной методике, но к белку в составе реакционной смеси EMSA добавляли антитела. При специфическом взаимодействии белка с антителами, связывающимися с меченым фрагментом, формируется комплекс ДНК с белком и антителом, обладающий меньшей электрофоретической подвижностью, чем исходный ДНК-белковый комплекс. Дополнительный сдвиг электрофоретической подвижности (supershift) свидетельствует о том, что добавленные антитела специфически связываются с белком, входящим в комплекс с ДНК. В присутствии антител (рис. 31) образуется комплекс ДНК-белок-антитело, который имеет меньшую электрофоретическую подвижность, чем комплекс ДНК-белок.

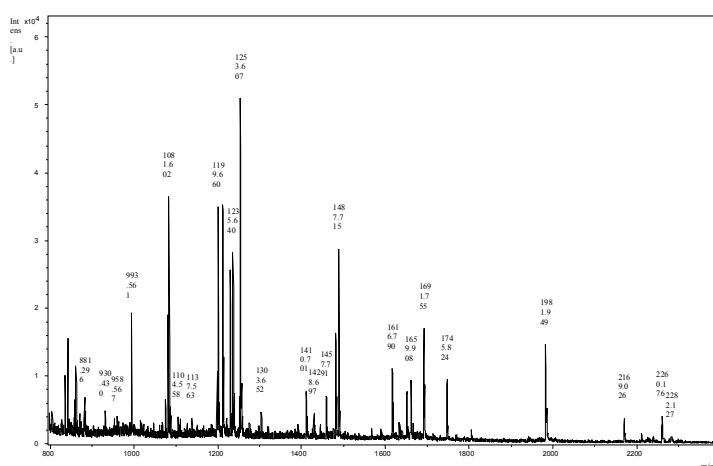


Рис. 29. Масс-спектр триптического гидролизата белкового фактора ERLBF-1.

mskgpavgidlgttyscvgvfqhgkveiiandqgnrtt<sup>s</sup>psyvaftdterligdaaknqvamnptntvfdakrlig  
rrfdдавvqsdmkhwpmvvdagrpkvqv<sup>e</sup>ykgetksfypeevssmvltkmkeiaeylgktvtnavvtv  
payfndsqrqatkdagtiaglnvriineptaaaiaygl<sup>d</sup>kkvgaernvlifldgggtdvsiltiedgifevkst<sup>a</sup>gd  
thlggedfdnrmvnhfiaefkrkhk<sup>k</sup>disenkravrrl<sup>t</sup>acerakrtlsstqasieidslvegidfyt<sup>s</sup>itarfeel  
nadlfrgtdpvekalrdakldksqihdivlv<sup>g</sup>gstripkiqkllq<sup>d</sup>ffngkelnksinpdeavaygaavqaails<sup>g</sup>  
dksenvqdlldvtplslgietaggvmtvlikr<sup>n</sup>ttiptkqtq<sup>t</sup>fttysdnqpgvliqv<sup>y</sup>egeramtkdnnllgkfel  
tgippaprgvpqievtfdidangilnvsavdkstgkenkitit<sup>nd</sup>kg<sup>r</sup>lskediermvqeaekykaedekqrdkv  
sskslesyafnmkatvedeklqgk<sup>i</sup>ndedkqkildk<sup>ne</sup>iiinwldknqtae<sup>ke</sup>efehqq<sup>ke</sup>lekv<sup>cn</sup>piitkly  
qsaggmpgg mp<sup>g</sup>gfp<sup>g</sup>ggappsggassgptievd

Рис. 30. Аминокислотная последовательность белка Hsc70 (AAH16660, Heat shock 70kDa protein 8, isoform 1, Homo sapiens). Идентифицированные триптические пептиды подчеркнуты.

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

**Таблица 6. Пептиды, совпадающие с картой конститутивного белка теплового шока массой 70 кДа (NM\_031165, heat shock 70kD protein 8; heat shock protein cognate 70, Mus musculus).**

Mr (набл.)	Mr (ожд.)	Пептид
1197.68	1196.67	FELTGIPPAPR
1199.68	1198.68	DAGTIAGLNVL
1228.64	1227.63	VEIANDQGNR
1251.62	1250.61	MVNHFIAEFK
1253.62	1252.61	FEELNADLFR
1426.64	1425.63	RFDDAVVQSDMK
1481.76	1480.75	SQIHDIVLVGGSTR
1487.67	1486.66	TTPSYVAFTDTER
1665.75	1664.74	NQVAMNPTNTVFDK
1669.78	1668.77	HWPFMVVNDAGRPK
1691.68	1690.67	STAGDTHLGGEDFNR
1981.97	1980.96	TVTNAVVTVPAYFNSQR
1457.71	1456.70	ELEKVCNPIITK
1616.74	1615.73	SFYPEEVSSMVLTK
1659.84	1658.83	IINEPTAAAIAYGLDK
1745.79	1744.78	NQTAEKEEFEHQQK
2260.26	2259.26	SINPDEAVAYGAAVQAAILSGDK
2774.64	2773.63	QTQTFTTYSNQPGLIQVYGER

Для подтверждения полученных выше результатов мы исследовали возможность взаимодействия белков фракции ERLBF с моноклональными антителами к Hsc70 (любезно предоставлены Б.А.Маргулисом, Институт Цитологии РАН, С-Петербург, Россия). Белки фракции ERLBF разделяли электрофорезом в ПААГ (рис. 32а), переносили их на PVDF-мембрану и инкубировали с антителами к Hsc70 (рис. 32б). На дорожке 3 в элюате белков



ядерного экстракта Krebs-II с гепарин-агарозной колонки, наблюдается отчетливый сигнал комплекса на уровне контрольного сигнала на дорожке 2, на которую нанесен маркер из трех белков: Hsc (70 kDa), BSA (67 kDa) и GDG (60 kDa). Сигнал соответствует белку Hsc70. Во фракции белков тотального ядерного экстракта Krebs-II (дорожки 4,5) сигнал комплекса обнаруживается в той же зоне. Таким образом, во фракции белков элюата ядерного экстракта с гепарин-агарозы в зоне подвижности ERLBF1 выявлен белок, гомологичный Hsc70 и образующий комплекс с его антителами. Данные western-блот анализа с моноклональными антителами к Hsc70 подтверждают результаты EMSA ERLBF и идентификации ERLBF1 методом MALDI-TOF-MS. Следовательно, в составе комплекса белков ERLBF1-3 с ДНК обнаруживается белок, который по триптической пептидной карте и антигенным детерминантам идентифицирован как конститутивный белок теплового шока Hsc70.

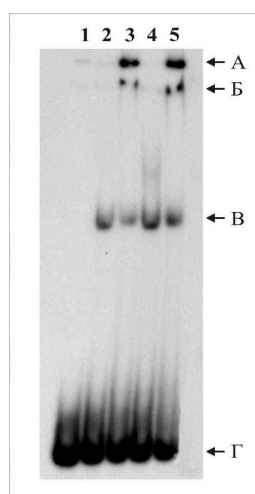


Рис. 31. Результат эксперимента по дополнительному сдвигу электрофоретической подвижности в присутствии антител. А, Б, - зоны сдвига комплексов белок-ДНК-антитело; В - зона сдвига комплекса белок-ДНК; Г свободный фрагмент ДНК. Дор. 1 - без добавления ядерного экстракта; 2 1 мкг ядерного экстракта из клеточной линии Jurkat; 3 - 1мкг экстракта Jurkat + 10 мкл анти-Hsc70 (15 мкг); 4 1 мкг ядерного экстракта из клеточной линии HeLa; 5 - 1мкг экстракта HeLa + 10 мкл анти-Hsc70 (15 мкг).

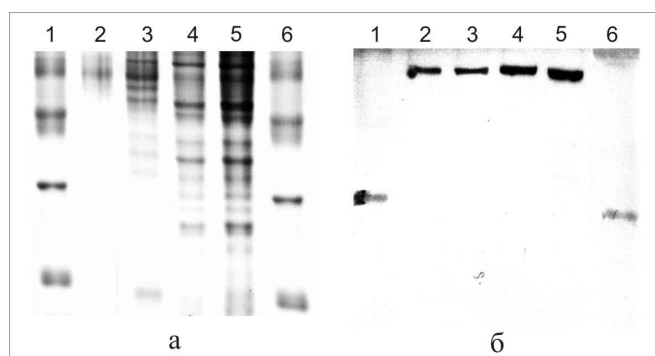


Рис. 32. Результат Western-блот анализа ERLBF с антителами к Hsc70. а - электрофореграмма белков фракции ERLBF в ПААГ; б - PVDF-мембрана с иммобилизованными белками, полученная как реплика с ПААГ (идентичного ПААГ на рис. 32а) и обработанная моноклональными антителами к Hsc70. Вторичные антитела конъюгированы с пероксидазой, в качестве субстрата использовали 4-хлор-1-нафтол. 1,6 -

белковые маркеры BioRad. 2 - смесь белков Hsc70 (70 кДа), BSA (67 кДа) и GDG (60 кДа). 3 - ERLBF, элюат с гепарин-агарозы. 4,5 - ядерный экстракт Krebs II (соотв., 20 и 40 мкг).

## 5. ОСНОВНЫЕ ВЫВОДЫ

1. Разработана стратегия идентификации и картирования потенциальных энхансеров в протяженных областях генома. С ее помощью идентифицировано и картировано 15 потенциальных энхансеров в локусе хромосомы 19 человека между генами *FXVD5* и *COX7A1*. Показано, что потенциальные энхансеры специфически взаимодействуют с ядерными белками.

13 из 15 обнаруженных последовательностей проявляют энхансерную активность в системе экспрессии репортерного гена. Таким образом, с помощью разработанной стратегии удается обнаруживать свыше 80% потенциальных энхансерных элементов в протяженных областях генома.

2. С помощью ранее разработанного нами метода позитивно-негативной селекции энхансер-блокирующих элементов генома идентифицировано и картировано 10 новых потенциальных инсультаторов в локусе хромосомы 19 человека между генами *FXYD5* и *COX7A1*. Проведен анализ энхансер-блокирующей активности обнаруженных нами ранее методом двумерного EMSA 10 CTCF-связывающих фрагментов ДНК. Все исследованные CTCF-связывающие последовательности проявляют энхансер-блокирующую активность в данной системе. Возможно, это является общим свойством CTCF-связывающих фрагментов ДНК.

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

4. На примере LTR семейства HERV-K (HML-2) исследован потенциальный спектр регуляторных элементов, принадлежащих рассеянным по геному мобильным элементам. Показано наличие промоторной активности LTR эндогенных ретровирусов человека семейства HERV-K в различных клеточных линиях человека и грызунов и выявлена ее тканеспецифичность. Обнаружено, что находящийся в составе LTR регуляторный элемент, подавляющий активность его промотора, остается активным во всех линиях клеток. Показана также тканеспецифичность энхансерной активности LTR. Таким образом, установлено, что LTR HERV-K обладают уникальным пакетом тканеспецифичных регуляторных элементов. Их обнаружение в области генов должно исследоваться с точки зрения возможного участия в регуляции активности данных генов.

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

К сохранили способность специфического взаимодействия с регуляторными белками клеток, что объясняет их функциональную активность.

#### СТАТЬИ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Nikolaev, L.G., S.B. Akopov, I.P. Chernov, E.D. Sverdlov. 2007. Maps of cis-regulatory nodes in megabase long genome segments are an inevitable intermediate step toward whole genome functional mapping. *Current Genomics* 8: 137-149.
2. Акопов, С.Б., И.П. Чернов, А.С. Ветчинова, С.С. Буланенкова, Л.Г. Николаев. 2007. Идентификация и картирование цис-регуляторных элементов внутри длинных геномных последовательностей. *Молек. биология* 41: 787-792
3. Акопов, С.Б., И.П. Чернов, С.С. Буланенкова, В. Скворцова, А.С. Ветчинова, Л.Г. Николаев. 2007. Методы идентификации эпигенетических элементов млекопитающих в длинных мультигенных геномных последовательностях. *Биохимия* 72: 725-732.
4. Chernov, I.P., K.A. Timchenko, S.B. Akopov, L.G. Nikolaev, E.D. Sverdlov. 2007. Identification of tissue-specific DNA-protein binding sites by means of two-dimensional electrophoretic mobility shift assay display. *Anal. Biochem.* 364: 60-66.
5. Vetchinova, A.S., S.B. Akopov, I.P. Chernov, L.G. Nikolaev, E.D. Sverdlov. 2006. Two-dimensional electrophoretic mobility shift assay: identification and mapping of transcription factor CTCF target sequences within an *FXVD5-COX7A1* region of human chromosome 19. *Anal. Biochem.* 354: 85-93.
6. Akopov, S.B., V.M. Ruda, V.V. Batrak, A.S. Vetchinova, I.P. Chernov, L.G. Nikolaev, J. Bode, E.D. Sverdlov. 2006. Identification, genome mapping, and CTCF binding of potential insulators within the *FXVD5-COX7A1* locus of human chromosome 19q13.12. *Mamm. Genome* 17: 1042-1049.
7. Сасс, А.В., В.М. Руда, С.Б. Акопов, Е.В. Снежков, Л.Г. Николаев, Е.Д. Сverdlov. 2005. Исследование регуляторного потенциала S/MAR-элементов при временной экспрессии. *Биоорганич. химия* 31: 77-81.
8. Iarovaia, O.V., S.B. Akopov, L.G. Nikolaev, E.D. Sverdlov, S.V. Razin. 2005. Induction of transcription within chromosomal DNA loops flanked by MAR elements causes an association of loop DNA with the nuclear matrix. *Nucleic Acids Res.* 33: 4157-4163.
9. Чернов, И.П., С.Б. Акопов, Л.Г. Николаев. 2004. Структура и функции участков прикрепления ДНК к ядерному матриксу (S/MARs). *Биоорганич. химия* 30: 3-14.
10. Акопов, С.Б., Л.Г. Николаев, О.Ю. Тырсин, А.С. Рузов, Е.Д. Сverdlov. 1997. Идентификация и характеристики 14 последовательностей китайского хомячка, предпочтительно связывающихся с ядерным матриксом. *Биоорганич. химия* 23: 727-731.
11. Nikolaev, L.G., T. Tsevegiyn, S.B. Akopov, L.K. Ashworth, and E.D. Sverdlov. 1996. Construction of a chromosome specific library of human MARs and mapping of matrix attachment regions on human chromosome 19. *Nucleic Acids Res.* 24: 1330-1336.
12. Akopov SB, Nikolaev LG, Khil PP, Lebedev YB, Sverdlov ED. 1998. Long terminal repeats of human endogenous retrovirus K family (HERV-K) specifically bind host cell nuclear proteins. *FEBS Lett* 421(3):229-233
13. Vinogradova, T. V., Leppik, L. P., Nikolaev, L. G., Akopov, S. B., Kleiman, A. M., Senyuta, N. B., and Sverdlov, E. D. 2001. Solitary human endogenous retroviruses-K LTRs retain transcriptional activity in vivo, the mode of which is different in different cell types, *Virology* 290, 83-90.
14. Trubetskoy DO, Zavalova LL, Akopov SB, Nikolaev LG. 2002. Purification of proteins specifically binding human endogenous retrovirus long terminal repeat by affinity elution chromatography. *J Chromatogr A. Nov* 8;976(1-2):95-101
15. А.Н.Доманский, С.Б.Акопов, Ю.Б.Лебедев, Л.Г.Николаев, Е.Д.Сverdlov. 2002.

- Энхансерная активность вневирусного длинного концевго повтора эндогенного ретровируса человека семейства HERV-K. *Биоорганич. химия* 28: 341-345.
16. Ruda VM, Akopov SB, Trubetskoy DO, Manuylov NL, Vetchinova AS, Zavalova LL, Nikolaev LG, Sverdlov ED. 2004. Tissue specificity of enhancer and promoter activities of a HERV-K(HML-2) LTR. *Virus Res. Aug;104(1):*11-6.
  17. Igor P. Chernov, Elena A. Stukacheva, Sergey B. Akopov, Dmitry A. Didych, Lev G. Nikolaev, and Eugene D. Sverdlov. 2008. A new technique for selective identification and mapping of enhancers within long genomic sequences. *Biotechniques. May;44(6):*775-84
  18. С.Б.Акопов, И.П.Чернов, Т.Вальстрем, М.Б.Костина, Г.Кляйн, М.Хенрикссон, Л.Г.Николаев. 2008. Идентификация участков узнавания для белков системы Мус/Мах/Мхд на хромосоме 19 человека путем селективного связывания. *Биохимия* 73: 1569-1579.
  19. Д.А.Дидыч, Н.А. Смирнов, Е.С.Котова, С.Б.Акопов, Л.Г.Николаев, Е.Д.Свердлов. 2011. Функциональная диссекция энхансерного элемента, расположенного во втором интроне гена U2AF1L4 человека. *Биохимия* 76: 951-957.
  20. Дидыч Д.А., Акопов С.Б., Снежков Е.В., Скапцова Н.В., Николаев Л.Г., Свердлов Е.Д. 2009. «Идентификация и картирование 10 новых потенциальных инсуляторов в области *FXVD5-COX7A1* хромосомы 19q13.12 человека». *Биохимия* 74(7), 728-733.
  21. Н.А. Смирнов, Д.А. Дидыч, С.Б. Акопов, Л.Г. Николаев, Е.Д. Свердлов. 2013. Оценка энхансер-блокирующей активности инсуляторов при помощи транзientной трансфекции. *Биохимия* 78, 895-903.
  22. D.A. Didych, E.S. Kotova, S.B. Akopov, L.G. Nikolaev<sup>s</sup>, E.D. Sverdlov. 2012 DNA fragments binding CTCF *in vitro* and *in vivo* are capable of blocking enhancer activity. *BMC Res Notes. Apr 5;5(1):*178.
  23. Nikolaev L.G, Akopov S.B., Didych D.A., Sverdlov E.D. 2009. Vertebrate protein CTCF and its multiple roles in a large-scale regulation of genome activity *Current Genomics*, , 8:137-149
  24. Котова Е., Дидыч Д., Акопов С., Николаев Л., Свердлов Е. 2013. Экспрессия гена CTCF кур в клетках COS-1 и получение обогащенной белком CTCF фракции. *Биохимия* 78, 879-83.
  25. E. S. Kotova, S. B. Akopov, E. D. Sverdlov, L. G. Nikolaev Transcription factor CTCF and mammalian genome organization. 2014. *Biopolymers and Cell. 30*, 260–272.