

## ДОМЕНЫ АЛЬФА- И БЕТА-ГЛОБИНОВЫХ ГЕНОВ В КОНТЕКСТЕ СТРУКТУРНО- ФУНКЦИОНАЛЬНОЙ ОРГАНИЗАЦИИ ЭУКАРИОТИЧЕСКОГО ГЕНОМА

©2012 г. С. В. РАЗИН<sup>1,2\*</sup>, С. В. УЛЬЯНОВ<sup>2</sup>,  
Е. С. ЮДИНKOBA<sup>1</sup>, Е. С. ГУЩАНСКАЯ<sup>1,2</sup>,  
А. А. ГАВРИЛОВ<sup>1,3</sup>, О. В. ЯРОВАЯ<sup>1</sup>

<sup>1</sup> Институт биологии гена Российской академии наук, Москва;

<sup>2</sup> Биологический факультет МГУ им М. В. Ломоносова, Москва;

<sup>3</sup> Университет Осло, Центр медицинских исследований в России,  
Москва

I. Введение. II. Домен бета-глобиновых генов. III. Домен альфа-глобиновых генов. IV. Происхождение и эволюция доменов разного типа. V. Заключение.

### I. ВВЕДЕНИЕ

В прокариотических клетках функционально связанные гены сгруппированы в опероны, являющиеся мишенями для регуляторных систем, контролирующих транскрипцию. В эукариотических клетках функционально связанные гены нередко располагаются на разных хромосомах. Соответственно, в течение многих лет было принято считать, что регуляция транскрипции осуществляется на уровне промоторов индивидуальных генов. В настоящее время очевидно, что

---

*Принятые сокращения:* DHS(s) – сайт(ы) гиперчувствительности к ДНКазе I (DNase I Hypersensitive Site); LCR – зона контроля локуса (Locus Control Region); п. н. (т. п. н.) – пары нуклеотидов (тысячи пар нуклеотидов); MCS – консервативный участок генома (Multispecies Conserved Sequence);  $\alpha$ -MRE – главный регуляторный элемент (Major Regulatory Element) домена  $\alpha$ -глобиновых генов; OR – ген обонятельного рецептора (Olfactory Receptor).

*Адрес для корреспонденции:* sergey.v.razin@usa.net

Работа выполнена при финансовой поддержке грантов Министерства наук и образования РФ (ГК № 16.740.11.0483; 14.740.12.1344; 16.740.11.0353), грантов Российского фонда фундаментальных исследований (11-04-00361-а; 11-04-91334-ННИО\_а; 12-04-00313-а; 12-04-00036-а), гранта Совета при Президенте РФ для поддержки молодых ученых (МК-3813.2012.4), грантов программы «Молекулярная и клеточная биология» Президиума РАН и фонда Дмитрия Зимина «Династия».

ситуация является более сложной. Изучение хроматина и эпигенетических механизмов, работающих на уровне упаковки генетического материала в клеточном ядре, привели к раскрытию многоуровневой системы регуляции транскрипции в эукариотической клетке. Наряду с контролем активности индивидуальных промоторов не менее важную роль играет определение статуса хроматинового домена (активный либо неактивный), в границах которого расположен данный ген [1]. Новые результаты, в частности демонстрация роли модификаций гистонов, совокупность которых принято называть гистоновым кодом, в контроле способа упаковки ДНК в хроматине, возродили интерес к доменной гипотезе организации эукариотического генома. Эта гипотеза была сформулирована в конце 70-х годов прошлого века [2, 3]. В классическом варианте гипотеза предполагает, что весь геном состоит из единообразных структурно-функциональных единиц – хроматиновых доменов. Хроматиновый домен можно определить как достаточно протяженный участок генома, в пределах которого могут происходить локальные, не распространяющиеся на фланкирующие области, изменения способа упаковки хроматиновой фибриллы. Хроматиновый домен может включать один или несколько генов. Активация хроматинового домена означает такое изменение его статуса, когда все присутствующие в домене гены становятся потенциально транскрипционно активными. Во многих случаях активация хроматинового домена коррелирует с переходом его из ДНКазо-устойчивой в ДНКазо-чувствительную конфигурацию. Одним из оснований для формулировки доменной гипотезы организации эукариотического генома как раз и послужили результаты работ по изучению чувствительности к ДНКазе I активных генов и фланкирующих их последовательностей. В пионерских работах лаборатории Вайнтрауба было продемонстрировано, что активно транскрибирующиеся тканеспецифичные гены предпочтительно перевариваются ДНКазой I при обработке этим ферментом пермеабелизированных ядер [4]. Далее было показано, что предпочтительная чувствительность к ДНКазе I характерна для довольно протяженной геномной области, включающей один или несколько транскрибирующихся генов и фланкирующие их последовательности ДНК [5, 6]. Два обстоятельства привлекли особое внимание исследователей: (1) одни и те же геномные области были предпочтительно чувствительны к ДНКазе I или относительно устойчивы к ней в клетках, дифференцированных по разным путям, и (2) границы ДНКазо-чувствительных областей были достаточно четкими [6]. Именно это и позволило предположить, что геном



Рис. 1. Схема, иллюстрирующая основную идею доменной гипотезы организации генома.

построен из единообразных структурно-функциональных блоков, в рамках которых транскрипционный статус генов может регулироваться посредством изменения способа упаковки хроматина (рис. 1). Молекулярные механизмы, контролирующие способ упаковки и статус хроматиновой фибриллы, долгое время оставались предметом изучения. В настоящий момент установлено, что основным фактором, контролирующим декомпактизацию хроматиновой фибриллы, является ацетилирование гистонов [7, 8]. Доменная модель организации генома предполагает, что должны существовать некие пограничные структуры, изолирующие домен от его окружения и ограничивающие процесс реконфигурации хроматина рамками индивидуальных доменов. Такие пограничные элементы были найдены и получили название инсуляторов. Впервые инсуляторы были обнаружены на границе локуса генов теплового шока *hsp70* у *Drosophila melanogaster* [9]. Впоследствии они были найдены у всех изученных групп эукариот [10, 11].

Решающим событием в становлении модели доменной организации генома явилось обнаружение области контроля локуса домена  $\beta$ -глобиновых генов человека, а затем и областей контроля локуса ряда других геномных доменов. Хотя результаты экспериментов по изучению дифференциальной чувствительности протяженных хроматиновых доменов к нуклеазам и обнаружение особого класса регуляторных элементов (инсуляторов), одной из функций которых является разграничение геномных доменов, хорошо объяснялись доменной гипотезой организации генома, оставалось не ясным, каким образом регулируется транскрипционный статус домена в целом. Логично было предположить, что существуют некие регуляторные элементы, контролирующие транскрипционный статус геномного/хроматинового домена. Такой регуляторный элемент был обнаружен в домене  $\beta$ -глобиновых генов человека. Идентификации данного элемента способствовали две линии наблюдений. При анализе природных делеций, влияющих на экспрессию  $\beta$ -глобиновых генов человека

(рис. 2А), было обнаружено, что так называемая «испанская делеция», в результате которой из генома удаляется достаточно протяженный фрагмент ДНК, включающий последовательности, локализованные в 5'-фланкирующей области кластера β-глобиновых генов, но не включающий, собственно, β-глобиновые гены и их промоторы, приводит к прекращению транскрипции всех β-глобиновых генов, расположенных на содержащей делецию хромосоме. Кроме того, домен β-глобиновых генов утрачивает предпочтительную чувствительность к ДНКазе I в эритроидных клетках, и изменяется время репликации домена (рис. 2Б). В нормальной ситуации репликация домена β-глобиновых генов в эритроидных клетках происходит в начале S-фазы (что является характерным признаком большинства транскрибирующихся генов), тогда как в хромосоме с испанской делецией домен реплицируется в конце S-фазы (что характерно для гетерохроматина и тканеспецифичных генов, неактивных в данном типе клеток) [12]. При испанской делеции из генома удаляется кластер участков гиперчувствительности к ДНКазе I (DNase I Hypersensitive Site, DHS), расположенный на расстоянии 6–22 т.п.н. перед кластером β-глобиновых генов [12] (рис. 2Б). Было высказано предположение о том, что данный кластер DHSs содержит регуляторный элемент, контролирующий транскрипционный статус домена β-глобиновых генов. Для проверки этого предположения были получены трансгенные мыши, в геном которых был интегрирован минидомен, содержащий один из β-глобиновых генов человека и упомянутый кластер DHSs, который удаляется из генома при испанской делеции (рис. 2В). Анализ этих трансгенных мышей продемонстрировал, что присутствие в составе конструктора изучаемого кластера DHSs обеспечивает высокий и не зависящий от позиции интеграции (пропорциональный числу интегрированных копий) уровень экспрессии β-глобинового гена человека в эритроидных клетках мыши. Авторы сделали вывод о том, что изучаемый ими кластер DHSs содержит регуляторный элемент, способный создавать автономный эритроид-специфичный минидомен в геноме трансгенных мышей [13]. Обнаружение регуляторного элемента доменного уровня, получившего название LCR (Locus Control Region, Область Контроля Локуса) в доменах β-глобиновых генов других позвоночных животных [14], способствовало широкому признанию основных положений доменной гипотезы организации генома [2, 3]. Одновременно домен β-глобиновых генов человека и аналогичным образом устроенные домены β-глобиновых генов различных позвоночных животных приобрели статус наиболее популярных моделей, на примере которых анализировались различные механизмы регуляции экспрессии генов. Концентрация усилий

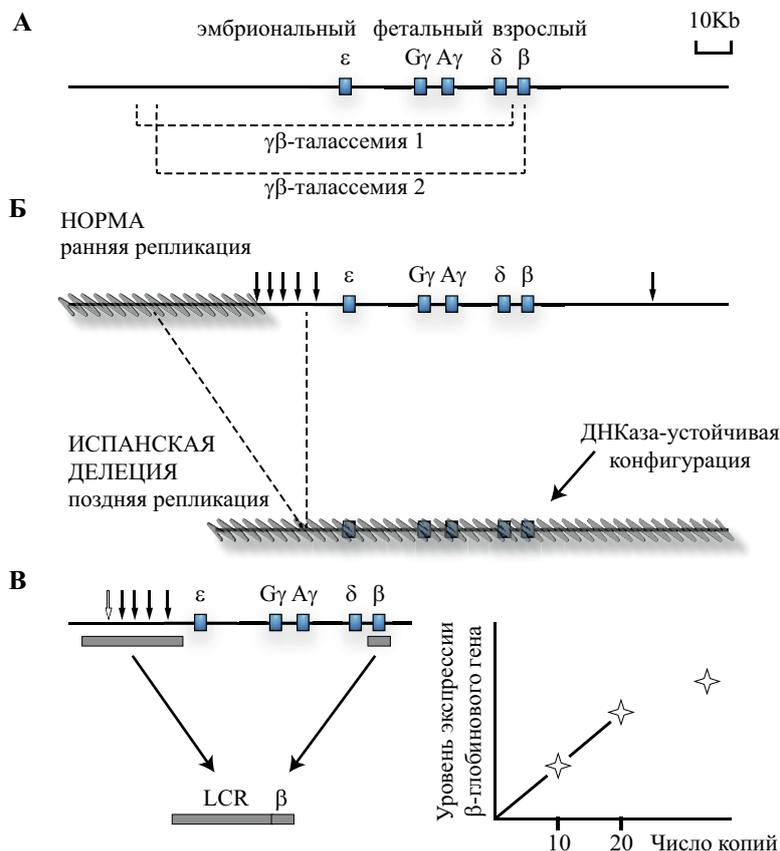


Рис. 2. Идентификация области контроля локуса (LCR)  $\beta$ -глобиновых генов человека.

(А) Природные делеции, приводящие к появлению различных талассемий в силу физического удаления всех или большинства  $\beta$ -глобиновых генов. Делетированные фрагменты показаны пунктирными линиями под схемой домена.

(Б) Схема, иллюстрирующая позицию испанской делеции и последствия этой делеции. Вертикальные стрелки показывают позиции участков гиперчувствительности к ДНКазе I. Неактивные области генома обозначены наклонным серым зигзагом. В результате испанской делеции, удаляющей кластер участков гиперчувствительности к ДНКазе I и сегмент генома, примыкающий к этому кластеру с 5'-конца, домен  $\beta$ -глобиновых генов полностью инактивируется.

(В) Схема, демонстрирующая стратегию создания эритроид-специфичного мини-домена, и результаты анализа уровня экспрессии находящегося в составе такого мини-домена  $\beta$ -глобинового гена человека в геноме трансгенных мышей. При относительно низком числе копий мини-домена суммарное количество продукта транскрипции гена  $\beta$ -глобина человека пропорционально числу интегрированных в геном трансгенной мыши копий содержащего этот ген мини-домена.

многих исследовательских коллективов на характеристике одной и той же группы модельных систем позволила раскрыть многие фундаментальные принципы, лежащие в основе регуляции транскрипции у эукариот. Об этом мы будем говорить более подробно в следующих разделах настоящего обзора. Здесь же мы хотели бы подчеркнуть, что сфокусированность исследований на одной группе модельных систем в течение ряда лет препятствовала осознанию того факта, что представление об эукариотическом геноме, как мозаике из единообразно организованных хроматиновых доменов, которые могут быть либо активными, либо неактивными, является упрощенным и, в общем случае, неверным. Несостоятельность классической модели доменной организации генома [2, 3] стала очевидной, когда было продемонстрировано, что устройство доменов  $\alpha$ -глобиновых генов млекопитающих и птиц не может быть объяснено в рамках данной модели. Эти домены локализуются в сегментах хромосом, обогащенных постоянно экспрессирующимися генами (генами «домашнего хозяйства»), и демонстрируют высокий уровень чувствительности к ДНКазе I как в эритроидных, так и в неэритроидных клетках (т.е. в независимости от транскрипционного статуса  $\alpha$ -глобиновых генов) [15]. Домен  $\alpha$ -глобиновых генов не изолирован инсульторами от соседних областей генома. Более того, главный регуляторный элемент домена  $\alpha$ -глобиновых генов расположен в интроне расположенного рядом с доменом гена домашнего хозяйства [16]. Все эти особенности организации тканеспецифичного домена  $\alpha$ -глобиновых генов трудно объяснить в рамках классической модели доменной организации генома. Перекрывание геномных доменов и даже транскрипционных единиц оказалось не таким уж редким явлением в геноме высших эукариот [17]. Ясно, что такие гены не могут быть помещены в изолированные структурно-функциональные домены, подобные домену  $\beta$ -глобиновых генов. Все это потребовало переосмысления доменной модели организации генома. В результате была сформулирована концепция открытых и закрытых геномных доменов (не следует путать с «открытой» и «закрытой» конформацией хроматина) [14, 18]. Открытый геномный домен (домен с размытыми границами) определяется прежде всего на функциональном уровне и может перекрываться на хромосоме с другими такими же доменами. Часто такие домены включают гены, которые функционально и филогенетически не связаны. Понятно, что регуляторные системы открытых геномных доменов (доменов с размытыми границами) должны существенно отличаться от регуляторных систем доменов с фиксированными границами (закрытых доменов).

В последующих главах настоящего обзора мы подробно проанализируем особенности регуляторных механизмов транскрипции в доменах открытого и закрытого типа на примерах доменов  $\alpha$ - и  $\beta$ -глобиновых генов.

## II. ДОМЕН БЕТА-ГЛОБИНОВЫХ ГЕНОВ

### ОБЩАЯ ХАРАКТЕРИСТИКА

Наиболее хорошо изучены домены  $\beta$ -глобиновых генов человека, мыши и курицы (рис. 3). У человека и мыши домены  $\beta$ -глобиновых генов расположены соответственно на 11 и 7 хромосомах и находятся внутри существенно более протяженного кластера генов обонятельных рецепторов. Протяженность этих доменов точно не известна в силу того, что эритроидспецифичные DHSs, которые предположительно показывают позиции регуляторных элементов домена  $\beta$ -глобиновых генов, присутствуют на значительном расстоянии перед кластером  $\beta$ -глобиновых генов внутри сегмента хромосомы, содержащего гены обонятельных рецепторов [14]. Эnhансеры, способные активировать транскрипцию  $\beta$ -глобиновых генов, присутствуют и в 3'-концевой фланкирующей области кластера  $\beta$ -глобиновых генов человека. Приближение этих энхансеров к генам, кодирующим фетальные  $\beta$ -глобиновые цепи, происходящее в результате ряда природных делеций, приводит к развитию особой группы талассемий (болезней, связанных с отсутствием или дисбалансом экспрессии различных субъединиц гемоглобина) [19]. В эритроидных клетках мыши область предпочтительной чувствительности к ДНКазе I существенно превосходит по размерам ограниченный инсульторами кластер  $\beta$ -глобиновых генов и включает целый ряд генов обонятельных рецепторов, расположенных как в 5'-, так и в 3'-концевой фланкирующей областях [20]. Следует, однако, подчеркнуть, что присутствующие в границах этого ДНКазо-чувствительного домена гены обонятельных рецепторов не экспрессируются в эритроидных клетках, что хорошо иллюстрирует принцип многоуровневой регуляции активности генов. В эритроидных клетках человека домен предпочтительной чувствительности к ДНКазе I, включающий кластер  $\beta$ -глобиновых генов, не был детально проанализирован. Тем не менее, ряд косвенных свидетельств указывает на то, что и у человека этот домен является весьма протяженным и существенно превосходит по размерам геномный сегмент, ограниченный областью контроля локуса и 3'-концевым инсультором [21].

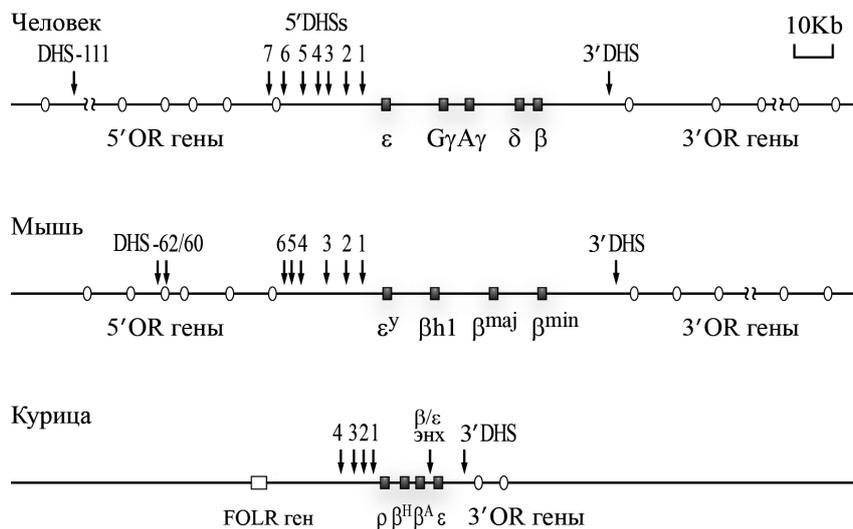


Рис. 3. Геномное окружение кластеров  $\beta$ -глобиновых генов человека, мыши и курицы.

DHSs, локализованные в пределах LCR, обозначены номерами, возрастающими по мере удаления от LCR. 3'DHS – консервативный DHS, локализованный на 3'-конце кластера  $\beta$ -глобиновых генов и колокализующийся с инсулятором. Эритроид-специфичные DHSs, локализованные на больших расстояниях перед кластером  $\beta$ -глобиновых генов в геномах человека и мыши, обозначены стрелками с указанием расстояний (в т.п.н.) от кластера  $\beta$ -глобиновых генов. На схемах использованы традиционные названия  $\beta$ -глобиновых генов, которые широко используются в современной литературе, несмотря на то что в базах данных используются другие обозначения этих генов (см. текст). OR – гены обонятельных рецепторов.  $\beta/\epsilon$  энх – энхансер, локализованный между генами  $\beta$  и  $\epsilon$ , присутствующий только в домене  $\beta$ -глобиновых генов кур, FOLR – ген фолатного рецептора.

Для позвоночных животных характерным является наличие нескольких глобиновых генов (как  $\alpha$ -, так и  $\beta$ -типа) которые дифференциально экспрессируются на различных стадиях развития организма. Важной особенностью кластеров  $\beta$ -глобиновых генов человека и млекопитающих животных является расположение генов на хромосоме в порядке их активации по ходу развития организма (рис. 3). Эмбриональный ген  $\epsilon$  расположен ближе всего к LCR [22].

Домен  $\beta$ -глобиновых генов кур является, пожалуй, наиболее типичным представителем доменов закрытого типа. Этот домен, имеющий протяженность порядка 33 т.п.н., расположен на первой хромосоме

между геном фолатного рецептора и кластером генов обонятельных рецепторов (рис. 3). Функциональный домен, включающий область контроля локуса и ограниченный 5'- и 3'-инсуляторами, практически точно колокализуется с эритроидспецифичным доменом предпочтительной чувствительности к ДНКазе I и эритроидспецифичным доменом гиперацетилирования гистонов [23]. Между LCR домена  $\beta$ -глобиновых генов и геном фолатного рецептора находится permanently конденсированная область (факультативный гетерохроматин) [24]. Домен включает кластер из четырех  $\beta$ -глобиновых генов:  $\rho$  (HBG1),  $\beta$ H (HBE1),  $\beta$ A (HBG2) и  $\epsilon$  (HBE), а также ряд регуляторных элементов, совпадающих с сайтами гиперчувствительности к ДНКазе I и необходимых для регуляции транскрипции, времени репликации и статуса хроматинового домена [25].

#### ПОГРАНИЧНЫЕ ЭЛЕМЕНТЫ ДОМЕНА $\beta$ -ГЛОБИНОВЫХ ГЕНОВ

О пограничных элементах генома (инсуляторах) уже говорилось в разделе I настоящего обзора (см. также [11, 26]). Классические инсуляторы обладают двумя активностями – энхансер блокирующей (препятствующей действию энхансера на промотор, расположенный за инсулятором [9]) и барьерной (препятствующей процессивному распространению ковалентных модификаций гистонов в составе хроматина [27]). Существуют неполноценные инсуляторы, обладающие лишь одной из перечисленных активностей. Их правильно называть либо энхансер-блокирующими, либо барьерными элементами. К сожалению, не все авторы следуют данному правилу, в связи с чем в литературе термин «инсулятор» нередко используется как для обозначения энхансер-блокирующих элементов, не обладающих барьерной активностью, так и для обозначения барьерных элементов, не обладающих энхансер-блокирующей активностью.

Наиболее хорошо изученным инсулятором позвоночных животных является инсулятор, расположенный на 5'-конце LCR домена  $\beta$ -глобиновых генов кур и колокализирующийся с DHS4 (см. рис. 3). Данный инсулятор является полноценным, т.е. обладает как энхансер-блокирующей, так и барьерной активностями [28, 29]. Минимальный фрагмент ДНК, обладающий инсуляторной активностью (коровый элемент), колокализирующийся с DHS4, имеет размер 250 п. н. и представляет собой CpG-островок, обладающий ограниченным структурным сходством с промоторами генов домашнего хозяйства [30]. В границах минимального инсулятора было картировано 5 сайтов связывания различных белков [30]. Один из этих сайтов (так называемый футпринт II, FII) является необходимым и достаточным для

проявления энхансер-блокирующей активности. С этим сайтом связывается многофункциональный транскрипционный фактор CTCF [31]. При делеции FII энхансер-блокирующая активность инсулятора утрачивается. Работы последних лет демонстрируют, что в работе энхансер-блокирующего элемента наряду с CTCF участвует белок когезин. Сайт связывания CTCF перекрывается с сайтом связывания когезина. При нокдаунах как по CTCF, так и по когезину, энхансер-блокирующая активность DHS4 исчезает [32]. Показано, что для проявления энхансер-блокирующей активности необходимо физическое взаимодействие CTCF и когезина [33].

Барьерная активность инсулятора сохраняется при делеции сайта связывания CTCF [31]. Это означает, что барьерная и энхансер-блокирующая активность обеспечиваются разными структурными элементами инсулятора, причем барьерная активность инсулятора обеспечивается взаимодействием с белками, отличными от CTCF [29], в частности, белком USF1, связывающимся с инсулятором и привлекающим к нему комплексы H3K4- и H4R3-специфичных гистонметилтрансфераз, гистонацетилтрансфераз и комплексы ремоделирования хроматина [34, 35]. Таким образом, представление об инсуляторе, как о пассивном элементе, препятствующем распространению всяческих сигналов (теория «дорожной пробки»), мало соответствует действительности. На деле инсулятор является точкой нуклеации, где собираются различные ферментные комплексы, осуществляющие ремоделирование хроматина и модификации гистонов.

3'-граница домена β-глобиновых генов кур изучена гораздо хуже. Показано, что на расстоянии ~ 5 т. п. н. после гена ε расположен конститутивный сайт гиперчувствительности к ДНКазе I (3'-DHS). Он проявляет ограниченное структурно-функциональное сходство с 5'-концевым инсулятором домена (DHS4). Внутри 3'-DHS картирован ряд сайтов связывания CTCF. Соответственно, 3'-DHS обладает энхансер-блокирующей активностью [36]. Однако, в отличие от DHS4, 3'-DHS не способен защитить трансген от эффекта положения, т.е. не обладает барьерной функцией, присущей классическим инсуляторам.

В доменах β-глобиновых генов человека и мыши были идентифицированы CTCF-зависимые энхансер-блокирующие элементы, которые расположены примерно так же, как и в домене β-глобиновых генов кур – на 5'-конце области контроля локуса и в 3'-концевой фланкирующей области кластера β-глобиновых генов [37–39]. Другими авторами было показано, что 5'-инсулятор домена β-глобиновых генов человека способен защищать трансгены от позиционных эффектов, т.е., подобно 5'-DHS4 домена β-глобиновых генов кур, обладает одновременно энхансер-блокирующей и барьерной активностями [40].

Функциональная нагрузка энхансер-блокирующих элементов, расположенных на 5'-конце LCR доменов  $\beta$ -глобиновых генов человека и мыши, не вполне понятна. Как уже говорилось выше (первый подраздел Раздела II), регуляторные элементы этих доменов присутствуют и за пределами области, ограниченной этими энхансер-блокирующими элементами.

#### РЕГУЛЯЦИЯ ТРАНСКРИПЦИИ $\beta$ -ГЛОБИНОВЫХ ГЕНОВ

Главным регуляторным элементом, контролирующим экспрессию  $\beta$ -глобиновых генов, является LCR. Хотя LCR прежде всего заинтересовал исследователей в связи с предположением о том, что этот регуляторный элемент контролирует способ упаковки геномного домена в хроматине, результаты, касающиеся этой активности LCR, остаются крайне противоречивыми. С одной стороны, не вызывает сомнений тот факт, что LCR способен сформировать активный эритроидспецифичный хроматиновый домен в эктопической позиции [41, 42]. С другой стороны, направленная делеция всего LCR из нормальной геномной позиции приводит лишь к резкому снижению уровней экспрессии глобиновых генов, в то время как предпочтительная чувствительность домена  $\beta$ -глобиновых генов к ДНКазе I в эритроидных клетках и повышенный уровень ацетилирования гистонов сохраняются [21, 43]. Причины очевидных различий между детектируемыми активностями LCR в нормальной и эктопических позициях не вполне понятны. Наиболее вероятное предположение состоит в том, что в 5'-фланкирующей области домена  $\beta$ -глобиновых генов присутствует несколько регуляторных элементов, каждый из которых способен поддерживать активный статус этого домена в эритроидных клетках. В этом случае удаление собственно LCR не будет приводить к утрате ДНКазо-чувствительного домена, в то время как удаление более протяженного фрагмента ДНК, содержащего LCR и не идентифицированные пока дополнительные регуляторные элементы, будет иметь более драматические последствия, что и наблюдается при испанской делеции [12].

LCR домена  $\beta$ -глобиновых генов человека состоит из ряда модулей, колокализующихся с индивидуальными участками гиперчувствительности к ДНКазе I (DHS2–5, рис. 3). DHS1 не удаляется при испанской делеции (рис. 2Б) [12], в силу чего его обычно не считают частью LCR. DHS2–5 можно рассматривать как субдомены LCR. Один из этих субдоменов (DHS5) является инсулятором, о чем уже говорилось в предыдущем разделе. Фрагменты ДНК, содержащие DHS2–4, оказались эритроид-специфичными энхансерами, что было

продемонстрировано в экспериментах по трансфекции конструкторов с репортерным геном [44]. Наиболее сильный энхансер был картирован в области DHS2 [44]. В DHS 2–4 были картированы многочисленные сайты связывания различных транскрипционных факторов, в частности GATA-1, NF-E2, AP1, CP2 [45, 46]. Короткие (~1 т.п.н.) фрагменты ДНК, происходящие из DHS2 и DHS4, оказались способными направлять образование эритроид-специфичных DHS в эктопических позициях в геноме трансгенных мышей [47]. При более детальном анализе фрагмента ДНК из DHS4 удалось идентифицировать минимальный фрагмент (101 п.н), способный направлять образование эритроид-специфичного DHS в эктопических позициях. В составе этого фрагмента ДНК были идентифицированы сайты связывания SP1, GATA1, AP1, NF-E2, причем сайты связывания GATA1 были представлены тандемом обращенных повторов, который располагался на расстоянии ~50 п.н. после сайта связывания AP1/NF-E2 [47]. Позднее такая организация сайтов связывания AP1/NF-E2 и GATA1 была выявлена и в DHS1–3 [48]. Мутации, препятствующие связыванию NF-E2 и AP1, приводили к утрате способности направлять образование DHSs в эктопических позициях [48].

В доменах  $\beta$ -глобиновых генов различных позвоночных животных LCR организован так же, как и в домене  $\beta$ -глобиновых генов человека [49–51]. Во всех этих доменах LCR располагается перед генным кластером (по ходу транскрипции глобиновых генов) и состоит из нескольких функциональных блоков, позиции которых совпадают с позициями DHSs. Существует значительная степень гомологии между последовательностями ДНК, присутствующими в индивидуальных функциональных блоках как в рамках одного LCR, так и между LCR доменов  $\beta$ -глобиновых генов различных организмов. Прежде всего, это касается набора и взаимного расположения сайтов связывания ключевых эритроидспецифичных транскрипционных факторов [48].

После того как стало ясно, что LCR представляет собой блок тканеспецифичных энхансеров, вопрос о механизме действия LCR редуцировался до вопроса о механизме действия энхансеров. Применительно к домену  $\beta$ -глобиновых генов существует также необходимость объяснить избирательную активацию разных генов по ходу развития организма (так называемое переключение экспрессии глобиновых генов). Наиболее популярные в настоящее время модели действия энхансеров можно разделить на две большие группы: модели прямого взаимодействия энхансера с промотором и модели формирования активаторного комплекса на энхансере с последующим перемещением его на промотор. Первая группа моделей основана на предположении,

что энхансер физически взаимодействует с промотором, способствуя стабилизации формирующегося пре-инициационного комплекса транскрипции [52]. Так как энхансеры нередко расположены на значительном расстоянии от промоторов, прямое взаимодействие этих регуляторных элементов возможно лишь при выпетливании разделяющего их фрагмента хроматиновой фибриллы. В силу этого в англоязычной литературе модели, постулирующие прямое взаимодействие энхансера с промотором, получили название «looping models» («модель выпетливания»).

С чисто статистической точки зрения, вероятность установления прямого контакта между LCR и тем или иным промотором должна быть обратно пропорциональна расстоянию между промотором и LCR. В этой связи последовательное расположение  $\beta$ -глобиновых генов в порядке их активации по ходу развития организма, типичное для кластеров  $\beta$ -глобиновых генов млекопитающих животных [53], может иметь важное функциональное значение. В силу своей близости к LCR эмбриональный ген  $\epsilon$  должен транскрибироваться предпочтительно по сравнению с другими  $\beta$ -глобиновыми генами. Эффективная транскрипция последних будет возможна лишь после принудительного выключения гена  $\epsilon$  с помощью некоего эпигенетического механизма. Действительно, в экспериментах с трансгенными мышами, в геном которых были внедрены перестроенные копии домена  $\beta$ -глобиновых генов человека, было продемонстрировано, что позиция гена внутри кластера определяет время экспрессии этого гена по ходу индивидуального развития [39, 54].

В рамках классической модели прямого взаимодействия LCR с промотором довольно трудно объяснить тот факт, что в эритроблестах «взрослого типа» одновременно экспрессируются по меньшей мере два глобиновых гена (у человека – гены  $\sigma$  и  $\beta$ ). Было высказано предположение, что LCR поочередно формирует короткоживущие альтернативные комплексы с промоторами активируемых генов. Это предположение получило определенные подтверждения при анализе транскрипции  $\beta$ -глобиновых генов в индивидуальных клетках, который проводился с использованием техники флуоресцентного иммуноокрашивания интронных последовательностей в составе первичных транскриптов [55, 56].

Некоторое время назад был разработан экспериментальный подход, позволяющий прямо анализировать пространственную конфигурацию различных хромосомных локусов. Эта процедура, получившая название фиксации конформации хромосомы (Chromosome Conformation Capture, 3C), позволяет определить относитель-

ные частоты взаимодействий между различными парами геномных элементов [57]. С использованием метода 3С было продемонстрировано, что в домене  $\beta$ -глобиновых генов мыши LCR действительно прямо взаимодействует с промоторами активируемых генов. Одновременно оказалось, что эти гены находятся в комплексе не только с LCR, но и друг с другом. Это послужило основой модели активаторного хроматинового блока (хаба; *hub* (англ.) – узловая станция) [58, 59]. Активаторный хроматиновый блок представляет собой сложный комплекс регуляторных последовательностей ДНК и связанных с ними транскрипционных факторов. Удаленные регуляторные элементы и промоторы подконтрольных генов находятся в непосредственном физическом контакте, будучи связанными посредством взаимодействий транскрипционных факторов, одни из которых узнают промоторы генов, а другие располагаются на регуляторных элементах. Активаторные хроматиновые блоки представляют собой динамичные комплексы, стабильность которых зависит от спектра транскрипционных факторов, принимающих участие в их формировании [58, 60]. Сборка хроматинового хаба не происходит одновременно. В качестве примера можно рассмотреть активаторный хаб домена  $\beta$ -глобиновых генов мыши [61]. Этот домен включает четыре гена, активность которых зависит от стадии развития. Гены  $\beta^{\text{maj}}$  и  $\beta^{\text{min}}$  экспрессируются во взрослом организме, гены  $\epsilon$  и  $\beta_{\text{H1}}$  активны на эмбриональных стадиях развития. На стадии дифференцировки эритроидных клеток, которая предшествует активации транскрипции глобиновых генов, в локусе  $\beta$ -глобиновых генов собирается незрелый хроматиновый хаб, включающий в себя LCR и ряд удаленных регуляторных последовательностей (5'DHS – 85, 5'DHS – 60/–62). После перехода клеток к терминальной дифференцировке к нему привлекаются также промоторы активных  $\beta$ -глобиновых генов. Строго говоря, неким «минимальным» хабом можно считать комплекс энхансера с промотором, однако, во всех изученных системах зрелый хроматиновый хаб, формирующийся с участием области контроля локуса, всегда содержит дополнительные модулирующие работу хаба регуляторные элементы [61]. Являясь в настоящее время общепризнанной [61–63], модель хроматинового хаба не объясняет ряда известных фактов. Так, в терминах предложенной модели значение строго определенного расположения глобиновых генов в порядке их активации в онтогенезе объяснить сложно. Механизм переключения экспрессии глобиновых генов, собственно говоря, тоже не объясняется. Авторы ограничиваются констатацией факта, что в эритроидных клетках эмбрионального типа в составе хаба присут-

ствуют эмбриональные  $\beta$ -глобиновые гены, тогда как в эритроидных клетках взрослого типа в составе хаба присутствуют «взрослые» глобиновые гены [58, 59].

Характерной чертой LCR является высокая концентрация сайтов связывания транскрипционных факторов [51]. Будучи связанными с LCR, эти транскрипционные факторы могут привлечь к LCR различные мультиферментные комплексы, осуществляющие ацетилирование гистонов и ремоделирование хроматина [64, 65]. Таким образом, LCR может стать центром формирования активного хроматинового домена [66, 67]. В общем случае, активный домен должен распространяться в обе стороны. Однако, охарактеризованные LCR доменов  $\beta$ -глобиновых генов содержат на 5'-конце инсультор. Уже в силу этого распространение активирующих модификаций хроматина будет полярным. Полярность распространения активирующего сигнала может обеспечиваться и по-другому. Согласно ряду данных, LCR является местом инициации так называемой межгенной транскрипции [68]. Известно, что транскрипционный комплекс привлекает к транскрибируемой области гистон-ацетилазы и комплексы ремоделирования хроматина [69, 70]. Перемещаясь в сторону кластера глобиновых генов, транскрипционный комплекс будет обеспечивать реорганизацию всего транскрибируемого домена [71]. Процессивное распространение активирующего сигнала от LCR к генам является основным постулатом моделей действия LCR, включающих этап сканирования [72]. Следует сказать, что модели действия LCR, постулирующие этап сканирования, и модели, постулирующие прямой контакт LCR с промоторами активируемых генов, не являются взаимоисключающими. Модель сканирования скорее объясняет механизм формирования активного хроматинового домена. Этот процесс может происходить на тех стадиях дифференцировки, которые предшествуют началу активной транскрипции глобиновых генов.

#### СУБДОМЕННАЯ ОРГАНИЗАЦИЯ ДОМЕНА $\beta$ -ГЛОБИНОВЫХ ГЕНОВ

Анализ относительного уровня чувствительности к ДНКазе I различных сегментов домена  $\beta$ -глобиновых генов человека позволил идентифицировать три субдомена в рамках данного домена: регуляторный субдомен, включающий LCR, эмбрионально-фетальный субдомен, включающий гены  $\epsilon$ ,  $\zeta\gamma$ ,  $\Lambda\gamma$ , и взрослый субдомен, включающий гены  $\sigma$  и  $\beta$  [68]. Хотя весь домен  $\beta$ -глобиновых генов отличался повышенной чувствительностью к ДНКазе I в эритроидных клетках, уровни чувствительности эмбрионально-фетального и взрослого субдоменов являются существенно более высокими в тех клетках, где

соответствующие гены транскрибируются. Что касается субдомена, содержащего LCR, то он характеризуется очень высоким уровнем чувствительности к ДНКазе I во всех эритроидных клетках, в том числе и на стадии дифференцировки, предшествующей началу транскрипции глобиновых генов [68]. Те же субдомены можно различить и посредством анализа уровней ацетилирования гистонов в разных типах эритроидных клеток [21, 73]. Логично считать, что дифференциальная активация эмбрионально-фетального и взрослого субдоменов существенна для переключения экспрессии глобиновых генов по ходу развития организма. Было продемонстрировано, что статус субдоменов регулируется посредством так называемой интергенной транскрипции [68]. Делеция промотора, контролирующего транскрипцию «взрослого» субдомена домена  $\beta$ -глобиновых генов человека, приводила к тому, что гены  $\sigma$  и  $\beta$  не транскрибировались, несмотря на то что промоторы этих генов и LCR не были повреждены. Это позволяет предполагать, что активный статус взрослого субдомена обеспечивается процессивным ацетилированием гистонов в границах этого субдомена, опосредованным процессом низкоуровневой транскрипции РНК-полимеразой II [71].

### III. ДОМЕН АЛЬФА-ГЛОБИНОВЫХ ГЕНОВ

#### ОБЩАЯ ХАРАКТЕРИСТИКА

Подобно доменам  $\beta$ -глобиновых генов, наиболее полно изучены домены  $\alpha$ -глобиновых генов человека, мыши и курицы. Гены, кодирующие  $\alpha$ -субъединицу гемоглобина, находятся у человека на 16-й хромосоме, у курицы и мыши на 14-й и 11-й хромосомах, соответственно. Домен  $\alpha$ -глобиновых генов построен у всех этих видов по единому образцу: один эмбриональный ген ( $\zeta$ ,  $\pi$ ), по крайней мере один  $\alpha$ -глобиновый ген взрослого типа ( $\alpha$ ), и один или два минорных  $\alpha$ -глобиновых гена взрослого типа ( $\alpha^D$ ,  $\theta$ ,  $\mu$ ) [74] (рис. 4). Домены  $\alpha$ -глобиновых генов можно отнести к доменам открытого типа, с присущими им особенностями: они располагаются в богатых генами областях хромосом, являются предпочтительно-чувствительными к нуклеазам во всех типах клеток и реплицируются в начале S-фазы клеточного цикла [15, 75, 76]. Кластер  $\alpha$ -глобиновых генов фланкирован генами «домашнего хозяйства», которые активно транскрибируются во всех изученных типах клеток [16, 77].

Сравнительный анализ расположения генов и консервативных участков последовательности ДНК вокруг кластера  $\alpha$ -глобиновых

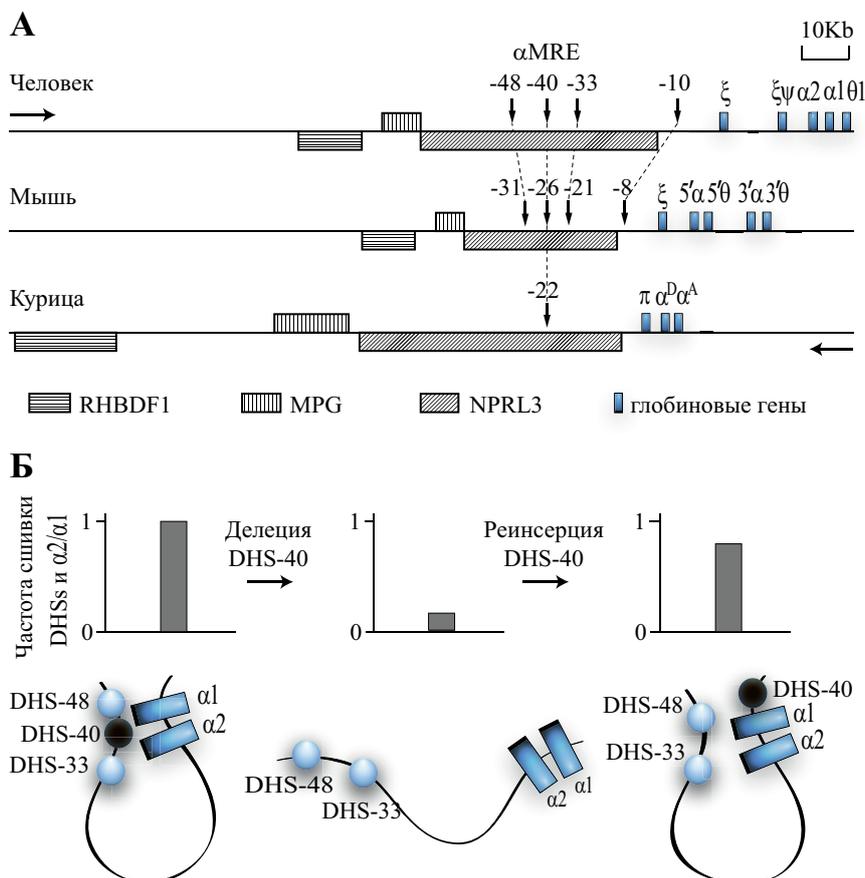


Рис. 4. Домены  $\alpha$ -глобиновых генов человека, мыши и курицы.

(А) Область генной синтении, фланкирующая кластеры  $\alpha$ -глобиновых генов с 5'-конца. Вертикальными стрелками показаны DHSs, колокализующиеся с MCS-R1–MCS-R4 (см. текст). Цифры над стрелками показывают расстояние в т.п.н. между DHSs и кластером  $\alpha$ -глобиновых генов. На схемах использованы традиционные названия  $\alpha$ -глобиновых генов. Горизонтальными стрелками показано направление транскрипции: «глобиновое» – гены над линией, «антиглобиновое» – гены под линией.

(Б) Структура активаторного комплекса домена  $\alpha$ -глобиновых генов человека и эксперименты, демонстрирующие ключевую роль  $\alpha$ -MRE в сборке активаторного комплекса (согласно [96]). На графиках суммированы результаты экспериментов по анализу пространственного взаимодействия DHSs домена с промоторами  $\alpha 2$ - и  $\alpha 1$ -глобиновых генов с использованием метода 3С-анализа, под графиками – интерпретация 3С-данных.

генов человека, мыши, курицы и рыбки-зебры выявил область синтении, которая включает весь кластер  $\alpha$ -глобиновых генов и протяженный 5'-концевой фланкирующий участок (рис. 4) [16, 75]. Во всех изученных геномах позвоночных животных непосредственно перед кластером  $\alpha$ -глобиновых генов расположен ген, который транскрибируется в направлении, противоположном направлению транскрипции глобиновых генов. Этот ген в течение долгого времени фигурировал в различных базах данных под названием C16orf35, но в последнее время был переименован в NPRL3 (NPR3 like) [78]. NPRL3 транскрибируется в клетках, дифференцированных по разным путям (т. е. является геном «домашнего хозяйства»). В пятом интроне NPRL3 расположен главный регуляторный элемент домена  $\alpha$ -глобиновых генов, обладающий некоторыми характеристиками области контроля локуса [16, 79]. Удаление этого элемента в результате спонтанной делеции 16-т.п.н.-фрагмента ДНК приводит к тяжелой форме талассемии. Уровень транскрипции гена  $\alpha$  в поврежденной хромосоме составляет менее 1% от нормального уровня [80].

Впервые удаленный регуляторный элемент был идентифицирован в домене  $\alpha$ -глобиновых генов человека и был назван DHS40, так как располагался в одном из гиперчувствительных к ДНКазе I участков, на расстоянии 40 т.п.н. перед первым геном домена [79]. Впоследствии он получил название  $\alpha$ -MRE (major regulatory element) [81, 82]. DHS40 представляет собой сильный эритроид-специфичный энхансер [83], его «коровый» элемент, фрагмент длиной 350 п.н., содержит несколько сайтов связывания эритроид-специфичных факторов транскрипции, в частности GATA1/2 и NF-E2 [84]. Наиболее важными элементами  $\alpha$ -MRE у всех изученных организмов являются два расположенных на расстоянии 21 п.н. участка связывания NF-E2 [16], которые в литературе нередко называют участками связывания Maf (YGCTGASTCAY; Maf recognition sequences, MARE), по названию малой субъединицы NF-E2 [85]. Пара MARE-элементов фланкирована сайтами связывания эритроид-специфичного транскрипционного фактора GATA.

Наряду с  $\alpha$ -MRE в доменах  $\alpha$ -глобиновых генов различных позвоночных животных присутствуют и другие регуляторные элементы. На это указывает, в частности, присутствие в этих доменах целого ряда консервативных участков гиперчувствительности к ДНКазе I и просто консервативных участков генома (так называемые MCS, multispecies conserved sequences). Наиболее консервативными являются MCSs, локализующиеся с промоторами (MCS-P) и регуляторными элементами (MCS-R) [74]. Для нашего обсуждения наибольший

интерес представляют MCS-R1–MCS-R4, расположенные в 5'-фланкирующей области кластера  $\alpha$ -глобиновых генов. В геноме человека они колокализуются с DHS-48 (MCS-R1), DHS-40 (MCS-R2), DHS-33 (MCS-R3) DHS-10 (MCS-R4);  $\alpha$ -MRE расположен в MCS-R2 (рис. 4А) [74].

Расположение различных регуляторных элементов хорошо изучено в домене  $\alpha$ -глобиновых генов кур. Помимо  $\alpha$ -MRE в этом домене присутствует еще два эритроид-специфичных энхансера, один из которых, расположенный сразу после  $\alpha^A$  гена [86], активирует экспрессию глобиновых генов взрослого типа, тогда как другой, расположенный в интроне гена TMEM8, обеспечивает интеграцию данного гена в регуляторную сеть домена  $\alpha$ -глобиновых генов [87]. Перед кластером  $\alpha$ -глобиновых генов кур расположен CpG-островок, в границах которого находятся участок начала репликации ДНК [88], промотор гена NPRL3 [89] и CTCF-зависимый сайленсер, стабилизирующий уровень экспрессии гена NPRL3 в эритроидных клетках [89]. В неэритроидных клетках часть данного CpG-островка метилируется, что, согласно одной из моделей, инициирует инактивацию кластера  $\alpha$ -глобиновых генов [90]. Любопытно, что между  $\alpha$ -MRE и кластером  $\alpha$ -глобиновых генов кур находятся полноценный инсулятор, обладающий как энхансер-блокирующей [91], так и барьерной [92] активностями. Биологическая роль данного инсулятора не понятна. Не вызывает сомнения то обстоятельство, что в эритроидных клетках  $\alpha$ -MRE стимулирует транскрипцию  $\alpha$ -глобиновых генов. Таким образом, по крайней мере в этих клетках энхансер-блокирующий элемент должен быть инактивирован. Нельзя исключить также и возможности того, что данный элемент, активность которого была продемонстрирована в составе конструкторов с репортерным геном [91], вообще не работает в геномном контексте. Что касается барьерного элемента, то его функция в домене  $\alpha$ -глобиновых генов еще менее понятна в силу того, что весь этот сегмент генома находится в активной хроматиновой конфигурации в клетках, дифференцированных по разным путям (см. выше). В этой связи уместно напомнить модель Диллона и Гросвельда, постулирующую, что в доменах открытого типа (функциональных доменах) инсуляторы вообще не выполняют каких-либо функций и в ходе эволюции сохраняются в тех позициях, где они просто ничему не мешают [18].

РЕГУЛЯЦИЯ ТРАНСКРИПЦИИ  $\alpha$ -ГЛОБИНОВЫХ ГЕНОВ

Как уже говорилось выше, в доменах закрытого типа активация транскрипции тканеспецифичных генов происходит поэтапно: сначала активируется хроматиновый домен (изменяется способ его упаковки в хроматине), потом стимулируется работа промоторов тех или иных генов, находящихся в данном домене [1]. В доменах открытого типа, в том числе и в доменах  $\alpha$ -глобиновых генов, регуляция транскрипции на уровне хроматинового домена, скорее всего, тоже происходит, хотя она и не связана с кардинальным изменением способа упаковки домена в хроматине. Действительно, домены  $\alpha$ -глобиновых генов находятся в открытой (ДНКазо-чувствительной) конфигурации как в эритроидных, так и в неэритроидных клетках [15]. Тем не менее, в эритроидных клетках происходят некие, менее кардинальные, изменения структуры хроматина в границах функционального домена  $\alpha$ -глобиновых генов. Об этом свидетельствует, в частности, существенное повышение уровня ацетилирования гистонов H3 и H4 по ряду специфических позиций [8]. Есть определенные указания на то, что в домене  $\alpha$ -глобиновых генов кур этот процесс контролируется  $\alpha$ -MRE посредством стимуляции низкоуровневой транскрипции всего домена  $\alpha$ -глобиновых генов [93]. Иными словами, в данном случае работают те же регуляторные принципы, что и при активации субдоменов домена  $\beta$ -глобиновых генов человека [68].

Ключевую роль в регуляции транскрипции  $\alpha$ -глобиновых генов играет адресная активация промоторов индивидуальных генов. Это достигается, в частности, посредством формирования активаторных хроматиновых блоков. Хорошо изучен процесс поэтапной сборки таких блоков в эритроидных клетках кур. Продемонстрировано, что уже в предшественниках эритробластов собирается неполноценный активаторный блок, содержащий  $\alpha$ -MRE, CpG-островок, расположенный перед кластером  $\alpha$ -глобиновых генов, и промотор  $\alpha^D$ -гена. В эритробластах взрослого типа, транскрибирующих гены  $\alpha^D$  и  $\alpha^A$ , к этому блоку присоединяются участок гиперчувствительности к ДНКазе I, расположенный на расстоянии 9 т.п.н. перед кластером  $\alpha$ -глобиновых генов (-9 DHS), и эритроид-специфичный энхансер, локализованный за кластером  $\alpha$ -глобиновых генов [94]. Прямые взаимодействия  $\alpha$ -MRE и ряда других эритроид-специфичных эритроидных элементов с глобиновыми генами были продемонстрированы и в доменах  $\alpha$ -глобиновых генов мыши и человека [95, 96]. Более того, было продемонстрировано, что ключевую роль в сборке активаторного комплекса играет  $\alpha$ -MRE. Удаление этого элемента приводило к дезинтеграции всего комплекса

одновременно с прекращением транскрипции  $\alpha$ -глобиновых генов. Реинсерция  $\alpha$ -MRE в другую геномную позицию в границах домена  $\alpha$ -глобиновых генов обеспечивала сборку нормального активаторного комплекса и восстановление транскрипции  $\alpha$ -глобиновых генов (рис. 4Б) [96]. На мышинной модели было сделано еще одно интересное наблюдение. Оказалось, что активация транскрипции  $\alpha$ -глобиновых генов коррелирует с их привлечением (в комплексе с эритроид-специфичными регуляторными элементами) к предсуществующей транскрипционной фабрике, осуществляющей транскрипцию генов домашнего хозяйства, локализованных в области геномной синтении перед кластером  $\alpha$ -глобиновых генов [62]. В домене  $\alpha$ -глобиновых генов кур, по-видимому, реализуется тот же сценарий. Об этом свидетельствует тот факт, что в эритроблестах промотор гена домашнего хозяйства NPRL3 прямо взаимодействует с  $\alpha$ -MRE и глобиновыми генами [94]. Помимо этого, промотор NPRL3 взаимодействует с промоторами других генов домашнего хозяйства, расположенных рядом с кластером  $\alpha$ -глобиновых генов [97].

Активация экспрессии  $\alpha$ -глобиновых генов по ходу развития контролируется рядом эритроид-специфичных транскрипционных факторов [98]. В мультипотентных гематопозитических клетках ключевую роль играет фактор GATA-2, места посадки которого служат центрами нуклеации, куда привлекаются транскрипционный фактор NF-E2 и так называемый пентамерный комплекс, включающий SCL. Начиная со стадии эритроидных предшественников GATA-2 вытесняется родственным фактором GATA-1, который образует комплекс с корегулятором FOG-1 [99] и связывается с целым рядом дополнительных регуляторных сайтов, включая и промоторы глобиновых генов. Подобно GATA-2, GATA-1 также привлекает NF-E2 и SCL. Места связывания GATA-1 колокализуются с большинством охарактеризованных участков гиперчувствительности к ДНКазе I [98], в том числе с теми, которые прямо участвуют в формировании единого активаторного комплекса домена  $\alpha$ -глобиновых генов [95]. В модельных экспериментах продемонстрировано, что посадка транскрипционных факторов на удаленные регуляторные элементы и промоторы  $\alpha$ -глобиновых генов осуществляется независимо. В то же время для привлечения РНК-полимеразы II (Pol II) к промоторам  $\alpha$ -глобиновых генов необходим  $\alpha$ -MRE. Pol II привлекается и к  $\alpha$ -MRE, причем для этого процесса присутствие промотора не является необходимым. В этой связи представляется вероятным предположение о том, что Pol II сначала привлекается к  $\alpha$ -MRE и затем тем или иным образом переносится на промоторы  $\alpha$ -глобиновых генов [95].

ПЕРЕКЛЮЧЕНИЕ ЭКСПРЕССИИ  $\alpha$ -ГЛОБИНОВЫХ ГЕНОВ

У позвоночных животных существует особый эмбриональный ген  $\alpha$ -типа ( $\zeta$  у человека и мыши,  $\pi$  у курицы), который экспрессируется на стадии желточного мешка, после чего инактивируется. Что касается глобиновых генов взрослого типа (как основного, так и минорных), то они экспрессируются на всех стадиях развития организма [100, 101]. Таким образом, вопрос о механизмах переключения экспрессии  $\alpha$ -глобиновых генов сводится к вопросу о способах инактивации эмбрионального гена. Эксперименты с трансфекцией различных конструкторов, содержащих ген  $\pi$  курицы, в эритроидные клетки раннеэмбрионального и взрослого типов [101], равно как и эксперименты по созданию трансгенных мышей, в геном которых был интегрирован ген  $\zeta$  человека вместе с фланкирующими его последовательностями, продемонстрировали, что все регуляторные элементы, необходимые для своевременной инактивации эмбриональных  $\alpha$ -глобиновых генов, располагаются в непосредственной близости от этих генов. Для инактивации гена  $\pi$  курицы необходимы последовательности ДНК, присутствующие в промоторной области (~350 п.н.) этого гена. Инактивация гена  $\pi$  коррелирует с метилированием промоторной области [102]. Инактивация гена  $\zeta$  человека по завершении раннеэмбриональной стадии развития также целиком определяется последовательностями, локализованными в непосредственной близости от этого гена [103]. Детальный анализ продемонстрировал, что для этого необходимы как последовательности, локализованные в промоторной области (~550 п.н. перед геном  $\zeta$ ), так и последовательности, локализованные в 3'-концевой фланкирующей области этого гена, где предположительно расположен сайленсер [100]. Вопрос о том, являются ли эмбриональные глобиновые гены полностью автономными (т.е. не нуждаются в активации удаленными регуляторными элементами) остается открытым. Анализ структуры активаторного хроматинового блока домена  $\alpha$ -глобиновых генов курицы продемонстрировал, что промотор гена  $\pi$  не взаимодействует ни с  $\alpha$ -MRE, ни с энхансерным элементом, расположенным на 3'-конце кластера  $\alpha$ -глобиновых генов [104]. С другой стороны, было продемонстрировано, что расположение гена  $\zeta$  первым от  $\alpha$ -MRE принципиально важно для нормальной экспрессии этого гена на раннеэмбриональной стадии развития. Перемещение гена  $\zeta$  в наиболее удаленную позицию посредством инверсии всего генного кластера относительно  $\alpha$ -MRE приводило к полной инактивации гена  $\zeta$  на всех стадиях развития [105].

#### ИНАКТИВАЦИЯ $\alpha$ -ГЛОБИНОВЫХ ГЕНОВ В НЕЭРИТРОИДНЫХ КЛЕТКАХ

В доменах закрытого типа, например в домене  $\beta$ -глобиновых генов, инаktivация генов в неэритроидных клетках достигается посредством упаковки домена в компактный (ДНКазо-устойчивый) хроматин. В домене  $\alpha$ -глобиновых генов и других доменах открытого типа данный механизм не используется. В то же время, проблема надежной репрессии глобиновых генов в неэритроидных клетках остается актуальной. Эта проблема могла бы быть решена посредством обеспечения высокой тканевой специфичности промоторов  $\alpha$ -глобиновых генов. Однако, промотор гена  $\alpha$  человека вообще не является тканеспецифичным в строгом смысле этого слова. Он расположен в CpG-островке и, по крайней мере, в модельных экспериментах с трансфекцией репортерного гена, поставленного под контроль данного промотора, проявляет высокую активность как в эритроидных, так и в неэритроидных клетках [106]. Каким образом инактивируются  $\alpha$ -глобиновые гены в неэритроидных клетках человека, долгое время оставалось неясным. Недавно было продемонстрировано, что в этом процессе участвуют репрессорные белки группы Polycomb [107]. Более того, оказалось, что именно метилированные CpG-островки участвуют в привлечении репрессорных комплексов группы Polycomb к промоторам  $\alpha$ -глобиновых генов человека [108].

Хорошо известно, что репрессорные комплексы Polycomb обеспечивают долговременную инаktivацию генов, экспрессирующихся на ранних стадиях развития, и являются важным элементом системы эпигенетической памяти. Участие белков группы Polycomb в репрессии тканеспецифичных  $\alpha$ -глобиновых генов явилось достаточно неожиданным. Будет интересно выяснить, участвуют ли эти белки в подавлении транскрипции других тканеспецифичных генов, расположенных в доменах открытого типа.

#### IV. ПРОИСХОЖДЕНИЕ И ЭВОЛЮЦИЯ ДОМЕНОВ РАЗНОГО ТИПА

Принципиально разная организация доменов  $\alpha$ - и  $\beta$ -глобиновых генов, наряду с жестко координированной экспрессией этих генов, делают особенно интересными вопросы о происхождении и закономерностях эволюции доменов разного типа. Один из подходов к анализу данного вопроса состоит в сравнительном изучении организации доменов  $\alpha$ - и  $\beta$ -глобиновых генов и их геномного окружения у разных таксономических групп. С помощью этого подхода можно понять насколько

«открытыми» для интеграции чужеродных генов являются оба типа доменов, решить вопрос о том, какой тип доменов – открытые (функциональные) или закрытые – первичен с эволюционной точки зрения. Наконец, данный подход позволяет получить представление о том, насколько координированной была эволюция собственно  $\alpha$ - и  $\beta$ -глобиновых генов и их геномного окружения.

Сравнительный анализ геномного окружения кластера  $\alpha$ -глобиновых генов у различных видов показал, что у кур участок, непосредственно соседствующий с  $\alpha$ -глобиновыми генами (с 3'-конца), претерпел хромосомную перестройку, в результате которой несколько генов ортологов, граничащих с кластером  $\alpha$ -глобиновых генов, оказались расположены в инвертированном порядке у курицы по сравнению с человеком и рядом других позвоночных. При этом у кур ген TMEM8 оказался приближен к кластеру  $\alpha$ -глобиновых генов, тогда как у человека он расположен на расстоянии 170 т.п.н. от него (рис. 5). Это перемещение коррелирует с существенным изменением характера экспрессии и структуры гена TMEM8. У человека этот ген кодирует трансмембранный белок, который экспрессируется в поджелудочной железе, плаценте, селезенке и лимфоцитах, где он является маркером их неактивного состояния. У кур ген TMEM8 экспрессируется предпочтительно в эритроидных клетках, в то время как в лимфоидных клетках он вообще не экспрессируется. Более того, оказалось, что индукция эритроидной дифференцировки сопровождается существенным увеличением уровня транскрипции гена TMEM8 [87]. Кардинальное изменение тканевой специфичности экспрессии TMEM8 коррелирует с приобретением им собственного эритроид-специфичного энхансера. Изучение конфигурации домена методом ЗС позволило заключить, что ген TMEM8 оказался интегрированным в регуляторную систему домена  $\alpha$ -глобиновых генов, сохранив при этом определенный уровень независимости [87]. Эти наблюдения свидетельствуют о том, что в геноме кур произошло расширение функциональных границ домена  $\alpha$ -глобиновых генов – по сравнению с геномом млекопитающих. К сожалению, в настоящее время невозможно сказать, почему произошла экспансия функционального домена после геномной перестройки, каковы причины приобретения геном TMEM8 эритроид-специфичного характера экспрессии, является ли это событие случайным или необходимым, первичным или вторичным по отношению к событию инверсии, и существовало ли давление отбора в направлении «эритроидспецифичности» TMEM8. Одно обстоятельство представляется немаловажным: экзоны, расположенные в конце гена, гораздо более консервативны в эволюции, чем экзоны, расположенные в начале, в «опасной бли-



зости» от кластера  $\alpha$ -глобиновых генов. Иными словами, эволюционное давление в направлении изменения гена TMEM8 было тем выше, чем ближе расположен экзон к  $\alpha$ -глобиновым генам [87].

Геномный контекст кластера  $\beta$ -глобиновых генов у млекопитающих и птиц также различен (рис. 5). У млекопитающих кластер  $\beta$ -глобиновых генов окружен генами обонятельных рецепторов (OR). У птиц гены обонятельных рецепторов фланкируют  $\beta$ -глобиновые гены с одной стороны, а с другой стороны расположен ген фолатного рецептора, FOLR1, который отделен от кластера глобиновых генов участком конденсированного гетерохроматина [14]. У человека ген FOLR1 находится на другом плече 11 хромосомы, т.е. удален от кластера  $\beta$ -глобиновых генов. FOLR1 кодирует рецептор фолиевой кислоты и у человека, согласно базе данных UCSC, экспрессируется исключительно в тканях эпителиального происхождения. У кур ген фолатного рецептора проявляет эритроидную специфичность и экспрессируется на ранних стадиях эритроидной дифференцировки (образование CFU-E), предшествующих экспрессии глобиновых генов. Кластер куриных  $\beta$ -глобиновых генов и ген фолатного рецептора разделены инсулятором, регулируются независимо и являются компонентами разных геномных доменов [24]. Более того, фрагмент ДНК протяженностью 3 т.п.н, расположенный перед геном FOLR1 содержит все регуляторные элементы, необходимые для корректной экспрессии этого гена на ранней стадии эритроидной дифференцировки (CFU-E).

Таким образом, сближение в результате транслокаций гена TMEM8 и домена  $\alpha$ -глобиновых генов кур, с одной стороны, и сближение гена FOLR1 с доменом  $\beta$ -глобиновых кур, с другой стороны, коррелировали с приобретением этими генами эритроид-специфичного характера экспрессии. При этом, сближение гена TMEM8 с эритроид-специфичным доменом открытого типа (домен  $\alpha$ -глобиновых генов) обусловило интеграцию данного гена в состав регуляторных сетей домена  $\alpha$ -глобиновых генов, тогда как ген FOLR1, который оказался расположен за обозначенной инсулятором границей домена  $\beta$ -глобиновых генов, относящегося к числу доменов закрытого типа, сохранил независимость и регуляторную самостоятельность.

Какой тип доменов – открытого или закрытого типа – является эволюционно более древним, и насколько правомерна такая постановка вопроса? В настоящее время считается общепринятым, что функциональная сложность и разнообразие современных геномов возникла как в результате эволюции внутри простых геномов предковых организмов, так и в результате многочисленных вертикальных и горизон-

тальных переносов генов, фрагментов геномов и целых геномов. Предполагается, что предок глобиновых генов появился в геноме древних одноклеточных эукариот в результате горизонтального переноса из бактериальных геномов, одновременно с появлением эндосимбиотических митохондрий и пластид [109]. Временное сохранение структурной и функциональной обособленности генов-интервентов, очевидно, необходимо, как единственная разумная стратегия выживания нового гена в агрессивной среде чужеродного генома. С этой точки зрения домены закрытого типа, изолированные от геномного окружения, самодостаточные и независимые с точки зрения регуляции, аналогичны таким генам-интервентам сразу после их внедрения в геном хозяина.

Ключевым этапом в эволюции геномов хордовых явились два последовательных цикла полногеномной дупликации. Резко возросшая избыточность генома явилась топливом для эволюционного процесса, в результате чего появилась возможность диверсификации и специализации многих физиологических функций. В частности, возникли специализированные системы, осуществляющие транспорт и хранение кислорода (миоглобин, цитоглобин и гемоглобин). Четырехкратная избыточность генома в протогеноме организма, появившегося в результате полногеномных дупликаций, может быть прослежена и у современных организмов. Геномное окружение кластера  $\alpha$ -глобиновых генов, генов миоглобина и цитоглобина (паралогон [110]) обнаруживает некий общий план, черты макросинтезии [110]. Разумеется, многочисленные транслокации и дупликации в ходе эволюционного процесса размывают эту макросинтезию. Естественный отбор редактирует генный комплекс, в результате чего гены, которые хорошо кооперируются, образуют тесно сцепленную группу. Очевидно, что наряду с исходной функциональной изоляцией происходят и те или иные интеграционные процессы внутри геномов, сопровождающиеся приобретением определенных преимуществ при сохранении исходной функции. Возможно, и домены открытого типа, и транскрипционные фабрики, образованные координатно расположенными генами, являются продуктом таких интеграционных процессов. Хочется еще раз напомнить о том, что внутри открытого домена  $\alpha$ -глобиновых генов кур были обнаружены неактивные инсультатор-подобные элементы, по-видимому, утратившие свои функции в ходе эволюции. В рамках этой логики более древними представляются домены закрытого типа.

Реконструированная эволюция  $\alpha$ - и  $\beta$ -глобиновых генов, их регуляторных элементов и доменов в целом является неисчерпаемым

источником для различных гипотез, так как последовательность ДНК глобинового кластера и его геномного окружения определена у многих организмов. Сравнивая структуру генов, состав  $\alpha$ - и  $\beta$ -глобиновых кластеров и соответствующий геномный контекст у разных групп позвоночных, можно проследить эволюционную историю  $\alpha$ - и  $\beta$ -глобиновых генов до общего предка рыб, земноводных и высших позвоночных (рис. 5). В геноме земноводных (*Xenopus laevis*, *X. tropicalis*) и большинства рыб (*Danio rerio*, *Oryzias latipes*, *Salmo salar*)  $\alpha$ - и  $\beta$ -глобиновые гены не разделены, а располагаются рядом в направлении 5'- $\alpha$ - $\beta$ -3' [111–113] и, видимо, представляют собой единый геномный домен. В геноме тропической рыбки-зебры (*Danio rerio*) гены, кодирующие молекулы гемоглобина, находятся на двух хромосомах: 3 и 12. На 12-ой хромосоме расположен кластер эмбриональных  $\alpha$ - и  $\beta$ -глобиновых генов, продукты экспрессии которых обнаружены только на уровне мРНК [114]. На 3-й хромосоме *Danio rerio* в непосредственной близости расположены эмбриональный и взрослый кластеры  $\alpha$ - и  $\beta$ -глобиновых генов [114]. В обоих кластерах генов на 3-ей хромосоме транскрипция  $\alpha$ - и  $\beta$ -глобиновых генов происходит в противоположных направлениях. Анализ геномного окружения кластера и поиск консервативных последовательностей у различных организмов позволил картировать ряд возможных эритроидспецифичных регуляторных элементов. Перед кластером  $\alpha$ -глобиновых генов, вне зависимости от того, является ли он изолированным, или объединен с кластером  $\beta$ -глобиновых генов, расположен консервативный ген «домашнего хозяйства» NPRL3, содержащий  $\alpha$ -MRE. Таким образом, связка кластера глобиновых генов и гена NPRL3 сохраняется у всех позвоночных и представляется абсолютно необходимой (рис. 5). Регуляторный элемент, гомологичный  $\alpha$ -MRE человека, у рыб является энхансером, активирующим работу как  $\alpha$ -, так и  $\beta$ -глобиновых генов, а также переключателем экспрессии с эмбрионального на взрослый тип [111].

К какому типу доменов (открытому или закрытому) относится объединенный домен  $\alpha$ - и  $\beta$ -глобиновых генов рыб и земноводных, в настоящий момент неизвестно.

Предполагается, что  $\alpha$ - и  $\beta$ -глобиновые гены произошли в результате дубликации предкового гена, кодирующего белок-транспортер кислорода. Эти гены располагались рядом у общего предка высших позвоночных и сохранили такое расположение у рыб и земноводных (рис. 5). Дивергенция  $\alpha$ - и  $\beta$ -глобиновых генов была, по всей видимости, обусловлена необходимостью усложнения молекулы гемоглобина. Гены, кодирующие  $\alpha$ - и  $\beta$ -глобиновые цепи,

оказались на разных хромосомах либо в результате транслокаций, либо в результате хромосомных/геномных и *in trans* дупликаций и последующей инактивации генов [115]. Дальнейшая эволюция и  $\alpha$ - и  $\beta$ -глобиновых генов у высших позвоночных сопровождалась неоднократными дупликациями генов внутри кластеров и их последующей специализацией. Согласно одной из гипотез, кластеры  $\alpha$ - и  $\beta$ -глобиновых генов птиц и млекопитающих эволюционировали независимо после дупликации кластера у их общего предка [116]. Так, предполагается, что перед расхождением птиц и млекопитающих участок хромосомы, несущий  $\alpha$ - $\beta$ -глобиновый кластер, дублировался и дал начало двум  $\alpha$ - $\beta$ -глобиновым кластерам на разных хромосомах. Далее эти кластеры претерпевали эволюцию у птиц и млекопитающих независимо, посредством инактивации либо  $\alpha$ -, либо  $\beta$ -глобиновой составляющей кластеров. Согласно другой гипотезе, изоляция  $\beta$ -глобинового кластера произошла путем перемещения копии  $\beta$ -глобинового гена у предка высших позвоночных на другую хромосому [117]. Удивительным образом современные «живые ископаемые» сохраняют архаические черты организации домена  $\alpha$ -глобиновых генов. Так, в геноме утконоса кластер  $\alpha$ -глобиновых генов содержит не только эмбриональный и взрослый гены, но и  $\beta$ -подобный ген омега-глобина ( $\omega$ ), а также ген GB $\gamma$ , гомологичный цитоглобину, который обособился структурно и функционально задолго до разделения единого  $\alpha$ - $\beta$ -глобинового домена [117, 118].

Так или иначе, в ходе эволюции произошла экспансия, функциональная дивергенция и вторичная утрата продуктов дупликаций:  $\beta$ -глобиновые гены переместились на другую хромосому и оказались в окружении генов обонятельных рецепторов, также обладающих выраженной тканеспецифичностью. Сейчас невозможно выяснить, что явилось причиной изоляции  $\beta$ -глобиновых генов и формирования закрытого домена – транслокация неких барьерных элементов предкового объединенного домена одновременно с  $\beta$ -глобиновым геном или давление эволюции в направлении появления пограничных элементов *de novo*. Последнее представляется маловероятным. В любом случае, исчезновение барьерных элементов из окружения оставшихся  $\alpha$ -глобиновых генов (если они там присутствовали и предковый  $\alpha$ - $\beta$ -глобиновый домен относился к категории закрытых) по какой-то причине оказалось эволюционно выгодным. Результатом транслокации явилось формирование открытого домена, интегрированного с геномным окружением наличием эритроидспецифичных удаленных регуляторных элементов, образующих с этим окружением общую транскрипционную фабрику и открытого для интервенции новых генов (TMEM8).

## V. ЗАКЛЮЧЕНИЕ

В настоящее время не вполне ясно, какие домены (открытого или закрытого типов) являются наиболее типичными в геноме современных позвоночных животных. Не вызывает сомнения тот факт, что исследования, выполненные на домене  $\beta$ -глобиновых генов и ряде других доменов закрытого типа, сыграли исключительную роль в изучении механизмов регуляции транскрипции у высших эукариот. Одновременно это привело к определенной переоценке значения таких доменов. В последнее время появилось немало примеров того, что даже тканеспецифичные гены далеко не всегда сконцентрированы в доменах закрытого типа. Более того, оказалось, что некоторые геномные домены, которые рассматривались, как типичные примеры доменов закрытого типа (например, домен гена лизоцима курицы [119]), содержат ген(ы) домашнего хозяйства, которые постоянно экспрессируются во всех типах клеток. Таким образом, структурно-функциональная организация генома на деле является существенно более сложной, чем это описывает даже адаптированная к существованию доменов открытого типа гипотеза доменной организации генома. Это наглядно иллюстрирует тот факт, что в природе нет четких разграничительных граней. Понять принципы работы геномных доменов, которые не могут быть отнесены к открытому, либо закрытому типу, а представляют некий более сложный тип организации, – задача будущих научных исследований.

## ЛИТЕРАТУРА

1. Razin, S.V., Iarovaia, O.V., Sjakste, N., Sjakste, T., Bagdoniene, L., Rynditch, A.V., Eivazova, E.R., Lipinski, M., Vassetzky, Y.S. (2007) *J. Mol. Biol.*, **369**, 597–607.
2. Bodnar, J.W. (1988) *J. Theor. Biol.*, **132**, 479–507.
3. Goldman, M.A. (1988) *Bioessays*, **9**, 50–55.
4. Weintraub, H., Groudine, M. (1976) *Science*, **73**, 848–856.
5. Forrester, W.C., Thompson, C., Elder, J.T., Groudine, M. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1359–1363.
6. Lawson, G.M., Knoll, B.J., March, C.J., Woo, S.L.C., Tsai, M.-J., O'Malley, B.W. (1982) *J. Biol. Chem.*, **257**, 1501–1507.
7. Krajewski, W.A., Becker, P.B. (1998) *Proc. Natl. Acad. Sci. USA*, **95**, 1540–1545.
8. Anguita, E., Johnson, C.A., Wood, W.G., Turner, B.M., Higgs, D.R. (2001) *Proc. Natl. Acad. Sci. USA*, **98**, 12114–12119.
9. Kellum, R., Schedl, P. (1992) *Mol. Cell. Biol.*, **12**, 2424–2431.
10. Felsenfeld, G., Burgess-Beusse, B., Farrell, C., Gaszner, M., Ghirlando, R., Huang, S., Jin, C., Litt, M., Magdinier, F., Mutskov, V., Nakatani, Y., Tagami, H., West, A., Yusufzai, T. (2004) *Cold Spring Harb. Symp. Quant. Biol.*, **69**, 245–250.
11. Wallace, J.A., Felsenfeld, G. (2007) *Curr. Opin. Genet. Dev.*, **17**, 400–407.

12. Forrester, W.C., Epner, E., Driscoll, M.C., Enver, T., Brice, M., Papayan-nopoulou, T., Groudine, M. (1990) *Genes Dev.*, **4**, 1637–1649.
13. Grosveld, F., van Assandelt, G.B., Greaves, D.R., Kollias, B. (1987) *Cell*, **51**, 975–985.
14. Razin, S.V., Farrell, C.M., Recillas-Targa, F. (2003) *Int. Rev. Cytol.*, **226**, 63–125.
15. Craddock, C.F., Vyas, P., Sharpe, J.A., Ayyub, H., Wood, W.G., Higgs, D.R. (1995) *EMBO J.*, **14**, 1718–1726.
16. Flint, J., Tufarelli, C., Peden, J., Clark, K., Daniels, R.J., Hardison, R., Miller, W., Philipsen, S., Tan-Un, K.C., McMorrow, T., Frampton, J., Alter, B.P., Frischauf, A.M., Higgs, D.R. (2001) *Hum. Mol. Genet.*, **10**, 371–382.
17. Makalowska, I., Lin, C.F., Makalowski, W. (2005) *Comput. Biol. Chem.*, **29**, 1–12.
18. Dillon, N., Sabbatini, P. (2000) *Bioessays*, **22**, 657–665.
19. Forget, B.G. (1998) *Ann. N.Y. Acad. Sci.*, **850**, 38–44.
20. Bulger, M., Bender, M., Farrell, C., Felsenfeld, G., Wertman, B., Groudine, M., Hardison, R. (2000) *Blood Cells Mol. Dis.*, **26**, 490.
21. Schubeler, D., Francastel, C., Cimbora, D.M., Reik, A., Martin, D.I.K., Groudine, M. (2000) *Genes Dev.*, **14**, 940–950.
22. Hanscombe, O., Whyatt, D., Fraser, P., Yannoutsos, N., Greaves, D., Dillon, N., Grosveld, F. (1991) *Genes Dev.*, **5**, 1387–1394.
23. Hebbes, T.R., Clayton, A.L., Thorne, A.W., Crane-Robinson, C. (1994) *EMBO J.*, **13**, 1823–1830.
24. Prioleau, M.-N., Nony, P., Simpson, M., Felsenfeld, G. (1999) *EMBO J.*, **18**, 4035–4048.
25. Guerrero, G., Delgado-Olguin, P., Escamilla-Del-Arenal, M., Furlan-Magaril, M., Rebollar, E., De La Rosa-Velazquez, I.A., Soto-Reyes, E., Rincon-Arano, H., Valdes-Quezada, C., Valadez-Graham, V., Recillas-Targa, F. (2007) *Comp. Biochem. Physiol. A. Mol. Integr. Physiol.*, **147**, 750–760.
26. Geyer, P.K. (1997) *Curr. Opin. Genet. Dev.*, **7**, 242–248.
27. West, A.G., Gaszner, M., Felsenfeld, G. (2002) *Genes Dev.*, **16**, 271–288.
28. Chung, J.H., Whiteley, M., Felsenfeld, G. (1993) *Cell*, **74**, 505–514.
29. Recillas-Targa, F., Pikaart, M.J., Burgess-Beusse, B., Bell, A.C., Litt, M.D., West, A.G., Gaszner, M., Felsenfeld, G. (2002) *Proc. Natl. Acad. Sci. USA*, **99**, 6883–6888.
30. Chung, J.H., Bell, A.C., Felsenfeld, G. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 575–580.
31. Bell, A.C., West, A.G., Felsenfeld, G. (1999) *Cell*, **98**, 387–396.
32. Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., Yahata, K., Imamoto, F., Aburatani, H., Nakao, M., Imamoto, N., Maeshima, K., Shirahige, K., Peters, J.M. (2008) *Nature*, **451**, 796–801.
33. Xiao, T., Wallace, J., Felsenfeld, G. (2011) *Mol. Cell Biol.*, **31**, 2174–2183.
34. Huang, S., Li, X., Yusufzai, T.M., Qiu, Y., Felsenfeld, G. (2007) *Mol. Cell Biol.*, **27**, 7991–8002.
35. Li, X., Wang, S., Li, Y., Deng, C., Steiner, L.A., Xiao, H., Wu, C., Bungert, J., Gallagher, P.G., Felsenfeld, G., Qiu, Y., Huang, S. (2011) *Blood*, **118**, 1386–1394.
36. Saitoh, N., Bell, A.C., Recillas-Targa, F., West, A.G., Simpson, M., Pikaart, M., Felsenfeld, G. (2000) *EMBO J.*, **19**, 2315–2322.
37. Farrell, C.M., West, A.G., Felsenfeld, G. (2002) *Mol. Cell Biol.*, **22**, 3820–3831.
38. Li, Q., Stamatoyannopoulos, G. (1994) *Blood*, **84**, 1399–1401.

39. Tanimoto, K., Liu, Q., Bungert, J., Engel, J.D., (1999) *Nature*, **398**, 344–348.
40. Li, Q., Zhang, M., Han, H., Rohde, A., Stamatoyannopoulos, G. (2002) *Nucleic Acids Res.*, **30**, 2484–2491.
41. Talbot, D., Collis, P., Antoniou, M., Vidal, M., Grosveld, F., Greaves, D.R. (1989) *Nature*, **338**, 352–355.
42. Li, G., Lim, K.C., Engel, J.D., Bungert, J. (1998) *Genes Cells*, **3**, 415–429.
43. Cimbora, D.M., Schubeler, D., Reik, A., Hamilton, J., Francastel, C., Epner, E.M., Groudine, M. (2000) *Mol. Cell. Biol.*, **20**, 5581–5591.
44. Philipsen, S., Talbot, D., Fraser, P., Grosveld, F. (1990) *EMBO J.*, **9**, 2159–2167.
45. Ney, P.A., Sorrentino, B.P., McDonagh, K.T., Nienhuis, A.W. (1990) *Genes Dev.*, **4**, 993–1006.
46. Strauss, E.C., Orkin, S.H. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 5809–5813.
47. Lowrey, C.H., Bodine, D.M., Nienhuis, A.W. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 1143–1147.
48. Stamatoyannopoulos, G., Goodwin, A., Joyce, T., Lowrey, C.H. (1995) *EMBO J.*, **14**, 106–116.
49. Li, Q., Zhou, B., Powers, P., Enver, T., Stamatoyannopoulos, G. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 8207–8211.
50. Mason, M.M., Lee, E., Westphal, H., Reitman, M. (1995) *Mol. Cell. Biol.*, **15**, 407–414.
51. Hardison, R., Slightom, J.L., Gumicio, D.L., Goodman, M., Stojanovic, N., Miller, W. (1997) *Gene*, **205**, 73–94.
52. Bulger, M., Groudine, M. (1999) *Genes Dev.*, **13**, 2465–2477.
53. Peterson, K.R., Stamatoyannopoulos, G. (1993) *Mol. Cell. Biol.*, **13**, 4836–4843.
54. Dillon, N., Trimborn, T., Stronboulis, J., Fraser, P., Grosveld, F. (1997) *Mol. Cell*, **1**, 131–139.
55. Wijgerde, M., Grosveld, F., Fraser, P. (1995) *Nature*, **377**, 209–213.
56. Gribnau, J., de Boer, E., Trimborn, T., Wijgerde, M., Milot, E., Grosveld, F., Fraser, P. (1998) *EMBO J.*, **17**, 6020–6027.
57. Dekker, J., Rippe, K., Dekker, M., Kleckner, N. (2002) *Science*, **295**, 1306–1311.
58. de Laat, W., Grosveld, F. (2003) *Chromosome Res.*, **11**, 447–459.
59. Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., de Laat, W. (2002) *Mol. Cell*, **10**, 1453–1465.
60. West, A.G., Fraser, P. (2005) *Hum. Mol. Genet.*, **14**, 101–111.
61. de Laat, W., Klous, P., Kooren, J., Noordermeer, D., Palstra, R.J., Simonis, M., Splinter, E., Grosveld, F. (2008) *Curr. Top. Dev. Biol.*, **82**, 117–139.
62. Zhou, G.L., Xin, L., Song, W., Di, L.J., Liu, G., Wu, X.S., Liu, D.P., Liang, C.C. (2006) *Mol. Cell. Biol.*, **26**, 5096–5105.
63. Kooren, J., Palstra, R.J., Klous, P., Splinter, E., von Lindern, M., Grosveld, F., de Laat, W. (2007) *J. Biol. Chem.*, **282**, 16544–16552.
64. Armstrong, J.A., Bieker, J.J., Chen, X. (1998) *Cell*, **95**, 93–104.
65. Blobel, G. (2000) *Blood*, **95**, 745–755.
66. Bresnick, E.H., Tze, L. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 4566–4571.
67. Forsberg, E.C., Johnson, K., Zaboikina, T.N., Mosser, E.A., Bresnick, E.H. (1999) *J. Biol. Chem.*, **274**, 26850–26859.
68. Gribnau, G., Diderich, K., Pruzina, S., Calzolari, R., Frazer, P. (2000) *Mol. Cell*, **5**, 377–386.
69. Wilson, C.J., Chao, D.M., Imbalzano, A.N., Schnitzler, G.R., Kingston, R.E., Yong, R.A. (1996) *Cell*, **84**, 235–244.
70. Cho, H., Orphanides, G., Sun, X., Yang, X.J., Ogryzko, V., Lees, E., Nakatani, Y., Reinberg, D. (1998) *Mol. Cell. Biol.*, **18**, 5355–5363.

71. Travers, A. (1999) Proc. Natl. Acad. Sci. USA, **96**, 13634–13637.
72. Herendeen, D.R., Kassavetis, G.A., Geiduschek, E.P. (1992) Science, **256**, 1298–1303.
73. Forsberg, E.C., Bresnick, E.H. (2001) Bioessays, **23**, 820–830.
74. Hughes, J.R., Cheng, J.F., Ventress, N., Prabhakar, S., Clark, K., Anguita, E., De Gobbi, M., de Jong, P., Rubin, E., Higgs, D.R. (2005) Proc. Natl. Acad. Sci. USA, **102**, 9830–9835.
75. Tufarelli, C., Hardison, R., Miller, W., Hughes, J., Clark, K., Ventress, N., Frischauf, A.M., Higgs, D.R. (2004) Genome Res., **14**, 623–630.
76. Klochkov, D.B., Gavrilov, A.A., Vassetzky, Y.S., Razin, S.V. (2009) Genomics, **93**, 481–486.
77. Tufarelli, C., Frischauf, A.M., Hardison, R., Flint, J., Higgs, D.R. (2001) Genomics, **71**, 307–314.
78. Neklesa, T.K., Davis, R.W. (2009) PLoS Genet., **5**, e1000515.
79. Higgs, D.R., Wood, W.G., Jarman, A.P., Sharpe, J., Lida, J., Pretorius, I.-M., Ayyub, H. (1990) Genes Dev., **4**, 1588–1601.
80. Viprakit, V., Harteveld, C.L., Ayyub, H., Stanley, J.S., Giordano, P.C., Wood, W.G., Higgs, D.R. (2006) Blood, **107**, 3811–3812.
81. Jarman, A.P., Wood, W.G., Sharpe, J.A., Gourdon, G., Ayyub, H., Higgs, D.R. (1991) Mol. Cell. Biol., **11**, 4679–4689.
82. Anguita, E., Sharpe, J.A., Sloane-Stanley, J.A., Tufarelli, C., Higgs, D.R., Wood, W.G. (2002) Blood, **100**, 3450–3456.
83. Chen, H., Lowrey, C.H., Stamatoyannopoulos, G. (1997) Nucleic Acids Res., **25**, 2917–2922.
84. De Gobbi, M., Anguita, E., Hughes, J., Sloane-Stanley, J.A., Sharpe, J.A., Koch, C.M., Dunham, I., Gibbons, R.J., Wood, W.G., Higgs, D.R. (2007) Blood, **110**, 4503–4510.
85. Igarashi, K., Itoh, K., Hayashi, N., Nishizawa, M., Yamamoto, M. (1995) Proc. Natl. Acad. Sci. USA, **92**, 7445–7449.
86. Knezetic, J., Felsenfeld, G. (1989) Mol. Cell. Biol., **9**, 893–901.
87. Philonenko, E.S., Klochkov, D.B., Borunova, V.V., Gavrilov, A.A., Razin, S.V., Iarovaia, O.V. (2009) Nucleic Acids Res., **37**, 7394–7406.
88. Razin, S.V., Kekelidze, M.G., Lukandin, E.M., Scherrer, K., Georgiev, G.P. (1986) Nucleic Acids Res., **14**, 8189–8207.
89. Klochkov, D., Rincon-Arango, H., Ioudinkova, E.S., Valadez-Graham, V., Gavrilov, A., Recillas-Targa, F., Razin, S.V. (2006) Mol. Cell. Biol., **26**, 1589–1597.
90. Razin, S.V., Ioudinkova, E.S., Scherrer, K. (2000) J. Mol. Biol., **209**, 845–852.
91. Valadez-Graham, V., Razin, S.V., Recillas-Targa, F. (2004) Nucl. Acids Res., **32**, 1354–1362.
92. Furlan-Magaril, M., Rebollar, E., Guerrero, G., Fernandez, A., Molto, E., Gonzalez-Buendia, E., Cantero, M., Montoliu, L., Recillas-Targa, F. (2011) Nucleic Acids Res., **39**, 89–103.
93. Razin, S.V., Rynditch, A., Borunova, V., Ioudinkova, E., Smalko, V., Scherrer, K. (2004) J. Cell. Biochem., **92**, 445–457.
94. Gavrilov, A.A., Razin, S.V. (2008) Nucleic Acids Res., **36**, 4629–4640.
95. Vernimmen, D., De Gobbi, M., Sloane-Stanley, J.A., Wood, W.G., Higgs, D.R. (2007) EMBO J., **26**, 2041–2051.
96. Vernimmen, D., Marques-Kranc, F., Sharpe, J.A., Sloane-Stanley, J.A., Wood, W.G., Wallace, H.A., Smith, A.J., Higgs, D.R. (2009) Blood, **114**, 4253–4260.
97. Gavrilov, A.A., Zukher, I.S., Philonenko, E.S., Razin, S.V., Iarovaia, O.V. (2010) Nucleic Acids Res., **38**, 8051–8060.

98. Anguita, E., Hughes, J., Heyworth, C., Blobel, G.A., Wood, W.G., Higgs, D.R. (2004) *EMBO J.*, **23**, 2841–2852.
99. Pal, S., Cantor, A.B., Johnson, K.D., Moran, T.B., Boyer, M.E., Orkin, S.H., Bresnick, E.H. (2004) *Proc. Natl. Acad. Sci. USA*, **101**, 980–985.
100. Liebhaber, S.A., Wang, Z., Cash, F.E., Monks, B., Russell, J.E. (1996) *Mol. Cell. Biol.*, **16**, 2637–2646.
101. Knezetic, J.A., Felsenfeld, G. (1993) *Mol. Cell. Biol.*, **13**, 4632–4639.
102. Singal, R., vanWert, J.M., Ferdinand, L. Jr. (2002) *Blood*, **100**, 4217–4222.
103. Sabath, D.E., Spangler, E.A., Rubin, E.M., Stamatoyannopoulos, G. (1993) *Blood*, **82**, 2899–2905.
104. Ioudinkova, E.S., Ulianov, S.V., Bunina, D., Iarovaia, O.V., Gavrilov, A.A., Razin, S.V. (2011) *Epigenetics*, **6**, 1481–1488.
105. Tang, Y., Wang, Z., Huang, Y., Liu, D.P., Liu, G., Shen, W., Tang, X., Feng, D., Liang, C.C. (2006) *Genes Cells*, **11**, 123–131.
106. Whitelaw, E., Hogben, P., Hanscombe, O., and Proudfoot, N. J. (1989) *Mol. Cell. Biol.*, **9**, 241–251.
107. Garrick, D., De Gobbi, M., Samara, V., Rugless, M., Holland, M., Ayyub, H., Lower, K., Sloane-Stanley, J., Gray, N., Koch, C., Dunham, I., Higgs, D.R. (2008) *Blood*, **112**, 3889–3899.
108. Lynch, M.D., Smith, A.J., De Gobbi, M., Flenley, M., Hughes, J.R., Vernimmen, D., Ayyub, H., Sharpe, J.A., Sloane-Stanley, J.A., Sutherland, L., Meek, S., Burdon, T., Gibbons, R.J., Garrick, D., Higgs, D.R. (2011) *EMBO J.*, Nov 4. doi: 10.1038/emboj.2011.399.
109. Vinogradov, S.N., Hoogewijs, D., Bailly, X., Arredondo-Peter, R., Gough, J., Dewilde, S., Moens, L., Vanfleteren, J.R. (2006) *BMC Evol. Biol.*, **6**, 31.
110. Hoffmann, F.G., Opazo, J.C., Storz, J.F. *Mol. Biol. Evol.*, in press.
111. Maruyama, K., Yasumasu, S., Naruse, K., Mitani, H., Shima, A., Iuchi, I. (2004) *Gene*, **335**, 89–100.
112. Fuchs, C., Burmester, T., Hankeln, T. (2006) *Cytogenet. Genome Res.*, **112**, 296–306.
113. Chan, F.Y., Robinson, J., Brownlie, A., Shivdasani, R.A., Donovan, A., Brugnara, C., Kim, J., Lau, B.C., Witkowska, H.E., Zon, L.I. (1997) *Blood*, **89**, 688–700.
114. Brownlie, A., Hersey, C., Oates, A.C., Paw, B.H., Falick, A.M., Witkowska, H.E., Flint, J., Higgs, D., Jessen, J., Bahary, N., Zhu, H., Lin, S., Zon, L. (2003) *Dev. Biol.*, **255**, 48–61.
115. Jeffreys, A.J., Wilson, V., Wood, D., Simons, J.P., Kay, R.M., Williams, J.G. (1980) *Cell*, **21**, 555–564.
116. Wheeler, D., Hope, R.M., Cooper, S.J., Gooley, A.A., Holland, R.A. (2004) *J. Mol. Evol.*, **58**, 642–652.
117. Wheeler, D., Hope, R., Cooper, S.B., Dolman, G., Webb, G.C., Bottema, C.D., Gooley, A.A., Goodman, M., Holland, R.A. (2001) *Proc. Natl. Acad. Sci. USA*, **98**, 1101–6.
118. Patel, V.S., Cooper, S.J., Deakin, J.E., Fulton, B., Graves, T., Warren, W.C., Wilson, R.K., Graves, J.A. (2008) *BMC Biol.*, **6**, 34.
119. Chong, S., Riggs, A.D., Bonifer, C. (2002) *Nucleic Acids Res.*, **30**, 463–467.