

ПАРАДОКСЫ ОРГАНИЗАЦИИ ЦЕНТРОМЕРА И ГЕТЕРОХРОМАТИНА

© О. И. Подгорная, Д. И. Остромышенский, И. С. Кузнецова, И. В. Матвеев, А. С. Комиссаров

Институт цитологии РАН, Санкт-Петербург;
электронный адрес: aleksey.komissarov@gmail.com

Центромера (CEN) — это структура, обеспечивающая удержание хроматид, правильность выстраивания хромосом в метафазной пластинке и прикрепления к веретену. Известно, что у млекопитающих ДНК CEN относится к группе сателлитных ДНК (сатДНК) — тандемно организованных высокоповторяющихся последовательностей. Растет количество свидетельств того, что CEN- и пери-CEN-хроматин необходимы для нормального CEN. CEN и прителомерные районы остаются «белыми пятнами» на картах хромосом, появившихся в результате чтения геномов. «Библиотечная» гипотеза предполагает, что в гетерохроматине каждого вида (рода) есть фрагменты разных сатДНК, но при фиксации вида происходит избирательное размножение фрагментов, что обеспечивает видоспецифичность сатДНК. Анализ базы данных Chromosome Unknown (ChrUn) привел к предсказанию *in silico* новых классов тандемных повторов мыши. Классы семейств тандемных повторов характерны не только для генома мыши, но и для генома крысы, так же как сходно распределение повторов по содержанию GC. Выявление общих структурных особенностей фрагментов CEN и пери-CEN ДНК позволит разрешить парадокс, заключающийся в том, что наборы белков, специфичных для CEN-района, связываются с абсолютно, казалось бы, несходными последовательностями.

Ключевые слова: гетерохроматин, центромера, перицентромер, сателлитная ДНК, базы данных, геном, Chromosome Unknown, кривизна ДНК.

Принятые сокращения: пери-CEN — перицентромер, сатДНК — сателлитная ДНК, CEN — центромера, ЯМ — ядерный матрикс, ChrUn — Chromosome Unknown, неизвестная хромосома.

Центромеры (CEN) большинства эукариот состоят из сателлитной ДНК (сатДНК). CEN прямо вовлечены в процессы, ведущие к анеуплоидии и полипloidии. Изменение числа хромосом (дозы) часто связано с нарушением баланса генной экспрессии и свойственно многим типам рака. Изучение CEN-района движется медленно из-за трудностей клонирования высокоповторяющихся последовательностей, несмотря на понимание того, как важно хорошо представлять себе процесс нормальной сегрегации хромосом.

CEN — это структура, обеспечивающая удержание хромосом, правильность выстраивания хромосом в метафазной пластинке и их прикрепления к веретену; участок, ответственный за контроль наступления анафазы. Известно, что у млекопитающих ДНК CEN относится к группе сателлитных ДНК (сатДНК) — тандемно организованных высокоповторяющихся последовательностей. Неизвестно, какая именно последовательность определяет район формирования кинетохора. Популярна гипотеза о так называемом эпигенетическом определении положения кинетохора на хромосоме. Гипотеза предполагает, что центромеризация участка хромосомы определяется не его первичной структурой, а модификациями на уровне хроматина. Но как происходит первичное маркирование CEN, до сих пор неизвестно. В последние годы достигнут значительный прогресс в описании кинетохорных белков (Durand-Dubief, Ekwall, 2008), но отсутствие первичной последовательности ДНК этой области затрудняет как те-

оретические исследования, так и работу по конструированию искусственных хромосом.

История открытия и изучения сатДНК насчитывает более 50 лет. Однако до сих пор функции сатДНК не очевидны. В последние годы этот вопрос оставался в тени такого фундаментального проекта, как «Геном человека», использующего преимущественно кДНК-технологии. Вместе с тем от наших предков нас отличают не столько транскрипционные единицы, сколько огромные массивы некодирующей ДНК (Pederson, 2000). Предполагается, что при этом некодирующая ДНК обеспечивает укладку остальной части генома (Manuelidis, Borden, 1988; Haaf, Schmid, 1991). Происходящее при этом пространственное позиционирование транскрибуируемой ДНК приводит к формированию специфического паттерна экспрессии (Blobel, 1985; Pederson, 2000; Misteli, 2005). В рамках данной гипотезы постулируется наличие факторов, распознавающих разные типы сатДНК и участвующих в формировании трехмерной (3D) структуры ядра.

Если, как полагают, тандемные повторы в интерфазе вовлечены в поддержание трехмерной структуры хроматина, то следует ожидать, что взаимодействующие с ними белки находятся в составе ядерного матрикса (ЯМ). Концепция ЯМ как скелетной структуры оказала сильное влияние на современную биологию. Благодаря концепции ЯМ определены районы прикрепления ДНК к ЯМ в промоторах генов (*Matrix Attachment Regions — MAR*). Но со временем становится ясно, что ЯМ — это в основном

фракция ядерных белков, полученных определенным образом, а в живой клетке белки ЯМ окружают хромосомные территории (Podgornaya, 2003).

Дополнительная и широко распространенная концепция 3D — организация интерфазных хромосом в «территории» и последствия такой упаковки для регуляции генной экспрессии. Полагают, что постоянно экспрессирующиеся гены расположены на периферии хромосомных территорий (Tajbakhsh et al., 2000). Экспрессирующиеся гены находят в больших хроматиновых петлях, отходящих от периферии хромосомных территорий (Volpi et al., 2000). Напротив, обедненные генами районы никогда не формируют больших хроматиновых петель (Fakan, van Driel, 2007). Интеграция хроматиновых петель поддерживает структурное единство хромосомной территории. Полагают, что в основании петель находятся районы прикрепления к ЯМ — MAR, расположенные с периодичностью около 50—100 тыс. н. п. вдоль ДНК (Boulikas, 1995). Среди MAR есть постоянные и временные. Предполагают, что сатДНК представляет собой постоянные MAR, в то время как временные MAR, расположенные в промоторных районах транскрибуемых генов, обеспечивают прикрепление генов при необходимости. Полагают, что сатДНК является основным компонентом транскрипционно неактивного гетерохроматина, который остается конденсированным в интерфазном ядре и образует хромоцентры. Последовательности сатДНК видоспецифичны, их эволюционная вариабельность очень высока, но они обладают общими структурными особенностями и их положение на хромосомах строго фиксировано. Значительная часть сатДНК разных семейств расположена в CEN- и periцентромерном (peri-CEN) районах хромосом.

CEN—peri-CEN и прителомерные районы остаются «белыми пятнами» на картах хромосом, появившихся в результате чтения геномов человека и мыши (Lander et al., 2001; Venter et al., 2001; Waterston et al., 2002). На концах хромосом после точки начала сборки остается пустой район — Golden Path Gap, где среди мегабаз peri-CEN расположен и собственно CEN. Отсутствие ДНК-карты высокого разрешения CEN-района привело к отсутствию общепринятой модели его организации.

После чтения геномов биологическое сообщество занимается в основном организацией огромного массива знаний (data mining). Нельзя, однако, забывать о том, что значительная и несомненно важная часть геномов остается неизвестной. Собственно, это определяющие части хромосом CEN- и peri-CEN-районы. Организация генома не может быть понята без определения этих районов. Именно они отличают высших эукариот от низших. В них скрыты секреты 3D-организации хроматина и эволюционных изменений. Попыткам разрешения парадоксов организации CEN- и peri-CEN-районов посвящен настоящий обзор.

Последовательности periцентромерного района (peri-CEN)

Растет количество свидетельств того, что как CEN-, так и peri-CEN-хроматин необходимы для нормального функционирующего CEN (Bernard, Allshire, 2002; Durand-Dubief, Ekwall, 2008). Определение структурных особенностей, необходимых для формирования фланкирующего гетерохроматина CEN, затруднено тем, что это районы высокоповторяющейся ДНК.

Некоторые peri-CEN сатДНК обнаруживают свойства, характерные для сильно изогнутых ДНК (Radic et al., 1987; Martinez-Balbas et al., 1990; Hibino et al., 1992).

Мы показали структурно-специфичное связывание для сатДНК-связывающих белков ЯМ p68 хеликазы и SAF-A (Enukashvily et al., 2005). SAF-A связывает сатДНК и различает CEN и peri-CEN сатДНК в зависимости от параметра «изогнутости» ДНК (Lobov et al., 2001). Это и послужило основанием для попытки обнаружить общие структурные мотивы CEN. До сих пор все многочисленные попытки обнаружить общие для CEN мотивы, основанные на множественном выравнивании (multiple sequence alignment), не принесли успеха (см., например: Tyler-Smith, Floridia, 2000; Henikoff et al., 2001), а немногочисленные попытки обнаружить общие структурные особенности неубедительны, так как используются слишком короткие фрагменты, для которых CEN-активность не может быть показана экспериментально (Koch, 2000). У высших эукариот в области CEN не найдено ничего, кроме полей сатДНК, которая является видоспецифичным компонентом генома и сильно вариабельна в эволюции. В то же время группа белков CENP, маркерных для CEN, вполне успешно связывается с такими разнообразными последовательностями. У эукариот не нашли «магической» CEN-последовательности. Полагают, что структуру и функцию CEN определяет не собственно первичная последовательность, а определенная комбинация обычных фрагментов (Choo, 2000; Tyler-Smith, Floridia, 2000; Henikoff et al., 2001). Высказывалось предположение о том, что для формирования CEN вообще не нужна определенная ДНК, а происходит оно за счет эпигенетических факторов (Karpf, Allshire, 1997; Choo, 2000), хотя природа этих факторов остается неопределенной. Не удается обойтись без консервативной группы CENP (Centromere Protein) и белков, различающих фрагменты ДНК CEN и peri-CEN. CENP-A, CEN-аналог корового гистона H3, характерен для активного CEN и как гистон, т. е. нуклеосомный белок чувствителен к степени изогнутости фрагмента. Определенную последовательность нуклеотидов используют для картирования позиций нуклеосом. Теоретически предсказана и доказана в эксперименте преимущественная посадка нормального октамера гистонов на изогнутых участках ДНК (Trifonov, 1985). CENP-A может оказаться необходимым для образования нуклеосом на «прямых» CEN-фрагментах ДНК.

Структурно-специфичный механизм связывания сатДНК-связывающих белков (Lobov et al., 2000, 2001; Podgornaya et al., 2000) явился предпосылкой для моделирования структуры CEN и peri-CEN сатДНК (Podgornaya et al., 2003). Модель предсказывает, что CEN должен состоять из чередующихся фрагментов сатДНК, которые различаются по степени «изогнутости». Однако проверить предсказание модели невозможно из-за отсутствия полной последовательности CEN.

Клонирование новых последовательностей CEN

Отсутствие ДНК-карты высокого разрешения CEN-района привело к отсутствию общепринятой модели его организации. Попытки конструирования искусственной хромосомы основаны на предположении о том, что только один тип сатДНК входит в состав CEN (альфа-сатДНК человека или минорный — МиСат — мыши). Однако от-

носительно стабильно наследующиеся конструкции всегда содержат неизвестный материал материнского генома.

Ранее считалось, что у домовой мыши *Mus musculus* сатДНК, входящая в состав центромеров, достаточно однородна и представлена монорным (МиСат) сателлитом. МиСат состоит из мономеров в 120 н. п., 1 % генома состоит из МиСат, он находится в СЕН всех хромосом, кроме Y. Для того чтобы определить недостающие последовательности СЕН, мы использовали материал ДНК хромоцентров гепатоцитов мыши. Хромоцентры выделяли из печени взрослых лабораторных мышей *M. musculus* по методике, предложенной в России (Прусов, Зацепина, 2002).

Из фракции хромоцентров мыши клонировали, в частности, последовательность мышного сателлита 3 (MS3). Компьютерный анализ MS3 методом множественного выравнивания показал, что MS3 нет в базах данных. Около 1 % тотальной ДНК составляет MS3 с длиной мономера 160 н. п. «Лесенка», характерная для сатДНК, выявляется на Саузерн-блоте с рестрицированной тотальной ДНК и меченым MS3 (Kuznetsova et al., 2005). Показана локализация MS3 в области первичной перетяжки. Гибридизация на растянутом хроматине (fiberFISH) показала, что блоки МиСат и MS3 длиной около 2 тыс. н. п. перемежаются в районе СЕН. Однако на вытянутых фибрillах СЕН-района интерфазного хроматина остались места, не покрытые ни одним из клонированных фрагментов сатДНК (Kuznetsova et al., 2006). Предполагается клонировать недостающий кусок нормального СЕН после проведения ПЦР на ДНК хромоцентров с праймерами, специфичными для МиСат и MS3. Модель СЕН предполагает наличие не менее трех типов сатДНК, различаю-

щихся по содержанию GC и третичной структуре ДНК. Определение кривизны фрагмента MS3 по сравнению с МиСат показало, что он более «прямой», что согласуется с его положением в СЕН и предсказаниями модели. В соответствии с моделью минимальный СЕН должен содержать повторяющуюся единицу размером около 6 тыс. н. п. из трех типов сатДНК, различающихся по «кривизне».

Видоспецифичность последовательностей пери-SEN

Из библиотеки фрагментов ДНК хромоцентров мыши клонированы и пери-SEN-фрагменты, в частности MS4 (Kuznetsova et al., 2005). До сих пор считалось, что у домовой мыши *M. musculus* основную массу сатДНК составляют МиСат и мажорный (MaCat) сателлиты. Мономер MaCat в 234 н. п. картирован в пери-SEN-районах всех хромосом и составляет 5—10 % генома по биохимическим данным. По данным компьютерного анализа генома, поле tandemных повторов MaCat заканчивает плечи телоцентрических хромосом и дальше начинается несобранный район, в котором предположительно находится СЕН. MS4 составляет ~2 % генома, и вблизи СЕН есть поля, где фрагменты MaCat и MS4 перемежаются (Kuznetsova et al., 2006). Обладая наиболее полным набором проб сатДНК мыши, мы попытались картировать их на хромосомах видов рода *Mus* и близких родов.

В литературе принято мнение о том, что сатДНК видоспецифичны. Однако мнение это базируется в основном на том, что основные, мажорные, сатДНК (допустим, крысы и мыши) не имеют сходства при выравнивании

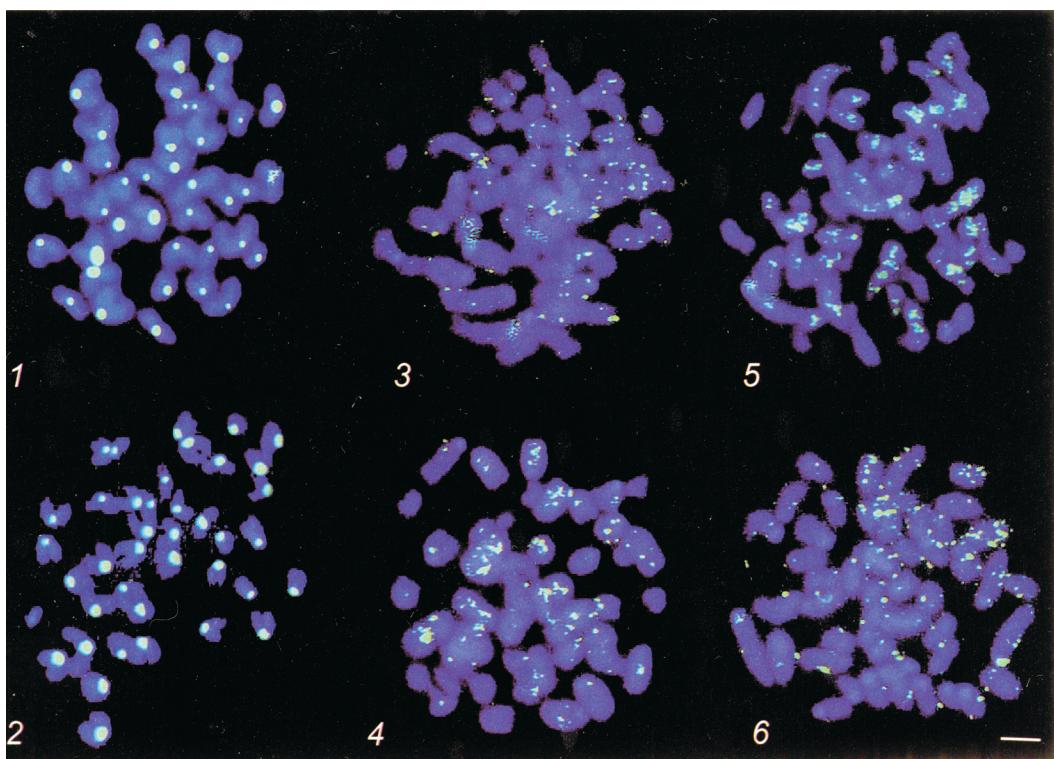


Рис. 1. Флуоресцентная гибридизация *in situ* (FISH) на метафазных пластинках мышей разных видов.

Помечена коммерческая пробы panCEN вида *Mus musculus*. 1 — *M. musculus*, 2 — *M. spicilegus*, 3 — *Apodemus agrarius*, 4 — *A. peninsula*, 5 — *Sylvaemus flavicollis*, 6 — *S. uralensis*. Масштабная линейка — 10 мкм.

(Беридзе, 1982). Не определено, все ли сатДНК гетерохроматина уникальны, потому что их полный набор неизвестен ни для одного вида. Однако, действительно, в гетерохроматиновом районе CEN и пери-CEN близких видов (родов) мало общего.

В качестве примера приведена флуоресцентная гибридизация *in situ* (FISH) с коммерческой пробой на все центромеры лабораторной мыши *M. musculus* (рис. 1). По условиям изготавления проба содержит как CEN, так и пери-CEN сатДНК. Если у видов одного рода *Mus* проба занимает сходные позиции в цитологически определяемом пери-CEN (рис. 1, 1, 2), то у видов других родов, хотя визуально это безусловно мыши, проба даже не ложится в области первичной перетяжки (рис. 1, 3—6). При использовании проб конкретных сатДНК картина еще более усложняется. На основании гибридизации на конденсированных хромосомах нельзя утверждать, что MS3 и MS4 у вида *Mus spicilegus* занимают позиции соответственно в CEN и пери-CEN, как это показано для *M. musculus*. Для утверждения о том, что CEN обоих видов построен из тех же последовательностей, нужно провести более тонкое картирование с использованием всех сатДНК CEN *M. musculus*, что невозможно до полной сборки CEN. С другой стороны, при использовании ПЦР с праймерами на конкретные сатДНК и дот-гибридизации можно обнаружить их небольшое количество в геномах других родов.

Основной современной гипотезой, которая описывает такое положение, является «библиотечная гипотеза» (Mravinaс, Plohl, 2007). Гипотеза предполагает, что в гетерохроматине каждого вида (рода) есть фрагменты самых разных сатДНК, но при фиксации вида происходит избирательное размножение фрагментов, т. е. мажорным фрагментом одного вида становится сатДНК1, а другого — сатДНК2. Так, гетерохроматин является «библиотекой» различных фрагментов, использующихся по необходимости. Гипотеза выдвинута после изучения гетерохроматина жуков. Материал для проверки гипотезы у мlekопитающих пока отсутствует.

И гипотеза, и экспериментальные данные приводят к очевидному выводу. В зоологии давно существует проблема абсолютного видового классификационного признака, желательно на основе ДНК (Tautz et al., 2003). Таким признаком был бы фрагмент ДНК, легко определяемый с помощью ПЦР при наличии известной последовательности и соответственно праймеров, абсолютно присущий данному виду и никакому другому. Наиболее успешно для решения этой задачи сейчас используют определенные гены митохондриальной ДНК (Waugh, 2007). Мажорные сатДНК вида, т. е. фрагменты пери-CEN ДНК, вряд ли обеспечат создание праймеров, потому что есть межвидовые перекрестья, о чем говорит и «библиотечная» гипотеза. Зато CEN сатДНК — вполне возможный кандидат на роль удобного классификационного признака. Для того чтобы это доказать, необходимо только определить полные CEN-последовательности близких видов.

Анализ базы данных Chromosome Unknown (ChrUn)

При секвенировании фрагментов генома и сборке из этих фрагментов непрерывных участков (контигов), из которых в дальнейшем собирают хромосомы, последовательности, для которых невозможно определить местоположение на хромосомах, помещают в дополнительную

хромосому — Chromosome Unknown (ChrUn, ftp://ftp.ncbi.nih.gov/genbank/genomes/M_musculus/CHR_Un/). ChrUn можно рассматривать как обогащенную различными классами повторяющихся элементов, в том числе сатДНК, и, вероятно, обогащенную последовательностями, входящими в состав гетерохроматина. Ведь именно из-за содержания большого количества повторяющихся элементов в секвенированной последовательности картировать ее на хромосоме практически невозможно.

Методы анализа уникальных последовательностей генома представлены солидным арсеналом; существуют и подходы к поиску повторяющихся последовательностей. Но набор методик анализа последовательностей, содержащих повторяющиеся элементы, на удивление скрупулезен, главным образом по причине того, что повторяющиеся последовательности рассматривают как мешающие анализу кодирующих белок регионов генома. Базу ChrUn анализировали собственной программой CHRUNTA (Kommisarov, Podgornaya, 2007) и программой Tandem Repeats Finder (Benson, 1999), позволяющей быстро идентифицировать вероятные tandemные повторы в геномной ДНК. СатДНК считали разновидностью tandemных повторов. ChrUn содержит более 50 000 фрагментов с общей длиной около $199 \cdot 10^6$ н. п. Из них 42 % содержат tandemные повторы. В ChrUn найдено более 35 000 вариантов последовательностей tandemных повторов. Для исключения мини- и микросателлитных повторов размеры повторяющегося мономера tandemного повтора задали в интервале 15—5000 н. п., также исключили все tandemные поля, где нет хотя бы одной из букв. В результате получили 2435 уникальных полей tandemных повторов. Для каждого поля была получена консенсусная последовательность. Для консенсусных последовательностей построен график распределения по содержанию GC (рис. 2).

Консенсусы AT-богатых сатДНК MaSat и MiSat обнаружили в 1 пике распределения; GC-богатых MS3 и MS4 — в пиках, лежащих за 50 % GC. Компьютерно предсказанные консенсусы для MS3 и MS4 не полностью совпадают с последовательностью экспериментально клонированных фрагментов, что естественно для сатДНК: консенсусы MaSat и MiSat уточняли несколько лет по мере клонирования все большего числа однотипных фрагментов.

Сходный анализ проведен и для некартированных последовательностей генома крысы (рис. 2, б). Несмотря на то что общая длина ChrUn крысы гораздо меньше, чем мыши (количество секвенированных букв ДНК генома крысы меньше), и количество найденных tandemных полей также меньше, общая тенденция распределения сохраняется. Это предварительный результат, но при высокой вариабельности tandemных повторов и отсутствии сходства удалось нашупать некие закономерности распределения: обязательным является наличие как AT-, так и GC-богатых повторов в определенных соотношениях.

Компьютерными методами, используя программы CHRUNTA и iGRAPH (Csárdi, Nepusz, 2006), проанализировали весь (собранный и ChrUn) геном мыши для выявления семейств tandemных повторов. Самым представительным является семейство MaSat, которое представлено в базах повторов REBASE под именем GSAT (см. таблицу). Именно полями GSAT заканчиваются собранные хромосомы мыши. Далее начинается несобранный участок (Golden Path Gap — GPG). Часто в tandemных полях мономеры сильно гетерогенны (MMTR3,4) — выяв-

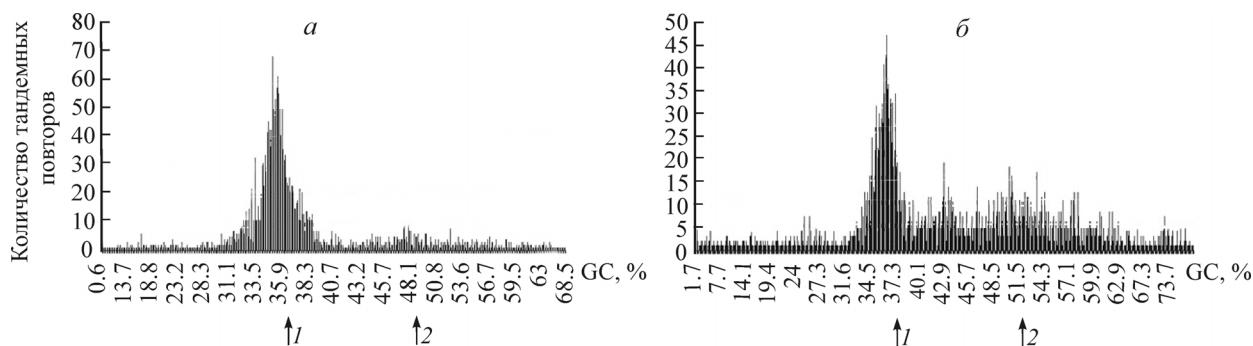


Рис. 2. Распределение tandemных повторов ChrUn мыши (а) и крысы (б) по содержанию GC (в %).

По горизонтали — доля GC в консенсусе, %; по вертикали — количество найденных консенсусов в каждой фракции. Стрелками отмечены пики распределения (AT-богатый — 1, GC-богатый — 2).

ляются мономеры в 30, 32, 35 и 62 н. п., хотя чаще всего встречается размер 30 н. п. В таблице приведен минимальный размер мономера, но, как свойственно сатДНК, внутри полей мономеры входят в иерархию повторяющихся единиц. Кроме SGAT (MaCat), консенсусные последовательности остальных повторов не содержатся в REPBASE. Среди них есть как AT- (MMTR1,7,8), так и GC-богатые (MMTR2,5). Есть как изогнутые, так и прямые фрагменты. Изогнутые фрагменты напоминают по структурным особенностям MAR и подтверждают гипотезу о том, что гетерохроматин содержит постоянные MAR, которые могут структурировать интерфазный хроматин.

В tandemных полях всех найденных tandemных повторов можно проследить гомологию с мобильными элементами, но в разной степени. По степени гомологии их можно разделить на два типа. Часть повторов первого типа (MMTR1,3,4,7,8) обнаруживает отчетливую гомологию. Так, например, поля MMTR1 и MMTR4 частично гомологичны фрагментам LINE и ERV соответственно. Эти повторы образуют небольшие tandemные поля, не превышающие 1 тыс. п. н. При компьютерном картировании фрагмента на собранной части генома повторы первого типа располагаются по плечам хромосом и в субтеломерной области. Повторы первого типа нельзя отнести ни к

мини-, ни к микро-, ни к собственно сателлитным tandemным повторам.

Для второго типа (GSAT, MMTR2 и MMTR5) гомология с диспергированными повторами менее очевидна, чем для первого. Для второго типа (MMTR2 и MMTR5) характерны большие tandemные поля, более 1 тыс. п. н., этот тип можно считать сатДНК в классическом понимании этого термина. В тех случаях, когда повторы расположены на собранных хромосомах, MMTR2 и MMTR5 лежат в областях, прилежащих к GPG, непосредственно переходя в поле GSAT, т. е. находятся на границе между GSAT и остальной хромосомой. Оба этих повтора GC-богатые и прямые (см. таблицу). Информация о них, полученная *in silico*, весьма напоминает экспериментально полученную картину с чередованием фрагментов MaCat и MS4 (Kuznetsova et al., 2006).

Один из выявленных повторов представлен только в ChrUn (MMTR3), и количество его там весьма значительно. Для него характерно наличие GCC-мотива. Для Ch1 не нашли полей с более чем 10 мономерами (~300 н. п.). Он относится к первому типу повторов, для которых неочевидна их принадлежность к сатДНК.

Предварительные данные показывают, что такая картина распределения семейств tandemных повторов характерна не только для генома мыши, но и для генома крысы,

Суперсемейства tandemных повторов *Mus musculus*

Условное название	Тип	Размер группы	Мономер, н. п.	GC, %	CENSOR—REPBASE	NCBI mouse genome	Кривизна фрагмента
MaSat	2	1853	59	36	GSAT	GPG	+
MMTR1	1	108	42	38	Unknown ~5'-L1-like	Везде	+
MMTR2	2	69	21	57	Unknown	GPG	-
MMTR3	1	30	30—62	34—65	»	ChrUn	-
MMTR4	1	25	22—45	34—59	Unknown ~ERV-like	Везде	+
MMTR5	2	20	31	52	Unknown	GPG	-
MMTR5	1	11	17	38	»	Везде	+
MMTR6	1	9	84	38	»	GPG	+

Примечание. Приведены первые классы по количественному содержанию в Chromosome Unknown. Новым повторам присвоен буквенный индекс MMTR — *Mus Musculus Tandem Repeat* и порядковые номера. В колонке «размер группы» приведено количество найденных в геноме tandemных полей. NCBI mouse genome — локализация в геноме; GPG — tandemные поля локализуются на собранных хромосомах вблизи от точки начала сборки (*Golden Path Gap*); ChrUn — есть только в ChrUn или есть везде. CENSOR — программа для поиска гомологии с REPBASE (базой данных повторов). Приведена предсказанная кривизна фрагмента.

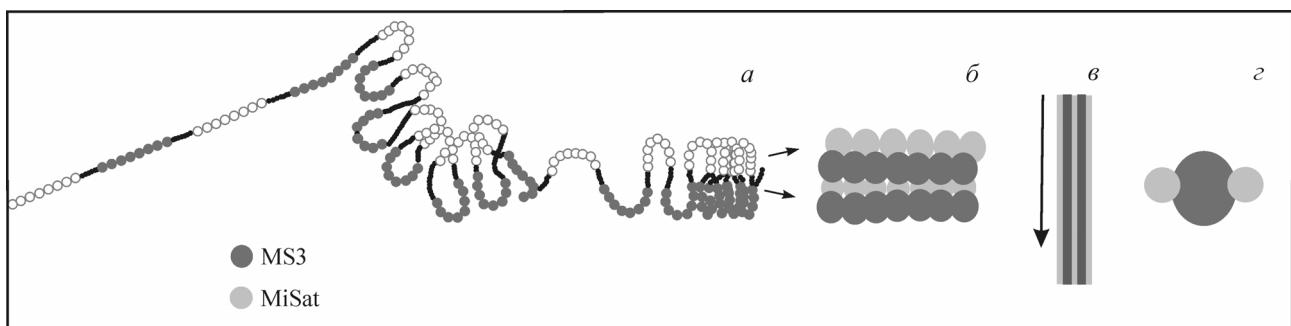


Рис. 3. Схема укладки CEN-хроматина мыши.

a—c — последовательные стадии конденсации начиная с растянутого хроматина (*a*). Темно-серым (MS3) и светло-серым (MiSat) отмечены картированные фрагменты, организованные в октамер нуклеосом (superbreads) длиной около 2 т. п. н.; стрелка указывает гипотетическое направление конденсации; фрагменты неизвестной природы показаны черной линией между октамерами (*a*). Построено по экспериментальным данным Кузнецовой с сотрудниками (Kuznetsova et al., 2006).

так же как сходно распределение повторов по содержанию GC (рис. 2).

Экспериментальное доказательство принадлежности предсказанных *in silico* тандемных повторов гетерохроматину и доказательство их сатДНК природы, или отнесение к отдельному классу повторов, являются предметом будущих исследований. Но и сейчас очевидно, что с помощью анализа ChrUn разных геномов можно найти особенности организации гетерохроматина, построенного из казалось бы абсолютно несходных фрагментов в разных геномах.

Заключение

CEN занимают поля в десятки тыс. н. п., окруженные полямиperi-CEN магабазных размеров, причем оба района богаты тандемными повторами. Как выделяются поля CEN в полях гетерохроматина и чем отличается их укладка, остается загадкой для молекулярной биологии. Предпринимаются попытки решить задачу как экспериментально, так и методами биоинформатики.

Полагают, что CEN человека содержит только альфа-сатДНК в виде хромосом-специфичных вариантов. Две независимые группы исследователей сообщили о реконструкции функционирующего CEN на основе конструкций с CEN-специфичным альфа-сатДНК (Ikeda et al., 1998; Henning et al., 1999; Alazami et al., 2004). Полученные в результате искусственные хромосомы человека (HAC) отличаются невысокой митотической стабильностью в неселективных условиях, их величина всегда больше введенной конструкции, а природа захваченной ДНК неизвестна (В. Ларионов, личное сообщение). Захват ДНК и структура CEN мыши (рис. 1) заставляют сомневаться в том, что для построения CEN человека достаточно одного типа сатДНК.

Биоинформатика может работать с большими фрагментами, содержащими поля альфа-сатДНК, которые есть в ChrUn. Среди мономеров альфа-сатДНК в полях можно выявить повторы высшего порядка — higher order repeats (HOR). Построен ряд моделей, который предполагает укладку HOR в структуру, подобную кристаллу. В основе модели лежат расчет положения нуклеосом и линкера в HOR и последующая укладка поля в фибрillлу диаметром 30 нм. оказывается, что в зависимости от количества мономеров в HOR и вариаций их длины фибрillла-кристалл приобретает хромосом-специфичные характеристики. Та-

ким образом, модель предполагает, что кристаллы CEN могут быть хромосом-специфичными и распознаваться как таковые (Rosandic et al., 2008).

В нашем компьютерном анализе ChrUn мы в основном задаемся вопросом о ДНК в составе гетерохроматина, не надеясь найти там истинно CEN-фрагменты. В гLOSSарии, приложенном к публикации генома человека, CEN и теломерные районы определены как «плохо поддающиеся клонированию» (Lander et al., 2001). При чтении геномов чаще всего используют клонирование в бактериальных или дрожжевых искусственных хромосомах (yeast artificial chromosome — YAC). В бактериальном геноме нет тандемных повторов; CEN дрожжей является уникальной последовательностью, а для построения peri-CEN дрожжи *S. pombe* использовали фрагменты тРНК (Takahashi et al., 1992), поэтому тандемные повторы высших эукариот плохо поддерживаются в YAC. CEN-поля млекопитающих, которые могут связывать аналоги CEN белков дрожжей, видимо, просто элиминируются вместе с несущим их YAC. Peri-CEN-поля поддерживаются лучше, потому что их сходство с дрожжевыми peri-CEN слабее. Мы видим из анализа ChrUn мыши (см. таблицу), что в тандемных повторах можно найти мотивы диспергированных повторов, а не тРНК. Пока это очевидно для мыши, но можно думать, что сходство выявится и для тандемных повторов в других прочитанных геномах.

Предсказанные тандемные повторы не случайно распределены у мыши и крысы (рис. 3). Найденные *in silico* классы тандемных повторов содержат как AT- так и GC-богатые, и соответственно их структура, «кривизна» различаются (см. таблицу). Так же как тандемные повторы, MAR не имеют консенсусной последовательности, но определяются по набору характерных, в том числе структурных, особенностей, обеспеченных и обогащением AT (программа MAR-Wizard). Все те предсказанные повторы, которые обнаружены не только в ChrUn, но и «везде» в собранном геноме (кроме сильно вариабельного MMTR4), являются AT-богатыми и «изогнутыми». При более точном их картировании они могут оказаться неизвестными MAR. Большинство хорошо охарактеризованных MAR находится в промоторных районах генов. Если AT-богатые тандемные повторы окажутся в других районах, то можно думать, что этот район нужен для структурирования генома.

В области картированных концов хромосом (GPG) есть «прямые» GC-богатые повторы типа сатДНК, непосредственно прилегающие к AT-богатому MaCat (GSAT;

см. таблицу), т. е. можно полагать, что есть чередование полей повторов с разными структурными характеристиками, как это наблюдается для CEN (рис. 3). Поля в этом случае гораздо больше. Гетерохроматин отличает «гистоновый код», в частности недоацетилирование и метилирование гистона H3 (H3K9me), что является меткой для связывания основного гетерохроматинового белка (HP1 — Heterochromatin Protein 1), который в свою очередь обеспечивает связывание когезина (Durand-Dubief, Ekwall, 2008; Loi et al., 2008). Белковые характеристики гетерохроматина едины для самых разных организмов, при том что даже у мышей близких родов ДНК гетерохроматина имеет мало общего (рис. 1). Однако примечательно, что, как в случае CEN, когда одним из специальных CEN-белков является CENP-A, модифицированный гистон H3, также и для гетерохроматина характерны модификации корового гистона, чувствительного к степени изогнутости ДНК.

Будут продолжаться дальнейшее изучение баз данных ChrUn, увеличение количества прочитанных последовательностей CEN и гетерохроматина, разработка способов их сборки. Кажется, что только выявление общих структурных особенностей последовательностей этих районов позволит разрешить парадокс, когда наборы белков, специфичных для CEN и района гетерохроматина, связываются с казалось бы абсолютно несходными последовательностями.

Работа выполнена при финансовой поддержке The Wellcome Trust (UK), Human Genome Program (HUGO) (DOE, USA), The Royal Society (UK) и Российского фонда фундаментальных исследований (проекты 05-04-49828-а и 05-04-49156-а).

Список литературы

- Beridze T. G. 1982. Сателлитная ДНК. М.: Наука. 121 с.
- Прусов А. Н., Зацепина О. В. 2002. Выделение фракции хромоцентров из ядер печени мыши. Биохимия. 67 (4) : 423—431.
- Alazami A. M., Mejia J. E., Monaco Z. L. 2004. Human artificial chromosomes containing chromosome 17 alphoid DNA maintain an active centromere in murine cells but are not stable. Genomics. 83 : 844—851.
- Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucl. Acids Res. 27 : 573—580.
- Bernard P., Allshire R. 2002. Centromeres become unstuck without heterochromatin. Trends Cell Biol. 12 : 419—424.
- Blobel G. 1985. Gene gating: a hypothesis. Proc. Nat. Acad. Sci. USA. 82 : 8527—8529.
- Boulikas T. 1995. Prediction of MAR sequences. Int. Rev. Cytol. 162A : 279—388.
- Choo K. H. A. 2000. Centromerization. Trends Cell Biol. 10 : 182—188.
- Csárdi G., Nepusz T. 2006. The igraph software package for complex network research. In: Proc. of the Int. Conference on Complex Systems. 1695—1696.
- Durand-Dubief M. I., Ekwall K. 2008. Heterochromatin tells CENP-A where to go. BioEssays. 30 : 526—529.
- Enukashvily N., Doney R., Sheer D., Podgornaya O. 2005. Satellite DNA binding and cellular localization of RNA helicase p68. J. Cell Sci. 118 : 611—622.
- Fakan S., van Driel R. 2007. The perichromatin region: a functional compartment in the nucleus that determines large-scale chromatin folding. Seminars in Cell & Develop. Biol. 18 : 676—681.
- Haaf T., Schmid M. 1991. Chromosome topology in mammalian interphase nuclei. Exp. Cell Res. 192 : 325—332.
- Henikoff S., Ahmad K., Malik H. S. 2001. The centromere paradox: stable inheritance with rapidly evolving DNA. Science. 293 : 1098—1102.
- Henning K. A., Novotny E. A., Compton S. T., Guan X. Y., Liu P. P., Ashlock M. A. 1999. Human artificial chromosomes generated by modification of a yeast artificial chromosome containing both human alpha satellite and single-copy DNA sequences. Proc. Nat. Acad. Sci. USA. 96 : 592—597.
- Hibino Y., Nakamura K., Asano S., Sugano N. 1992. Affinity of a highly repetitive bent DNA for nuclear scaffold proteins from rat liver. Biochem. Biophys. Res. Commun. 184 : 853—858.
- Ikeno M., Grimes B., Okazaki T., Nakano M., Saitoh K., Hoshino H., McGill N. I., Cooke H., Masumoto H. 1998. Construction of YAC-based mammalian artificial chromosomes. Nat. Biotechnol. 16 : 431—439.
- Karpen G. H., Allshire R. C. 1997. The case for epigenetic effects on centromere identity and function. Trends Genet. 13 : 489—496.
- Koch J. 2000. Neocentromeres and alpha satellite: a proposed structural code for functional human centromere DNA. Hum. Mol. Genet. 9 : 149—154.
- Komissarov A., Podgornaya O. 2007. CHRUNTA — tandem repeat search and classification program. In: Proceedings of the 3rd Moscow conference on computational molecular biology. 155—156.
- Kuznetsova I., Podgornaya O., Ferguson-Smith M. A. 2006. High-resolution organization of mouse centromeric and pericentromeric DNA. Cytogenet. Genome Res. 112 : 248—255.
- Kuznetsova I. S., Prusov A. N., Enukashvily N. I., Podgornaya O. I. 2005. New types of mouse centromeric satellite DNAs. Chrom. Res. 12 : 9—25.
- Lander E., Linton L., Birren B. et al. 2001. Initial sequencing and analysis of the human genome. Nature. 409 : 860—921.
- Lobov I. B., Tsutsui K., Mitchell A. R., Podgornaya O. I. 2000. Specific interaction of mouse major satellite with MAR-binding protein SAF-A. Eur. J. Cell Biol. 79 : 839—849.
- Lobov I. B., Tsutsui K., Mitchell A. R., Podgornaya O. I. 2001. SAF-A and lamin B binding specificity *in vitro* correlates with the satellite DNA bending state. J. Cell. Biochem. 83 : 218—229.
- Loi P., Beaujean N., Khochbin S., Fulka J., Jr, Ptak G. 2008. Asymmetric nuclear reprogramming in somatic cell nuclear transfer? BioEssays. 30 : 66—74.
- Manuelidis L., Borden J. 1988. Reproducible compartmentalization of individual chromosome domains in human CNS cells revealed by *in situ* hybridization and three-dimensional reconstruction. Chromosoma. 96 : 397—410.
- Martinez-Balbas A., Rodriguez-Campos A., Garcia-Ramirez M., Sainz J., Carrera P., Aymami J., Azorin F. 1990. Satellite DNAs contain sequence that induce curvature. Biochemistry. 9 : 2342—2348.
- Misteli T. 2005. Concepts in nuclear architecture. BioEssays. 27 : 477—487.
- Mravinač B., Plohl M. 2007. Satellite DNA junctions identify the potential origin of new repetitive elements in the beetle *Tribolium madens*. Gene. 394 : 45—52.
- Pederson T. 2000. Half a century of «the nuclear matrix». Mol. Cell. Biol. 11 : 799—805.
- Podgornaya O. I., Dey R., Lobov I. B., Enukashvily N. I. 2000. Human satellite 3 binding protein from the nuclear matrix: isolation and binding properties. Biochim. biophys. acta. 1497 : 204—214.
- Podgornaya O. I., Voronin A. P., Enukashvily N. I., Matveev I. V., Lobov I. B. 2003. Structure-specific DNA-binding proteins as the foundation for 3-dimensional chromatin organization. Int. Rev. Cytol. 224 : 227—296.
- Radic M. Z., Lundgren K., Hamkalo B. A. 1987. Curvature of mouse satellite DNA and condensation of heterochromatin. Cell. 50 : 1101—1108.
- Rosandić M. et al. 2008. The role of alphoid higher order repeats (HORs) in the centromere folding. J. Theor. Biol.
- Tajbakhsh J., Luz H., Bornfleth H., Lampel S., Cremer C., Lichten P. 2000. Spatial distribution of GC- and AT-rich DNA sequen-

- ces within human chromosome territories. *Exp. Cell Res.* 255 : 229—237.
- Takahashi K., Murakami Sh., Chikashige Y., Funabiki H., Niwa O., Yanagida M. 1992. A low copy number central sequence with strict symmetry and unusual chromatin structure in fission yeast centromere. *Mol. Biol. Cell.* 3 : 819—835.
- Tautz D., Archtander P., Minelli A., Thomas R. H., Vogler A. P. 2003. A plea for DNA taxonomy. *TREE*. 18 : 70—74.
- Trifonov E. N. 1985. Curved DNA. *CRC Crit. Rev. Biochem.* 19 : 89—106.
- Tyler-Smith C., Floridia G. 2000. Many paths to the top of mountain: diverse evolutionary solutions to centromere structure. *Cell.* 102 : 5—8.
- Venter C. et al. 2001. The sequence of the human genome. *Science*. 291 : 1304—1351.
- Volpi E. V., Chevret E., Jones T., Vatcheva R., Williamson J., Beck S., Campbell R. D., Goldsworthy M., Powis S. H., Ragoussis J., Trowsdale J., Sheer D. 2000. Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. *J. Cell Sci.* 113 : 1565—1576.
- Waterston R. H. et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature*. 420 : 520—562.
- Waugh J. 2007. DNA barcoding in animal species: progress, potential and pitfalls. *BioEssays*. 29 : 188—197.

Поступила 13 X 2008

HETEROCHROMATIN AND CENTROMERE STRUCTURE PARADOX

O. I. Podgornaya, D. I. Ostromishensky, I. S. Kuznetsova, I. V. Matveev, A. S. Komissarov

Institute of Cytology RAS, St. Petersburg;
e-mail: aleksey.komissarov@gmail.com

Centromere (CEN) is the structure responsible for the chromatid association, chromosome attachment to the spindle, and correct position in the plate. The only DNA found in the mammalian CEN belongs to the satellite DNA — high repeated tandem repeats. Mounting evidence indicates that both types of chromatin (CEN and peri-CEN) are required for proper centromere function. CEN, peri-CEN and peritelomeric regions remain «white spots» at the chromosome maps appeared after reading genomes of human, mouse, and rat. SatDNA is considered to be species-specific. «Library» hypothesis regards heterochromatin as the library of different satDNA one fragments of which became spread and fixed in species fixation. We have analyzed database Chromosome Unknown (ChrUn) and found several new classes of mouse tandem repeats. The features of these classes are similar with the ones from rat ChrUn, as well as their distributions according to GC-richness. We believe that similar fragments' structure, i. e. intermingling of fragments with different curvature rather than their primary sequence will help to solve the paradox, when CEN or peri-CEN fragments from different animals have nothing in common, but bind the same sets of proteins.

Key words: heterochromatin, centromere, pericentromere, satellite DNA, databases, genome, Chromosome Unknown, curved DNA.