

Остромышенский

Дмитрий Игоревич

Состав хромоцентров мышцы *in silico* и их основной компонент, тандемные
повторы, у мышевидных грызунов

03.01.03 – молекулярная биология

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

кандидата биологических наук

Санкт-Петербург

2018

Работа выполнена в Федеральном государственном бюджетном учреждении науки Институт цитологии Российской академии наук

Научный руководитель: доктор биологических наук, профессор
Подгорная Ольга Игоревна
заведующий группой некодирующей ДНК
Федерального государственного бюджетного
учреждения науки Институт цитологии
Российской академии наук

Официальные оппоненты: Доктор биологических наук, профессор
Родионов Александр Викентьевич
заведующий лабораторией биосистематики и
цитологии Федерального государственного
бюджетного учреждения науки Ботанический
институт имени В.Л. Комарова Российской
академии наук

Доктор биологических наук, профессор
Крамеров Дмитрий Александрович
заведующий лаборатории эволюции
эукариотических геномов Федерального
государственного бюджетного учреждения науки
Институт молекулярной биологии имени В.А.
Энгельгардта Российской академии наук

Ведущая организация: Федеральное государственное бюджетное
учреждение науки Институт молекулярной и
клеточной биологии Сибирского отделения
Российской академии наук, Новосибирск

Защита диссертации состоится «__» _____ 2018 г. в ____ часов на заседании
Диссертационного совета Д 002.230.01 на базе Института цитологии РАН по адресу:
194064, Санкт-Петербург, Тихорецкий проспект, д. 4.

Сайт института: <http://www.cytspb.rssi.ru>

Адрес электронной почты института: cellbio@incras.ru

Факс института: (812) 297-35-41

С диссертацией можно ознакомиться в библиотеке Института цитологии РАН и на
сайте <http://www.cytspb.rssi.ru>

Автореферат разослан «__» _____ 2018 г.

Ученый секретарь Диссертационного совета,
кандидат биологических наук

Е.В. Каминская

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность проблемы

Конститутивный гетерохроматин (ГХ) играет важную роль в функционировании генома. На ядерном уровне организации ГХ вовлечен в процессы поддержания пространственной организации хроматина в ядре; на хромосомном уровне - необходим в процессах мейоза и митоза. У многих видов ГХ образует четко очерченные структуры в ядре - хромоцентры. Считается, что в образовании одного хромоцентра может принимать участие ГХ разных хромосом (Wijchers et al., 2015).

В ядрах домыши хромосомы окрашиваются DAPI из-за присутствия в них большого количества АТ-богатого мажорного сателлита (MaSat) (Mayer et al., 2005; Probst, Almouzni 2008). Хромоцентры являются плотными структурами, сформированными из материала центромерных (ЦЕН) и перичентромерных регионов (периЦЕН) разных хромосом. Иммуногистохимическими методами в хромоцентрах выявляют белки, характерные для ГХ, например, белок HP1 (Guenatri et al., 2004; Snapp et al., 2013). Гены, располагающиеся в районах ГХ обычно находятся в транскрипционно-неактивном состоянии. Основной составной частью ГХ являются различные типы повторяющихся последовательностей ДНК. У высших эукариот разные классы повторяющейся ДНК, основными из которых являются тандемные повторы (ТП), диспергированные повторы разных классов – ERV (Endogenous retroviruses, эндогенные ретровирусы), LINE (Long interspersed nuclear element, длинные диспергированные повторы) и ДНК-транспозоны, вместе составляют до 2/3 генома (de Koning et al., 2011). Характерной чертой генома домыши и других грызунов является большое распространение одного из ERV, относящегося ко II-классу ERV (ERV2) — IAP (Intracisternal A-Particle), образующий внутриклеточные вирус-подобные частицы внутри цистерн эндоплазматического ретикула на ранних стадиях эмбриогенеза (Ribet et al., 2008, Magiorkinis et al., 2012).

Давно известно, что главным компонентом ГХ являются ТП, которые, например, у мыши и человека составляют до 10-15% генома. ТП играют важную роль в клеточных процессах, кроме того, известны транскрипты ТП, обнаруженные в раннем эмбриональном развитии разных организмов (Erukashvily et al., 2009; Probst et al., 2010) и в клетках опухолей (Tanne A, et al., 2015; Bersani et al., 2015, Подгорная, Остромышенский и др, 2018). ТП являются самой сложной для сборки и аннотирования частью генома. У подавляющего большинства высших эукариот известны лишь несколько мажорных ТП. Даже бурное развитие техник секвенирования и сборки геномов не дало полного представления о составе районов хромосом, образованных ГХ из-за отсутствия алгоритмов сборки ТП. В сборках геномов большинства высших эукариот на месте ЦЕН/периЦЕН находится незаполненный промежуток — Golden Path Gap (GPG). Однако, некартированные на хромосомах и неаннотированные поля ТП присутствуют в контигах Whole Genome

Shotgun (WGS) баз данных. Новые биоинформатические подходы, которые ранее разработаны в нашей лаборатории, позволяют найти и аннотировать большие ТП в сборках геномов различного качества. Это позволяет находить как новые ТП в собранных геномах, так и проследить их эволюцию.

Трудности сборки, картирования и аннотирования участков ГХ ограничивают возможности исследования функциональной значимости последовательностей, входящих в его состав. Возможность выделения хромоцентров из ядер мыши позволяет определить качественный и количественный состав ГХ и т.о. обеспечить реперные точки для будущей сборки.

Цели и задачи исследования

Цель работы – определение состава конститутивного гетерохроматина доменной мыши и основных компонентов, входящих в его состав и выяснение некоторых аспектов эволюции главного компонента гетерохроматина — tandemных повторов.

Для выполнения этой цели поставлены следующие задачи:

1. Выделение, высокопроизводительное секвенирование и анализ ДНК хромоцентров доменной мыши. Анализ распределения найденных последовательностей в доступных сборках генома мыши.

2. Проверка полученных *in silico* данных молекулярно-биологическими методами (*in situ*).

3. Анализ распределения и эволюции tandemных повторов в геномах мышевидных грызунов.

Основные положения, выносимые на защиту

1. Основной составной частью хромоцентров и конститутивного гетерохроматина мыши являются ТП — мажорный и минорный сателлит. Остальные ТП представлены в незначительных количествах, но разнообразны.

2. Важной частью конститутивного гетерохроматина мыши является фрагмент 3'-конца ORF2 и 3'-некодирующего региона LINE длиной ~2 т.п.н. Полноразмерные LINE в конститутивном гетерохроматине не найдены.

3. В конститутивном гетерохроматине мыши найдены внутренние части ERV (в особенности IAP) и соответствующие им LTR, внутренние части ERV сильно фрагментированы.

4. Внутри клады мышевидных грызунов происходит быстрая эволюция tandemных повторов. Большинство tandemных повторов видоспецифичны или встречаются только у очень близких видов. Эволюция tandemных повторов в этой группе животных не соответствует библиотечной модели.

Научная новизна работы. Теоретическое и практическое значение работы.

Впервые проведено высокопроизводительное секвенирование и анализ состава ДНК хромоцентров мыши. Показано сильное обогащение ДНК хромоцентров и, следовательно, ГХ tandemными повторами. С помощью биоинформатических и

молекулярно-биологических методов доказано присутствие значительного количество ~2 т.п.н. фрагмента LINE. Показано, что аналогичная картина присутствует и в доступных сборках центромерных районов генома человека. Найдено большое количество фрагментов ERV - как внутренних частей, так и LTR. Среди ERV наиболее представлены IAP и их LTR. Найденные фрагменты ERV являются важными для дальнейших исследований транскрипции ТП как кандидаты на роль промоторов транскрипции ТП.

Аннотирование новых ТП в геномах разных организмов, в том числе модельных, таких как домовая мышь и китайский хомячок позволит глубже исследовать транскрипты и их роль в эмбриогенезе и канцерогенезе, а также будет способствовать изучению функциональной роли этого типа последовательностей ДНК в формировании архитектуры генома и регуляции при помощи lncRNA (long non-coding RNA, длинная некодирующая РНК).

ГХ — бедная генами плотно упакованная и обычно транскрипционно-неактивная часть генома. Последовательности, входящие в состав ГХ, часто называют «мусорной ДНК», но, отношение к ГХ как к «мусорной ДНК» постепенно меняется по причине обнаружения новых функций, как у входящих в его состав последовательностей, так и у ГХ в целом. Исследования показывают, что ГХ является далеко не такой инертной фракцией генома, как считалось ранее. Расшифровка хранимой в ГХ информации и поиск ее функций и значения для функционирования клетки в целом, станет на ближайшие годы одной из важнейших задач молекулярной биологии, геномики, генетики. Исследование ГХ позволит ответить на огромное количество вопросов связанных с эволюцией, медициной и другими актуальными областями науки. Для подобных исследований могут потребоваться десятки лет. Однако понимания функций невозможно без понимания того, какие последовательности выполняют эти функции. Но первые шаги, к которым относится и настоящая работа, уже сделаны, и можно с уверенностью говорить об их важности.

Материалы диссертации используются в курсах лекций для бакалавров и магистров Биолого-почвенного факультета СПбГУ, Школы Биомедицины и Школы Естественных наук ДВФУ и могут быть использованы в общих и специальных курсах лекций биологических факультетов других университетов.

Апробация работы. Результаты работы были представлены на 2-й международных конференциях Современные проблемы биологической эволюции (Москва, 11-14 марта 2014), 7-ой Московской международной конференции по вычислительной молекулярной биологии (Москва, 16-19 июля 2015), «Хромосома-2015» (Новосибирск, 24-28 августа 2015), Bioinformatics of Genome Regulation and Structure\ Systems Biology — BGRS\SB-2016 (Новосибирск, 25 августа — 2 сентября 2016), 25th Wilhelm Bernhard Workshop on the Cell Nucleus, (Нижний Новгород, 19—22 июня 2017).

Вклад автора. Автором выполнен весь биоинформатический анализ. Молекулярно-биологический блок выполнен совместно с И.С. Кузнецовой и В.Н. Стефановой. Клонирование ДНК хромоцентров и центромеров выполнено совместно с И.С. Кузнецовой. Пробоподготовка и секвенирование на Illumina MiSeq выполнено Черняевой Е.Н. В работе использованы микродиссекционные пробы первичной перетяжки хромосом мыши, любезно предоставленные В. А. Трифионовым (Институт молекулярной и клеточной биологии СО РАН)

Объем и структура диссертации. Диссертация состоит из введения, обзора литературы, описания материалов и методов, результатов и обсуждения, а также выводов и списка цитируемой литературы, в которые входит 159 ссылок. Работа изложена на 114 страницах машинописного текста, содержит 9 таблиц, 17 рисунков и приложение на 7 страницах.

Публикации. По результатам и проблематике настоящего исследования опубликовано 18 работ, из них 9 статей в рецензируемых изданиях рекомендованного перечня ВАК.

Работа выполнена при поддержке грантов РФФИ 15-15-20026 и Президиума РАН МКБ № 01201457147

МАТЕРИАЛЫ И МЕТОДЫ

Молекулярно-биологические методы

1. Получение ядер и метафазных хромосом. Ядра и метафазные хромосомы получали из костного мозга и культур клеток первичных фибробластов по стандартной процедуре (Guo, Wu 2008).

2. Выделение ДНК хромоцентров проводили по методике, предложенной Прусовым (Prusov et al., 2002, Zatsepina et al., 2008), которая основана на центрифугировании фрагментированных ультразвуком ядер печени в градиенте плотности сахарозы. ДНК ЦЕН района, полученная с помощью микродиссекции любезно предоставлена В. А. Трифионовым (Институт молекулярной и клеточной биологии СО РАН). Для клонирования обе пробы амплифицировали с помощью DOP-праймера (5'-CCGACTCGAGNNNNNATGTGG-3') (Arneson et al., 2008).

3. Клонирование. Амплифицированная ДНК хромоцентров и ЦЕН клонирована в вектор pCR2.1 TOPO используя TA Cloning Kit (Life Technologies). Плазмиды со вставкой были трансфицированы в компетентные клетки штамма DH5a E. coli по стандартному протоколу (Sambrook and Russell, 2001). 32 клон отсеквенированы по Сэнгеру.

4. Флуоресцентная гибридизация in situ (fluorescent in situ hybridization, FISH). Для гибридизации использовали зонды меченые биотином или FITC. Препараты метафазных хромосом с зондом денатурировали в буфере для

денатурации (70%-ный формамид; 2-кратный SSC) в течении 2 мин при 70 °С. Гибридизацию проводили во влажной камере при 37 °С в течение ночи. Зонды, меченные биотином, детектировали стрептавидином, конъюгированным с Alexa 488 или Alexa 568 (Invitrogen, США). Препараты окрашивали DAPI и заключали в среду VectaShield (Dabco, США). Препараты анализировали на флуоресцентном микроскопе AXIOSKOP-DFC360.

Компьютерные (in silico) методы

5. Анализ клонов ДНК хромоцентров и ЦЕН проводили с использованием BLAST (Altschul et al., 1990). Последовательности клонов искали в референсном геноме мыши версии 38.3 (http://www.ncbi.nlm.nih.gov/assembly/GCF_000001635.23) и в контигах WGS ААНУ (<http://www.ncbi.nlm.nih.gov/nucore/ААНУ00000000.1/>) и СААА (<http://www.ncbi.nlm.nih.gov/nucore/СААА00000000.1/>), повторы и аннотирование последовательностей сходных с диспергированными повторами проводили по коллекции повторов млекопитающих из Repbase.

6. Для секвенирования ДНК хромоцентров сделали библиотеку с использованием кита Nextera. Библиотека отсеквенирована на платформе Illumina MiSeq (размер вставки 220 нп) с короткими ридами длиной 37 п.н. Полученные риды очистили от ридов, содержащих неизвестные нуклеотиды, технических последовательностей Illumina и ридов, короче 20 п.н. с помощью программы Trimmomatic (Bolger et al., 2014).

7. Содержание повторов в хромоцентрах (ChrmC) определяли как процент ридов с помощью программы Bowtie2 с параметром sensitive-local. Для анализа содержания повторов извлекли из эталонного генома мыши версии GRCh38.p6 все последовательности диспергированных повторов разных классов согласно их координатам в выходном файле Repeat Masker для этой версии генома мыши. Анализ содержания ТП проводили на основании рассчитанных ранее полей ТП разных семейств (Komissarov et al., 2011). Аналогичным методом рассчитывали содержание повторов в ридов полнотранскриптома секвенирования домашней мыши линии 129P2 (wgHTS) полученных на платформе Illumina Genome Analyzer II (study ERP000034, experiment ERX002258, ERX002259, ERX002260, ERX002261; Wellcome Trust Sanger Institute). Для расчета покрытия ридов использовали следующий подход. Риды ChrmC и wgHTS картировали на консенсусные последовательности L1_MM и ERV (RLTR6I_MM, IAPEZI, MMERVK10C и MERVL из Repbase с помощью Bowtie2 с параметром sensitive-local.

Риды, очищенные, от картировавшихся на поля MaCat или MiCat собрали в контиги. Сборку контигов осуществляли геномным ассемблером IDBA_UD с параметрами по умолчанию. Аннотировали только контиги с длиной больше 300 п.н. Контиги сравнивали с последовательностями известных повторов из Repbase с помощью BLAST. Контиги, которым не нашлось соответствие среди известных повторов аннотировали относительно референсного генома мыши.

8. Поиск тандемных повторов. ТП искали в следующих сборках геномов: 1) *C. griseus*: сборки АРМК (длина сборки ~2,3 Gb, N50 ~11 kb) и АМДС (длина сборки ~2,3 млрд. п.н. N50 ~27 т.п.н.); 2) *M. auratus* сборка АРМТ (длина сборки ~2,5 млрд. п.н., N50 ~22,5 т.п.н.). Кроме того, использовали ранее найденные ТП (Komissarov et al., 2011) *Mus musculus* из референсного генома мыши (длина сборки ~2,8 млрд. п.н., N50 ~32 м.п.н.) и сборки ААНУ (длина сборки ~3,2 млрд. п.н., N50 ~22 т.п.н.).

9. Определения количества тандемных повторов в геномах. Для оценки содержания ТП в геномах исследуемых видов использовали программу bowtie2 с параметром чувствительности sensitive-local. С помощью этой программы выравнивали на все поля ТП, принадлежащих к одному семейству исходные риды чтения геномов: 1) *M.auratus* — SRR396599, SRR396609; 2) *C. griseus* — SRR803174, SRR803182; 3) *M.caroli* — ERR019942, ERR019947; 4) *M.musculus* — описанные в пункте 7. В качестве содержания в геноме принимали процент ридов, выровнявшихся на поля ТП.

РЕЗУЛЬТАТЫ

1. Секвенирование ДНК хромоцентров

В результате секвенирования ДНК хромоцентров получено 4371191 парных (paired-end) ридов со средней длиной 37 п.н. Полученные риды загружены в базу данных Sequence Read Archive (<http://www.ncbi.nlm.nih.gov/sra>; SRR3414277). Для дальнейшего анализа парные риды были объединены в один набор. Для очистки ридов от содержащих технические последовательности Illumina и ридов с низким качеством прочтения использовали программу Trimmomatic. После очистки получено 6856380 ридов. Они составляют коллекцию ДНК хромоцентров (ChrmC, chromocenters). Обобщенная схема работы представлена на Рис. 1.



Рисунок 1. Схема секвенирования и аннотации ДНК хромоцентров

2. Содержание повторов в хромоцентрах.

Оценили содержание разных классов повторов в ChrmC и сравнили его с

содержанием повторов в полногеномном секвенировании на Illumina GAII и в эталонном геноме мыши (Таблица 1; Ostromyshenskii et al., 2018a).

В ChrмС наибольшее содержание показали ТП: ~66% MaCat, ~4% MiCat и около 1% остальных (включая теломерный повтор). Т.о. общее содержание ТП в ChrмС ~71%. Следующим по содержанию в ChrмС, среди классов повторов оказались Non-LTR ретропозоны (в число которых входят SINE и LINE) ~13 % со значительным преобладанием LINE ~11%. На третьем месте – ERV с содержанием ~9%. Среди ERV преобладают ERV2 ~5% (из них более 2% приходится на долю IAP). Содержание ERV3 около 3% и меньше 1% приходится на долю ERV1. Также обнаружено незначительное (менее 1%) содержание ДНК-транспозонов. По сравнению с эталонным геномом и wgHTS в ChrмС наблюдается относительное обогащение LINE относительно SINE (соотношение LINE:SINE в эталонном геноме 2,3:1, в wgHTS 3,2:1, а в ChrмС 5,2:1). Содержание ERV и разных их классов в wgHTS и в эталонном геноме почти одинаковое: всего около 12% из них 5-6% ERV2, ~6% ERV3 и около 1% ERV1. По сравнению с эталонным геномом и wgHTS в ChrмС наблюдается обогащение ERV2 и особенно IAP.

Таблица 1. Содержание повторов в ДНК хромоцентров, рядах полногеномного секвенирования и в эталонном геноме.

Повторы		ChrмС %	WgHTS %	Эталонный геном %	
ТП	MaCat	66,2	13	*	
	MiCat	4,4	0,6	*	
	Остальные	0,9	0,9	*	
Non-LTR ретропозоны	LINE	10,9	17,1	19,1	
	SINE	2,1	5,2	8,3	
ERV	Всего		8,8	12,7	12,3
	ERV1		0,5	0,9	1,2
	ERV2	Всего	5,1	5,9	5
		IAP	2,2	0,8	1
	ERV3		3,2	5,9	6,1
ДНК-транспозоны		0,6	1,2	1,2	
Всего		93,7	50,8	40,9	

ТП крайне скудно представлены в эталонном геноме — общее содержание полей ТП не превышает 0,1-0,2%, что является следствием трудности сборки этого типа последовательностей. В wgHTS среди ТП преобладает MaCat ~13%.

Общее содержание повторов в ChrмС ~94%, т.е. 6% ридов остались неаннотированными. Причиной может являться недопредставленность повторов в RepBase и недостаточная их аннотированность в собранном геноме. Для определения природы последовательностей в неаннотированной части ридов, а также более полного исследования отдельных аннотированных повторов провели сборку отдельных ридов в более протяженные последовательности (контиги) и аннотировали их.

3. Сборка контигов и их аннотация.

Для дальнейшего анализа очистили ChrмС от ридов, которые картируются на поля MaCat и MiCat, а оставшиеся риды собрали в контиги с помощью программы IDBA_UD (Ostromyshenskii et al., 2018a). В результате сборки получено 93 контига с длиной >300 п.н. Максимальная длина контига 4385 п.н., общая длина контигов ~57 т.п.н. Контиги загружены в Genbank (Genbank ID KX121610 — KX121702).

Сравнение контигов с последовательностями Repbase показало, что наиболее представленными среди контигов являются фрагменты ERV. К их числу относятся 35 из 93 контигов общей длиной ~ 26 т.п.н. Контиги относятся ко всем трем классам ERV, но наиболее представлен ERV2 (к числу которых относится IAP). По 6 контигов относятся к Non-LTR ретропозонам — LINE и SINE. Также 4 контига аннотированы как фрагменты псевдогенов rRNA.

Контиги, которым не нашлось соответствия в Repbase аннотировали с помощью BLAST против эталонного генома мыши. Наибольшее количество контигов (31) принадлежат к неаннотированным диспергированно-повторяющимся последовательностям половых хромосом. Из этих контигов 19 контигов являются специфичными для Y хромосомы и повторены на ней более 500 раз, а остальные 12 – кроме Y хромосомы, найдены также и на X хромосоме. При этом на Y-хромосоме они встречаются более 500 раз, в то время как на X хромосоме 10-20 раз.

8 контигов аннотированы как последовательности рибосомальной ДНК, 6 контигов -принадлежащие к imprinted gene in the Prader-Willi syndrome region, расположенного на 7-й хромосоме. Для всех этих контигов найдено большое число совпадений (>100) в этом регионе. Известно, что регион состоит из тандемно-организованных кластеров генов малых ядрышковых РНК (Ding et al., 2008). Один контиг является фрагментом гена *Sfi1 homolog*. Этот ген локализован в периЦЕН районе 11-й хромосомы мыши. Два контига относятся многократно повторены в кластере псевдогенов V2R, расположенном на 13-й хромосоме (Yang et al., 2005). Еще два неаннотированных контига расположены на 14 хромосоме. Они также многократно повторяются (количество совпадений (хитов BLAST) > 200) в области прилежащей к GPG на расстоянии 0,1-4.5 м.п.н.

Анализ контигов показывает, что большинство из них относятся к

повторяющимся последовательностям. Около половины из них не аннотированы в геноме мыши (Ostromyshenskii et al., 2018a).

4. Эндогенные ретровирусы

ERV оказались третьими по содержанию при определении содержания в исходных ридсах ChrmC и первым по содержанию среди собранных контигов. В отличие от мономеров большинства ТП повторов ERV являются относительно большими последовательностями (~1-8 т.п.н.), поэтому встает вопрос о том содержатся ли в хромоцентрах целые копии ERV, или их отдельные их фрагменты. Картирование контигов на консенсусные последовательности ERV разных классов ERV1 (RLTR6I_MM), ERV2 (IAPEZI, MMERVK10C) и ERV3 (MERVL) из Repbase показало, что контиги почти полностью покрывают последовательности ERV разных классов с небольшими промежутками или вовсе без них, как в случае MERVL. Анализ покрытия ридсами ChrmC последовательностей IAPEZI, MERVL, RLTR6I_MM и MTA показал, что покрытие достаточно равномерно для всех ERV (Рис. 2). Не найдено какого-либо обогащения отдельных фрагментов MTA ридсами ChrmC, что может означать, что полноразмерные MTA присутствуют в хромоцентрах. Отличие на Рис. 2 MTA от остальных ERV может быть объяснено разницей в масштабе для MTA, поскольку этот элемент является самым коротким среди других ERV. Аналогичную картину показал анализ покрытия тех же последовательностей ридсами wgHTS (Рис. 2, красные линии). По причине большего общего покрытия в wgHTS график более равномерный без ярко выраженных пиков, однако в целом картина близка к таковой для ChrmC (Ostromyshenskii et al., 2018a).

Полученные данные не дают ответа на вопрос присутствуют ли в конститутивном гетерохроматине (ГХ) полноразмерные копии ERV. Из WGS проекта ААНУ выбрали контиги, содержащих ЦЕН и периЦЕН ТП (MaCat, MiCat, TRPC-21A) вместе с ERV. Их анализ показал, что ERV в контигах с ТП сильно фрагментированы, а фрагменты могут располагаться в неправильном порядке. Исключение составляет короткий MTA. Таким образом, в конститутивном гетерохроматине ERV представлен отдельными фрагментами, однако преобладания частей ERV не обнаружено. В контигах WGS около половины всех фрагментов ERV из ТП-содержащих контигов относятся к IAP и соответствующим им LTR.

Для подтверждения наличия IAP в конститутивном гетерохроматине мы сделали FISH с синтетическим зондом к последовательности IAP. FISH на хромосомах линии L929 и хромосомах клеток красного костного мозга мыши свидетельствует о том, что на метафазных хромосомах сигнал тяготеет к

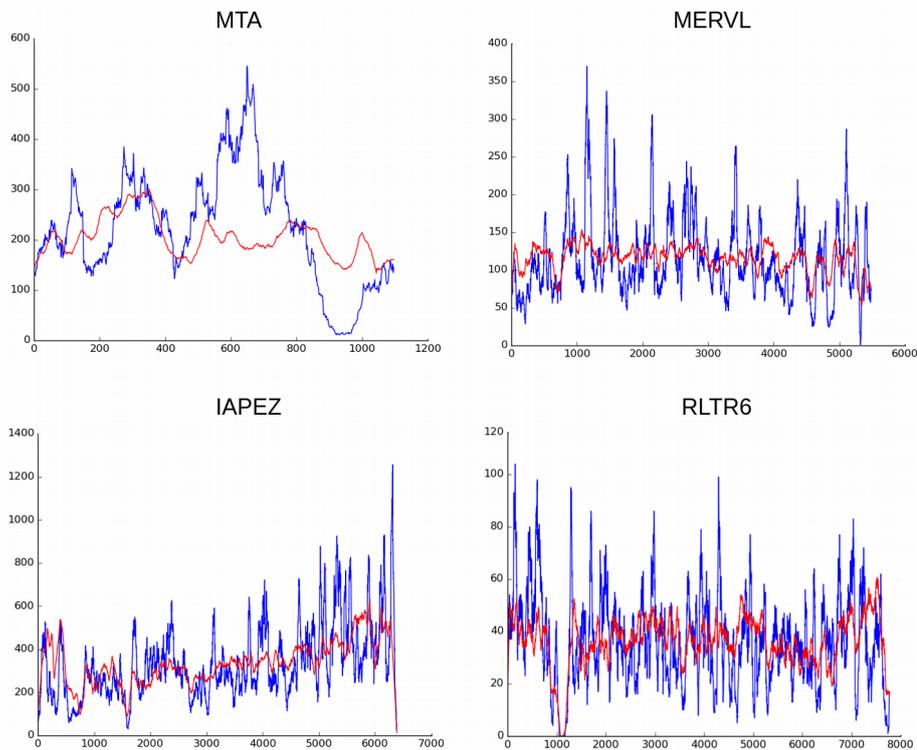


Рисунок 2. Покрытие ридами ChrmC (синий) и wgHTS (красный) консенсусных последовательностей ERV из RepBase. По оси абсцисс – длина ERV в п.н., по оси ординат – покрытие ридами ChrmC.

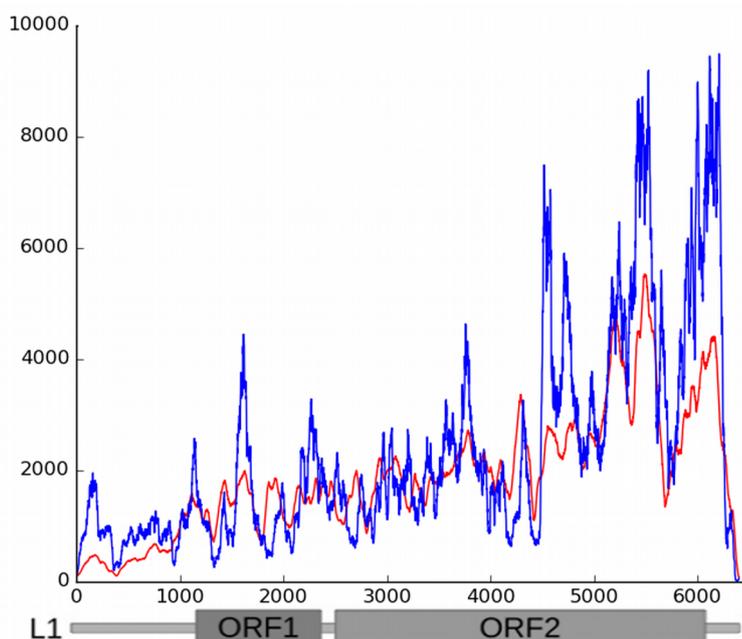


Рисунок 3. Покрытие ридами ChrmC консенсусной последовательности L1_MM из RepBase. Синяя линия – риды ChrmC, красная линия – риды wgHTS нормализованные на размер набора данных. По оси x – последовательность LINE в п.н., по оси Y – покрытие каждого нуклеотида ридами.

гетерохроматиновым областям. Если в перестроенной и сильно малигнизированной

линии L929 встречаются дисперсные сигналы вне гетерохроматиновых областей, то в клетках костного мозга с нормальным и неперестроенным кариотипом в области первичной перетяжки всех хромосом наблюдается сильный сигнал. Проведен fiber-FISH с тем же зондом. На fiber-FISH видно, что зонд к IAP присутствует на границе поля МиСат, отчасти перекрываясь с ним, т.е. в периЦЕН районе (Ostromyshenskii et al., 2018a).

Совокупность данных говорит о том, что ERV (особенно IAP (принадлежат к классу ERV2) являются регулярным компонентом конститутивного гетерохроматина.

5. Фрагмент LINE в хромоцентрах.

Для определения обогащения конститутивного гетерохроматина отдельными фрагментами LINE мы картировали риды ChrмС на консенсусную последовательность L1_MM из RepBase и сравнили с распределением ридов wgHTS (Рис. 3). В отличие от ERV, покрытие LINE ридами ChrмС сильно неравномерно с несколькими пиками и значительно большим покрытием в ~2 т.п.н. области 3' конца 2-го ORF и 3'-UTR (Рис. 3, синяя линия). В случае ридов wgHTS (Рис. 3, красная линия) также наблюдается большее покрытие ~2 т.п.н. фрагмента 3' конца 2-го ORF и 3'-UTR, однако разница между покрытием этой части остальной части L1 выражено не так сильно. Вероятно, именно этот фрагмент характерен для конститутивного гетерохроматина (Ostromyshenskii et al., 2018a).

6. Экспериментальное доказательство наличия фрагмента LINE в конститутивном гетерохроматине.

Для экспериментального доказательства наличия фрагмента LINE в гетерохроматине исследованы клоны, полученные из ДНК хромоцентров и центромер. Для работы использовали диссекционные пробы первичной перетяжки хромосом мыши, а также биохимически выделенная ДНК хромоцентров. ДНК хромоцентров и первичной перетяжки, амплифицированная с помощью DOP-праймера клонирована в плазмиду pCR2.1 TOPO. Полученные клоны секвенированы, последовательности доступны в GenBank (Kuznetsova, Ostromyshenskii et al., 2016). Всего исследовано 38 клонов. 18 из этих клонов имеют сходство с фрагментами LINE. 14 из них относятся к участку 3' конца ORF 2 и 3' — UTR L1_MM. Провели FISH с 7 мечеными клонами. Предсказание *in silico* предполагает дисперсное распределение этих зондов по хромосомам, однако *in situ* только 1 показал распределение сигнала, характерное для диспергированных повторов, а 6 остальных дают сигнал преимущественно в области первичной перетяжки всех аутосом и X хромосомы, а один клон также и в первичной перетяжке Y хромосомы. Пример гибридизации для клона С7 приведен (Рис. 4А).

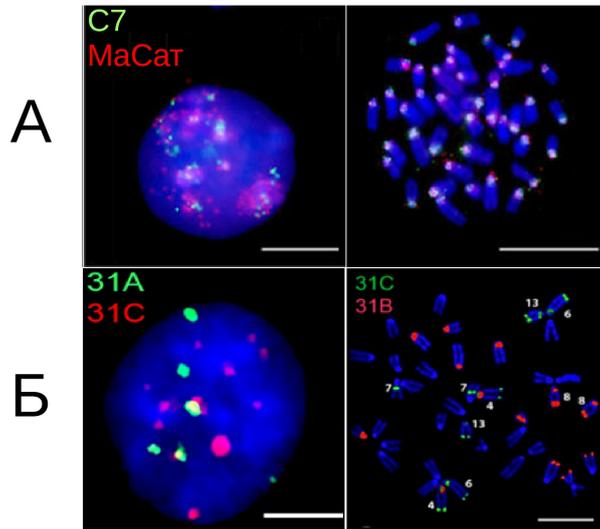


Рисунок 4. Гибридизация ТП и клонов ДНК хромоцентров на ядрах и хромосомах мыши. А — гибридизация С7 (зеленый) и МиСат (красный) на ядрах и хромосомах мыши. Б — гибридизация зондов к ТП TR-31А (зеленый) и TR-31С (красный) на ядрах мыши и зондов к ТП TR-31С (зеленый) и TR-31В (красный) на хромосомах мыши. Хромосомы контрастированы DAPI. Размерная линейка 10 мкм.

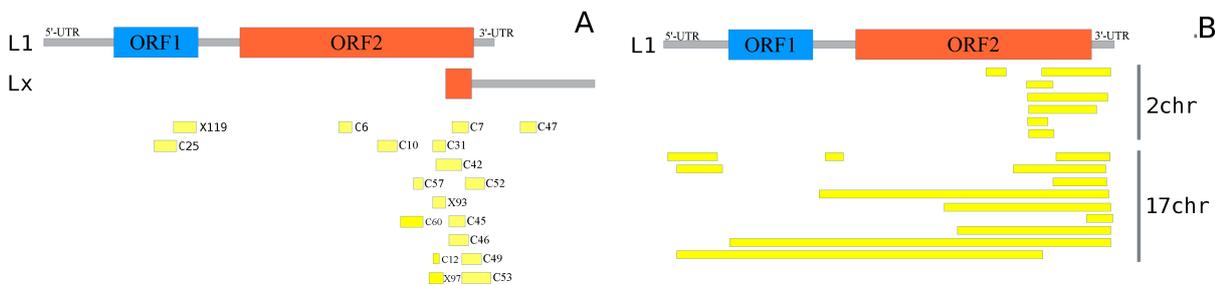


Рисунок 5. Рисунок 8. А – схема позиций клонов, содержащих фрагменты LINE на последовательности L1_MM и Lxs В – Фрагменты LINE, найденные в центромерах двух хромосом человека, картированные на L1.

Обогащение тем же фрагментом найдено и для недавно опубликованных ЦЕН человека (Рис.5, Остромышенский и др., 2015). Следовательно, не полноразмерный LINE, а его конкретный фрагмент характерны для ЦЕН/периЦЕН района человека и мыши.

7. Тандемные повторы в ДНК хромоцентров (ChrmC)

Таблица 2. Содержание тандемных повторов в ДНК хромоцентров

№	Семейство ТП	%	№	Семейство ТП	%
1	MaCat	66.22	18	TR-1870A-MM	0.015
2	MiCat	4.42	19	TR-6A-MM	0.015
3	TR-1908A-MM	0.310	20	TR-4B-MM	0.011
4	Tel	0.105	21	TR-22A-MM	0.006
5	TR-23A-MM	0.091	22	TR-1384A-MM	0.004
6	TR-57A-MM	0.076	23	TR-48A-MM	0.004
7	TR-31C-MM	0.065	24	TR-1521A-MM	0.002
8	TR-4A-MM	0.065	25	TR-31D-MM	0.001
9	TR-54A-MM	0.054	26	TR-814A-MM	0.001
10	TR-54B-MM	0.044	27	TR-234A-MM	0.001
11	TR-84A-MM	0.043	28	TR-24B-MM	0.001
12	TR-31A-MM	0.039	29	TR-36A-MM	0.001
13	TRPC-21A-MM	0.036	30	TR-24A-MM	0.001
14	TR-31B-MM	0.034	31	TR-29B-MM	0.001
15	TR-93A-MM	0.031	32	TR-29A-MM	0.001
16	TR-1149A-MM	0.023	33	TR-13A-MM	0.001
17	TR-18A-MM	0.019	34	TR-1164A-MM	0.001

Количество отдельных классов повторов подсчитывали по выравниванию рядов ChrMС на консенсусные последовательностях из Repbase. Известно, что последовательности ТП весьма вариабельны. Ранее проведенный поиск и анализ ТП в геноме мыши (Komissarov et al., 2011) позволил провести оценку содержания всех ТП и их отдельных семейств в ChrMС. Общее количество ТП, подсчитанных по полным полям ТП, составило 74,9% из них 68.6% приходится на долю MaCat, 6.2% - MiCat. Содержание остальных семейств ТП составляет 0.88% (Табл. 2).

Около 30 ТП из исходной классификации ТП обнаружены в ChrMС в количествах от 0.3% до 0.001%. Сравнение с содержанием в контигах WGS показывает, что почти все MaCat и MiCat генома находятся в хромоцентрах, но из вновь классифицированных семейств ТП в хромоцентрах находится не более половины семейств ТП (Ostromyshenskii et al., 2018a).

8. Сложный состав хромоцентров на основании позиций ТП

Синтезировали 13 олигонуклеотидных проб для найденных *in silico* семейств ТП по мере убывания их количества в геноме. 11 успешно гибридизовались и обнаружены в хромоцентрах или около них (Остромышенский и др., 2015). Ни одна из проб не показала специфичности к единственной паре хромосом; но двойной FISH позволяет идентифицировать некоторые хромосомы без трудоемкой процедуры кариотипирования (Рис. 4Б, правая панель). В интерфазных ядрах клеток первичной

культуры фибробластов при двойном FISH видно, что хромоцентры неоднородны— зонды к разным семействам ТП детектируются в разных частях хромоцентров (Рис. 4Б, левая панель). Ни один из сигналов зондов к «новым» семействам ТП не перекрывается с сигналом к МиСат. Даже если зонды дают сигнал во внутренней части хромоцентров, они почти не перекрываются между собой (Рис. 4Б, левая панель). Семейства ТП, выявленные при секвенировании действительно принадлежат хромоцентрам и занимают в них отдельные области.

9. ТП у разных родов мышевидных грызунов

Гибридизация метафазных хромосом трех родов мышей (*Mus*, *Apodemus* и *Sylvaemus*) с диссекционной центромерной пробой *M.musculus* показала, что если у *M.musculus* сигнал целиком покрывает область первичной перетяжки на хромосомах, то на хромосомах очень близкого вида *M. spicilegus* наблюдается меньшая степень гибридизации в районе ЦЕН/периЦЕН, а также полное отсутствие гибридизационного сигнала на Y хромосоме (Подгорная, Остромышенский и др., 2009; Остромышенский и др., 2011). На хромосомах филогенетически более далекого вида *M. caroli* гибридизационный сигнал еще слабее, чем у *M. spicilegus*. У представителей других родов наблюдается сигнал по плечам хромосом и полное отсутствие гибридизации в области первичной перетяжки. Гибридизация отдельных проб ТП показывает сходную картину. На хромосомах *M. spicilegus* сигналы дают все из проверенных зондов к ТП (МиСат, МаСат, TR-31A, TR-31B, TRPC-21A, TR-24B, TR-38C), но на меньшем числе хромосом, чем в случае *M. musculus*, на хромосомах *M. caroli* только МаСат, TR-31A и TR-31B. У представителей других родов гибридизационный сигнал к ТП, найденным в геноме мыши отсутствует (Остромышенский и др., 2011; 2015). Аналогичная картина наблюдается и при гибридизации проб, полученных путем хроматин-иммунопреципитации с антителами против конститутивного белка центромера CENP-A (ChpA). Проба ChpA *M. musculus* дает сигнал в области первичной перетяжки только у представителей рода *Mus*, а аналогичная проба, полученная из хроматина *A. peninsulae* — сигнал в области первичной перетяжки только у представителей рода *Apodemus* (Ostromyshenskii et al., 2018b).

В базах данных сборок геномов и SRA доступны несколько геномов и исходных прочтений геномов нескольких видов мышевидных грызунов. Для анализа содержания и сравнения ТП мы выбрали геномы мышей *M. musculus* и *M.caroli*, хомяков *Cricetulus griseus* и *Mesocricetus auratus*. Анализ ридов генома *M.caroli* (ERP000243) показал, что наиболее представленным из всех исследованных ТП является МаСат (~1% от всех последовательностей), вторым по количеству в геноме является ТП TR-31B (~0,2%). Количество другого обнаруженного с помощью FISH повтора – TR-31A в геноме *M.caroli* около 0,001%. Кроме этих ТП в геноме *M. caroli* найдено еще 14 ТП, сходных с ТП домашней мыши. Последовательности схожие с

известными ТП *M.caroli* — sat79 и sat60 (Kipling et al., 1995) в геноме *M. musculus* не найдены, хотя для генома *M.caroli* это мажоры (Табл. 3, левая панель).

Таблица 3. Сравнение содержания семейств ТП в геномах *M.musculus* и *M.caroli*; *C. griseus* и *M. auratus*.

Семейство ТП	<i>M. musculus</i>	<i>Mus caroli</i>	Семейство ТП	<i>C. griseus</i>	<i>M. auratus</i>
MaCat	13,5587	0,6627	TR-49A-CG	0.9317	0
МиCat	4,5683	0,0001	TR-11A-CG	0.6038	0
TR-1590A-MM	2,2426	0,0356	TR-79A-CG	0.4107	0
TR-6A-MM	0,2513	0,1548	TR-272A-CG	0.3422	0
TR-31B-MM	0,2155	0,1917	TR-33A-CG	0.2756	0
TR-107A-MM	0,0392	0,037	TR-6A-CG	0.1614	0
TR-21A-MM	0,0445	0	TR-25B-CG	0.1497	0
TR-194A-MM	0,0013	0	TR-304A-CG	0.1457	0
TR-57A-MM	0,0038	0,044	TR-77A-CG	0.0864	0
TR-13A-MM	0,0005	0	TR-17A-CG	0.0383	0
TR-93A-MM	0,0207	0	TR-49A-MA	0	3.1229
TR-30A-MM	0,0004	0	TR-44A-MA	0	0.6156
TR-31A-MM	0,0007	0	TR-85A-MA	0	0.071
TR-1149A-MM	0,0011	0	TR-42A-MA	0	0.0704
TR-22A-MM	0,0026	0	TR-161A-MA	0	0.0697
TR-1526A-MM	0,0004	0	TR-62A-MA	0	0.059
TR-24B-MM	0,0022	0	TR-32A-MA	0	0.0429
TR-27A-MM	0,0007	0	TR-163A-MA	0	0.0324
sat60	0	11	TR-73A-MA	0	0.0313
sat79	0	5	TR-1743A-MA	0	0.0048

Содержание ТП в геномах 4 видов мышевидных грызунов определено путем выравнивания ридов высокопроизводительного секвенирования на все поля (или мономеры для sat60 и sat 79) соответствующего семейства ТП.

В сборках геномах двух видов разных родов хомяков *Cricetulus griseus* и *Mesocricetus auratus* нашли семейства больших ТП (Михеев, ..., Остромышенский, 2015). Всего в сборках *C. griseus* найдено около 100 семейств ТП с длиной поля > 3 т.п.н., в сборке генома *M.auratus* только 19. При этом геном *M.auratus* содержит ярко выраженный мажорный ТП — TR-49A-MA, а в геноме *C. griseus* мажорный ТП обнаружить не удалось. Эти виды принадлежат к филогенетически близким родам, однако сравнение *in silico* ТП в их геномах показало, что последовательности ТП у этих видов разные и не обнаружены в геноме другого вида (Табл. 3, правая панель). Геномы мышей собраны в разной степени и это могло бы объяснять различия в наборе ТП, но геномы хомяков имеют один порядок критерия успешности сборки N50 (см. Материалы и методы), а различия в наборах ТП более показательны. Ранее были известны единичные ТП этих видов, которые, не имеют ничего общего между собой

по критериям BLAST; наши данные показывают, что полностью отличаются и полные наборы ГП.

ОБСУЖДЕНИЕ

Районы хромосом, образованные конститутивным гетерохроматином (ГХ) остаются несобранными, более того, не известен точный набор последовательностей, которые составляют эти функционально значимые районы хромосом. Отсутствие данных по составу ГХ приводит к трудностям в определении функций и особенностей организации. Данные, полученные на клеточных линиях мыши, показали, что кинетохор и периЦЕН районы хромосом мыши занимают разные домены хромоцентров (Guenatri et al., 2004). ПериЦЕН MaCat занимает центральную часть хромоцентров, в то время как ЦЕН MiCat располагается на периферии. Эти результаты предполагали достаточно однородную организацию хромоцентров при использовании ограниченного количества гетерохроматиновых зондов (MaCat и MiCat). Наши данные с использованием большего количества проб показывают, что хромоцентры устроены намного сложнее, чем считалось ранее. Как и для ядра (Cournac et al., 2016), для них характерна компертментализация сходных по последовательности областей.

Отсутствие методов для сборки полных последовательностей районов ГХ заставило нас объединить все доступные подходы для определения состава ДНК хромоцентров. Результаты работы дали ряд последовательностей, которые могут позволить прояснить вопрос об ассоциации хромоцентров и могут помочь разобраться в уже существующих данных, полученных, например, с использованием методики ЗС. В нашей работе мы нашли последовательности хромоцентров, специфичные исключительно для X-хромосомы, а также последовательности присутствующие и на X и на Y хромосомах. Они многократно диспергированно повторены на половых хромосомах. Эти последовательности могут быть использованы для разработки специфичных для X-хромосомы или для обеих половых хромосом гетерохроматиновых зондов.

Диспергированные повторы в хромоцентрах

ERV

ERV составляют порядка 10% генома мыши (Stocking, Kozak, 2008). Однако, представленность разных классов ERV отличается в геноме и в хромоцентрах. Наиболее распространенным в геноме мыши является MaLR-подобные ERV, принадлежащие к классу ERV3 (Smith, 1993). В геноме мыши наиболее представлены MaLR-подобные элементы MERVL, MTA и ORR1 (Lander et al., 2002). С другой стороны, наиболее активными в геноме мыши являются два типа ERV2 — IAP и ETn.

Около 15% спонтанных мутантов у домового мыши связаны с инсерцией IAP или ETn (Lander et al., 2002). В ChrмС мы нашли и MERVL и MTA, но преобладающим ERV является IAP.

Среди выявленных *in silico* семейств два семейства ТП мыши имеют сходство с диспергированными повторами (Komissarov et al., 2011). Одно из этих семейств образовано эндогенным ретровирусом MTA, который был найден нами в хромоцентрах, а также в полноразмерном виде в контигах, содержащих МиСат и МаСат. Однако тандемно-повторяющиеся MTA в сборках генома мыши обнаружены только по плечам хромосом (Komissarov et al., 2011). Анализ контигов WGS показал, что в сборке геномов присутствуют контиги в которых одновременно встречаются поля ТП и фрагменты ERV (в первую очередь IAP) и/или их полноразмерные LTR. Это относится и к ЦЕН ТП (МиСат) и к периЦЕН ТП (МаСат и TRPC-21A). Среди контигов с МаСат и TRPC-21A есть содержащие практически полноразмерные IAP, ограниченные LTR, либо контиги, содержащие фрагменты IAP, не образующие полной последовательности, но расположенные в правильном порядке. С учетом этих фактов, а также анализа покрытия ридями хромоцентров последовательностей ERV из Rerbase и картирования, полученных нами контигов на последовательности ERV мы предполагаем, что в конститутивном хроматине содержатся, по крайней мере крупные фрагменты ERV.

Известны транскрипты, состоящие из фрагмента ERV и ТП, являющиеся основой для образования одного из классов малых РНК. Подобные транскрипты известны для разных классов позвоночных животных (Carone et al., 2009; Ferreri et al., 2011) и растений (Neumann et al., 2007). LTR ERV являются сильными промоторами и в некоторых случаях могут запускать транскрипцию в двух направлениях (Dunn et al. 2006; Cowley, Oakey, 2012).

Мы не нашли обогащения конститутивного гетерохроматина ERV по сравнению с остальным геномом, но показали асимметрию расположения: если в геноме в целом наиболее распространенными являются MaLR-подобные элементы (принадлежат к ERV3), то в ГХ преобладают IAP и соответствующие им LTR (класс 2). Найденные в ГХ ERV и особенно их LTR могут выступать промоторами транскрипции для окружающих их ТП (Подгорная, Остромышенский и др., 2018).

LINE

Цитологически показано, что мажорные семейства диспергированных повторы в геноме мыши (LINE и SINE) занимают разные участки, которые на метафазных хромосомах относятся G и R бендам, соответственно (Boyle et al., 1990). LINE присутствуют в активных центромерах хромосом человека (Schueler et al., 2001), нецентромерах (Chueh et al., 2009), в комбинации с другими диспергированными повторами (Carbone et al., 2009). С другой стороны, анализ последовательностей в

сборках генома показывает, что LINE характерны для AT-богатых (G-позитивных), SINE для GC-богатых (R-позитивных) районов хромосом (Waterston et al., 2002). Зонды к MaCat, L1 (наиболее распространенный из LINE) и B1 (наиболее распространенный у грызунов SINE) на FISH маркируют те же районы хромосом, соответственно, C-бэндинг, G-бэндинг и R-бэндинг при дифференциальном окрашивании хромосом цитогенетическими методами; в области ЦЕН/периЦЕН зонды к полноразмерному L1 сигнала не дают (Solovei et al., 2009). Этот парадокс может быть разрешен если предположить, что в конститутивном гетерохроматине преобладают не полноразмерные LINE, а их отдельные фрагменты.

Наши результаты, полученные разными методами – анализом ридов высокопроизводительного секвенирования ДНК хромоцентров, анализом клонов, полученных из этой ДНК и их гибридизацией на хромосомах мыши, показывают, что хромоцентры обогащены ~2 т.п.н. фрагментом 3'-конца LINE. Одно из семейств ТП, родственных диспергированным повторам из исходной классификации ТП состоит из фрагментов LINE, той же природы, т.е. из фрагмента длиной ~2 т.п.н., который включает в себя 3'-конец ORF2 и 3'-некодирующий регион (Komissarov et al., 2011).

Аналогичный фрагмент нашли при анализе сборки центромера человека (Рис. 5). Мы использовали сборку центромеров человека, предсказанную биоинформатическими методами (Miga et al., 2015; LinearCen 1.1, GCA_000442335.2). Картирование фрагментов LINE найденных в центромерах двух хромосом человека из этой сборки на консенсус L1 из RepBase показало обогащение аналогичным ~2 т.п.н. фрагментом 3'-конца LINE. Т.о., именно этот фрагмент характерен для ГХ человека и мыши.

Тандемные повторы в ДНК хромоцентров

Исторически ТП считали «мусорной ДНК». Только в последние годы с открытием функциональной значимости транскриптов некоторых ТП отношение к этой части генома стало меняться. Как правило, ТП расположены в ЦЕН и периЦЕН районах хромосом, быстро эволюционируют, но при этом сохраняют свои функции в кинетохоре (Podgornaya et al., 2003). Несмотря на различия в последовательности центромерные ТП как правило имеют общие черты — организацию в длинные гомогенные поля и длину мономера, около размера нуклеосомной ДНК, склонность областей с ТП к формированию нетривиальных вторичных структур ДНК (Pavlek et al., 2015).

Наши данные показывают, что почти все MaCat и MiCat генома входят в состав хромоцентров. Остальные же семейства ТП вместе составляют лишь ~1% от состава хромоцентров, а 29 семейств из 62 семейств ТП вообще не найдены в хромоцентрах, что позволяет предположить их расположение на плечах хромосом. О наличии ТП на плечах хромосом сообщали у домового мыши (Podgornaya et al., 2013),

человека (Ames et al., 2008; Warburton et al., 2008) и жука *Tribolium castaneum* (Pavlek et al., 2015). In silico анализ собранного генома *Tribolium castaneum* показал наличие в эухроматиновой части плеч хромосом ТП с характерной длиной мономера 170 п.н. и количеством повторов в поле 5 и более (Pavlek et al., 2015). Анализ эухроматиновой части генома человека, показал широкий диапазон ТП с разной длиной мономеров, от нескольких п.н. до нескольких т.п.н., образующих поля ТП (Warburton et al., 2008).

Определение функций расположенных по плечам хромосом полей ТП требует дальнейших исследований, но точный состав ТП хромоцентров является одним из результатов настоящей работы.

Эволюция тандемных повторов у мышевидных грызунов

Белки, участвующие в формировании хроматина ЦЕН районов хромосом консервативны у разных групп животных. Последовательности ТП, с которыми связаны эти белки очень разнообразны и не имеют ничего общего по критериям BLAST. Разрешение этого парадокса возможно только путем исследования эволюции ТП в разных группах организмов и определение того, какие именно черты придают ТП функциональное значение. Имеющиеся в базах данных сборки геномов позволяют поставить вопрос о видоспецифичности и эволюции ТП у мышевидных грызунов.

Вопросы видоспецифичности и эволюции ТП изучены крайне слабо. Наиболее популярной моделью эволюции ТП является библиотечная гипотеза, предложенная почти 40 лет назад (Fry, Salser 1977). Основная идея гипотезы состоит в том, что в геноме предкового вида есть пул последовательностей, из которых у потомков образуются новые ТП путем выборочной амплификации, а последовательности, не подвергшиеся амплификации, остаются в низкокопийном состоянии. Экспериментально библиотечная гипотеза проверена на жуках нескольких родов (Mravinac, Plohl, 2007; Palameque et al., 2005; Meštrović et al., 1998).

В 2013 году опубликованы результаты масштабного исследования (Melters et al., 2013), главная цель которого поиск мажорного (по содержанию в геноме) ТП для каждого исследованного вида. Всего исследованы мажорные ТП у 282 видов организмов (78 представителей царства растений и 204 представителя царства животных). Почти все исследованные виды принадлежали к разным родам. В работе учитывали последовательности мономера ТП, их длину, GC-контент и их количество в геноме. Показано, что сходство этих параметров наблюдается только у очень близких организмов. Выводы работы основываются только на сравнении последовательностей мажорного сателлита каждого вида, но, например, для *Mus caroli* ранее показано in situ (Siracusa et al., 1983) и подтверждено нами in silico присутствие в геноме последовательностей сходных с MaCat домовой мыши, но мажорным по содержанию в геноме является другой ТП.

Известные ЦЕН и периЦЕН tandemные повторы *M. musculus* (МиCat и МаCat), *M. caroli* (sat79 и sat 60) (Kipling et al., 1995) и *A. peninsulae* (Matsubara et al., 2008) сильно отличаются друг от друга. При анализе *in silico* имеющихся сборок геномов *M. musculus* и ридов полногеномного секвенирования *M. caroli* у последнего вида найдены только несколько ТП, характерных для *M. musculus*, большинство ТП, характерных для *M. musculus* в геноме *M. caroli* не найдено (Табл. 3). *In situ* показано, что внутри рода *Mus* только у очень близкого вида *M. spicilegus* присутствуют те же ТП, что и у *M. musculus*. Еще более явная картина различий наборов ТП наблюдается в случае хомяков разных родов *C. griseus* и *M. auratus*, в геномах которых отсутствуют последовательности схожие с ТП другого вида (Табл. 3). Таким образом, у исследованных животных отличаются не только мажорные ТП, но и весь набор ТП. Отсутствуют даже единичные копии.

ГХ остается «темной материей» генома. Его состав и функции до сих пор остаются крайне слабо изученными. В последнее время началось активное изучение lncRNA, которые в основном считываются с повторов различных типов, в том числе входящих в состав гетерохроматина. В настоящей работе проведено как исследование состава контитутивного гетерохроматина домашней мыши, так и главного его компонента — ТП. Показано, что эволюция ТП у мышевидных грызунов не укладывается в рамки «библиотечной» гипотезы. Найдено необычное распределение ТП у китайского хомячка. Род *Cricetulus* характеризуется быстрой кариотипической эволюцией, что делает его объектом для дальнейших исследований эволюции ТП и роли ТП в эволюции.

ВЫВОДЫ

1. ДНК хромоцентров домашней мыши секвенирована на платформе Illumina MiSeq (ChrmC). Наиболее представленными в ДНК хромоцентров являются tandemные повторы — мажорный (~69%) и минорный (~4%) сателлиты, остальные tandemные повторы составляют ~1%. Около 6% ChrmC относятся к неаннотированной повторяющейся ДНК.
2. Среди неаннотированных последовательностей выявлены последовательности специфичные для половых хромосом, которые являются не охарактеризованными диспергированными повторами.
3. ERV составляют ~9% ДНК хромоцентров; наиболее представлены элементы IAP (ERV2 класс) и их LTR. Наличие IAP в конститутивном гетерохроматине доказано *in situ*.
4. Показано *in silico* и доказано *in situ*, что LINE в хромоцентрах представлен ~2 т.п.н. фрагментом 3'-конца ORF2 и 3'-некодирующего региона.
5. Хромоцентры являются сложно устроенными структурами, а не гомогенными, как считалось ранее.

6. Эволюция ТП в двух группах мышевидных грызунов не укладывается в рамки библиотечной гипотезы.

СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи

1. Подгорная О. И., **Остромышенский Д. И.**, Кузнецова И. С., Матвеев И. В., Комиссаров А. С. (2009). Парадоксы организации центромера и гетерохроматина. Цитология, 51(3), 204-211.
2. **Остромышенский Д. И.**, Кузнецова И. С., Голенищев Ф. Н., Маликов В. Г., Подгорная О. И. (2011). Применимость сателлитной ДНК как филогенетического маркера на примере трех родов подсемейства Murinae. Цитология. 53 (7), 564–571
Ostromyshenskii, D. I., Kuznetsova, I. S., Golenishev, F. N., Malikov, V. G., & Podgornaya, O. I. (2011). Satellite DNA as a phylogenetic marker: Case study of three genera of the murine subfamily. Cell and Tissue Biology, 5(6), 543-550. (перевод)
3. Михеев Д.Ю., Подгорная О.И., **Остромышенский Д.И.** (2015). Большие тандемные повторы сирийского хомячка *Mesocricetus auratus* in silico и in situ. Цитология, 57 (2), 95–101
Miheev, D. Y., Podgornaya, O. I., Ostromyshenskii, D. I. (2015). Large tandem repeats of *Mesocricetus auratus* in silico and in situ. Cell and Tissue Biology, 9(3), 226-232. (перевод)
4. **Остромышенский Д.И.**, Кузнецова И.С., Комиссаров А.С., Картавцева И.В., Подгорная О.И. (2015). Тандемные повторы геномов мышевидных грызунов в базах данных и их картирование. Цитология, 57 (2), 102–110
Ostromyshenskii D. I., Kuznetsova I. S., Komissarov A. S., Kartavtseva I. V., Podgornaya O. I. (2015). Tandem repeats in the rodent genome and their mapping. Cell and Tissue Biology, 9(3), 217-225. (перевод)
5. **Остромышенский Д.И.**, Комиссаров А.С., Кузнецова И.С., Черняева Е.Н., Вайсертрейгер И.Р., Подгорная О.И. (2016). Состав ДНК хромоцентров мыши in silico и in situ. Фрагменты LINE и ERV - обязательный компонент ДНК хромоцентров кроме тандемных повторов. Цитология, 58 (5), 389–392
6. Kuznetsova I. S., **Ostromyshenskii D. I.**, Komissarov A. S., Prusov A. N., Waisertreiger I. S., Gorbunova A. V., Podgornaya O. I. (2016). LINE-related component of mouse heterochromatin and complex chromocenters' composition. Chromosome Research, 24(3), 309–323
7. Подгорная О.И., **Остромышенский Д.И.**, Енукашвили, Н.И. (2018). Кому он нужен, этот мусор, или темная материя генома. Биохимия, 83(4), 610-628

8. **Ostromyshenskii, D.I.**, Chernyaeva E.N., Kuznetsova I.S., Podgornaya O.I. (2018) Mouse chromocenters DNA content: sequencing and *in silico* analysis. BMC Genomics, 19, 159. DOI: 10.1186/s12864-018-4534-z
9. **Ostromyshenskii D**, Kuznetsova I, Podgornaya O, Kartavtseva I. (2018) Appearance of B Chromosomes like Structures in Apodemus Peninsulae Primary Cell Culture. Research Journal of Zoology, 1:1, DOI: 10.4172/RJZ.1000105
10. Лебедев Е.Е., **Остромышенский Д.И.**, Соловьева А.И., Туренко А.С., Подгорная ОИ, Адонин ЛС. (2018) Транспозоны морского ежа *Strongylocentrotus intermedius*: *in silico* versus *in vitro*: определение методами биоинформатики и доказательство присутствия в геноме). Биология моря, в печати

Тезисы

1. **Д. И. Остромышенский**, И. С. Кузнецова, И. В. Картавцева, О. И. Подгорная. Исследование некоторых аспектов эволюции центромерного района хромосом у представителей трех родов Murinae// Хромосомы и эволюция. Симпозиум памяти Григория Андреевича Левитского. — 2008. — С. 78–79.
2. **D. I. Ostromyshensky**, O. I. Podgornaya, F. N. Golenichev et al. Satellite dna in genomes of the subfamily Murinae// 11th International Conference Rodens Et Spatium on Rodent Biology. Myshkin, Russia July 24-28, 2008.
3. **Остромышенский Д.И.**, Комиссаров А.С., Подгорная О.И. Тандемные повторы *in silico* и *in situ* у трех видов рода Mus//Современные проблемы биологической эволюции (Москва, 11-14 марта 2014)
4. **Dmitrii Ostromyshenskii**, Olga Podgornaya The genome wide analysis of the large tandem repeats in the closely related species//Седьмая Московская международная конференция по вычислительной молекулярной биологии MCCMB'15 Москва, 16-19 июля 2015 Сборник трудов. <http://mccmb.belozersky.msu.ru/2015/proceedings/abstracts/18.pdf>
5. **Ostromyshenskii D.I.**, Komissarov A.S., Kuznetsova I.S., Chernyaeva E.N., Vaysertreyger I., Podgornaya O.I. Mouse chromocenters' DNA content *in silico* and *in situ*. Line fragment and ervs are an essential chromocenters' components beside tandem repeats//Материалы международной конференции Хромосома-2015. Новосибирск, 2015. с. 42-43.
6. **Остромышенский Д.И.**, Подгорная О.И. Большие тандемные повторы в геномах млекопитающих *in silico* и *in situ*//Материалы международной конференции Хромосома-2015. Новосибирск, 2015. с. 133–134.
7. **D.I. Ostromyshenskii**, O.I. Podgornaya. The genome wide analysis of the large tandem repeats in the closely related genomes//The tenth international conference on bioinformatics of genome regulation and structure / systems biology. — Novosibirsk, Russia, 2016.

8. D.I. Ostromyshenskii, A.S. Komissarov, I.S. Kuznetsova, O.I. Podgornaya. In silico mouse chromocenters content//The tenth international conference on bioinformatics of genome regulation and structure / systems biology. — Novosibirsk, Russia, 2016.

Благодарности

Автор выражает огромную благодарность всем нынешним и бывшим сотрудникам группы Некодирующей ДНК Института Цитологии РАН и в особенности Вере Николаевне Стефановой, за помощь в проведении FISH. Сотрудникам Центра Геномной Биоинформатики им. Добржанского СПбГУ Екатерине Николаевне Черняевой за проведение секвенирования ДНК хромоцентров и Алексею Сергеевичу Комиссарову. Зав. лабораторией сравнительной геномики ИМКБ СО РАН Владимиру Александровичу Трифонову за предоставление пробы первичной перетяжки хромосом мыши и огромную помощь при написании англоязычных статей. Сотруднику ФНЦ Биоразнообразия ДФО РАН Ирине Васильевне Картавцевой.

Отдельную благодарность автор выражает своему первому научному руководителю — Инне Сергеевне Кузнецовой за обучение молекулярно-биологическим методикам и всестороннюю помощь и нынешнему научному руководителю — Ольге Игоревне Подгорной за всестороннюю помощь, понимание и поддержку как в научных, так и в житейских вопросах.

Цитированная литература

Wijchers et al., 2015, *Genome research*, 25(7): 958-969; Probst, Almouzni 2008, *Differentiation*, 76(1): 15-23; Mayer et al., 2005, *BMC cell biology*, 6(1): 44; Guenatri et al., 2004, *The Journal of cell biology*, 166(4): 493-505; Snapp et al., 2013, *Journal of anatomy*, 223(3): 255-261; de Koning et al., 2011, *PLoS Genet*, 7(12): e1002384; Ribet et al., 2008, *Genome research* 18(4): 597-609; Magiorkinis et al., 2012, *PNAS*, 109(19): 7385-7390; Erukashvily et al., 2009, *Cytogenetic and genome research*, 124(3-4): 277-287; Probst et al., 2010, *Developmental cell*, 19(4): 625-638; Tanne et al., 2015., *PNAS*, 112(49): 15154-15159; Bersani et al., 2015, *PNAS*, 112(49): 15148-15153; Guo, Wu 2008, Protocol Exchange :10.1038/nprot.2008.164; Prusov et al., 2002, *Biochemistry (Moscow)*, 67(4): 423-431; Zatsepina et al., 2008, *The Nucleus: Volume 1: Nuclei and Subnuclear Components*: 169-180; Arneson et al., 2008, *Cold Spring Harbor Protocols*, 2008(1): pdb-prot4919; Sambrook J., Russell D. W. *Molecular Cloning: A Laboratory Manual*, 2001; Altschul et al., 1990, *Journal of molecular biology*, 215(3): 403-410; Bolder et al., 2014, *Bioinformatics*: btu170; Ding et al., 2008, *PLoS one*, 3(3), e1709; Yang et al., 2005, *Genomics*, 86(3): 306-315; Miga et al., 2015, *Genome research*, 24(4): 697-707; Komissarov et al., 2011, *BMC genomics*, 12(1): 531; Kipling et al., 1995, *Mol. Cell. Biol.* 15: 4009-4020; Cournac et al., 2016, *Nucleic Acids Research*, 44(1): 245–255 Stocking, Kozak, 2008. *Cellular and molecular life sciences*, 65(21): 3383-3398; Mouse

Genome Sequencing Consortium, 2002, *Nature*, 420(6915): 520-562 ; Carone et al., 2009, *Chromosoma*, 118(1), 113-125; Ferreri et al., 2011, 85(10): 4761-4771; Neumann et al., 2007 *Genetics*, 176(2): 749-761; Dunn et al., 2006, *Gene*, 366(2): 335-342; Cowley, Oakey RJ, 2013, *PLOS Genetics*, 9(1):e1003234. Boyle et al., 1990, *PNAS*, 87(19): 7757-7761; Solovei et al., 2009, *Cell*, 137(2): 356-368; Waterstone et al., 2002, *PNAS*, 99(6): 3712-3716.; Schueler et al., 2001. *Science*, 294(5540): 109-115; Chueh et al., 2009, *PLoS Genet*, 5(1): e1000354; Carbone et al., 2009, *PLoS Genet*, 5(6): e1000538; Podgornaya et al., 2003 *International review of cytology*, 224: 227-296; Pavlek et al., 2015, *DNA research*, 22(6): 387-401; Ames et al., 2008, *Genetics*, 179(3): 1693-1704; Warburton et al., 2008, *BMC genomics*, 9(1): 533; Fry, Salser 1977, *Cell*, 12(4): 1069-1084; Mravinac, Plohl, 2007, *Gene*, 394(1-2):45-52; Palameque et al., 2005 *Chromosome Res.*, 13(8):795-807; Meštrović et al., 1998, *Mol. Biol. Evol.*, 15(8): 1062-1068; Melters et al., 2013, *Genome Biol.*, 14(1); Siracusa et al., 1983, *Development*, 73(1): 163-178; Matsubara et al., 2008, *Chromosome Research*, 16: 1013–1026.