

УДК 575.174.015.3

НОВОЕ В ГЕНОМИКЕ ЦЕНТРОМЕР: УРОКИ ПЕРВОЙ T2T-СБОРКИ ХРОМОСОМ ЧЕЛОВЕКА

© 2023 г. Л. И. Уральский^{1, 2, *}, И. А. Александров^{2, 3, 4}, Ф. Д. Рябов⁵, А. Л. Лапидус³, Е. И. Рогаев^{2, 6}

¹Научно-технологический университет “Сириус”, Научный центр генетики и наук о жизни,
пос. гор. типа Сириус, 354340 Россия

²Институт общей генетики им. Н.И. Вавилова Российской академии наук, Москва, 119991 Россия

³Центр биоинформатики и алгоритмической биотехнологии СПбГУ,
Санкт-Петербургский государственный университет, Санкт-Петербург, 199034 Россия

⁴Федеральный исследовательский центр “Фундаментальные основы биотехнологии” Российской академии наук,
Москва, 119071 Россия

⁵Национальный исследовательский университет “Высшая школа экономики”, Москва, 109028 Россия

⁶Медицинская школа Чан Массачусетского университета, Департамент психиатрии, Шрусбери, 01655 США

*E-mail: uralskiy.li@talantiuspeh.ru

Поступила в редакцию 01.12.2022 г.

После доработки 07.12.2022 г.

Принята к публикации 08.12.2022 г.

Технологии секвенирования с длинными прочтениями, появившиеся недавно, позволили впервые прочесть полную, без пробелов, последовательность генома человека. Результатом явилась первая T2T (от теломеры до теломеры) геномная сборка, опубликованная одноименным международным консорциумом ученых в 2022 г. Наиболее значительным вкладом новой сборки явились центромерные области, состоящие из высокоповторяющейся сателлитной ДНК. В этом обзоре мы кратко перечислим основные достижения T2T-консорциума, связанные с центромерами, и более подробно рассмотрим неожиданные находки цитогенетического масштаба, которые принес анализ впервые собранных центромер человека, такие как “расщепленные” центромеры хромосом 3 и 4, мега-инверсия в активном массиве центромеры 1-ой хромосомы, гаплотипические эпипаллеты в центромере X-хромосомы и макроповторы, найденные в нескольких центромерах.

Ключевые слова: альфа-сателлит, центромера, ПВП-гаплотипы, инверсии, макроповторы, центромерные эпипаллеты, T2T-консорциум

DOI: 10.31857/S0041377123030094, EDN: VDUPJC

Геномная ДНК эукариотических организмов представлена в клетках в виде набора хромосом, которые реплицируются и поровну разделяются между дочерними клетками в каждом клеточном цикле. Центромеры — это участки хромосом, которые ответственны за их равное расхождение при делении, направляя сборку микротрубочек, связывающих кинетохоры, и служа местом адгезии между сестринскими хроматидами. Первичная структура ДНК центромер долгое время оставалась нерасшифрованной из-за сложности их структуры.

В предлагаемом обзоре изложены основные достижения, связанные с расшифровкой нуклеотидной последовательности центромер, осуществленной международным консорциумом ученых в процессе работы над первой полной сборкой референсного генома че-

ловека T2T-CHM13. В ходе проведенных исследований было получено около 200 Мб новых ранее не расшифрованных последовательностей генома человека, включающих все прицентромерные и прителомерные участки, короткие плечи пяти акроцентрических хромосом, длинные центромерные тандемные повторы, кластеры генов рибосомной РНК и другие непрочитанные ранее элементы. Расшифровка хромосом “от теломеры до теломеры” является новым достижением, после которого в геноме человека уже не осталось “белых пятен”. Высокоточное секвенирование с длинными прочтениями позволило преодолеть технологический барьер и достичь не только полноты охвата, но и беспрецедентной аккуратности полученной последовательности. Целью данного обзора является обобщение знаний, полученных в ходе анализа и аннотации центромер, а также освещение работы над проектами, связанны-

Принятые сокращения: АС — альфа-сателлит; ПВП — повторы высокого порядка; T2T — от теломеры до теломеры (telomere-to-telomere); Мб — мегабаза (миллион нуклеотидов ДНК).

ми с анализом повторов в сборках, запланированных к производству.

ОБЩИЕ СВЕДЕНИЯ О ЦЕНТРОМЕРАХ ЧЕЛОВЕКА И ПРИМАТОВ

Центромеры являются важными хромосомными структурами, которые, связываясь с комплексом центромерных белков (кинетохором), формируют места прикрепления микротрубочек веретена и, в конечном счете, обеспечивают правильную сегрегацию хромосом в митозе и мейозе. Дефекты центромер, их наследования и поддержания могут вызывать неравное распределение хромосом в дочерних клетках и нестабильность генома, приводя к раку, врожденным дефектам развития и бесплодию (Cimini et al., 2001; Levine, Holland, 2018; Vázquez-Diez, FitzHarris, 2018). Особо надо отметить важность когорты с численными нарушениями X-хромосомы в выборках пациентов с проблемами развития нервной системы (Vorsanova et al., 2021; Vorsanova et al., 2022). Несмотря на их важную клеточную функцию, детальная организация центромер человека была исключена из исходных эталонных сборок генома человека (Lander et al., 2001; Venter et al., 2001; International Human Genome Sequencing Consortium, 2004) и в значительной степени игнорировалась генетическими и геномными исследованиями в последние два десятилетия. Центромерные области и связанный с ними прицентромерный гетерохроматин обычно отмечены наличием длинных массивов почти идентичных (гомогенных) tandemных повторов, называемых также сателлитными ДНК (Singer, 1982). Эти повторяющиеся последовательности исторически недопредставлены в геномных сборках эукариот из-за неизбежных ошибок клонирования и секвенирования методами первого поколения: нестабильности в *Escherichia coli* во время клонирования на основе ВАС (бактериальные искусственные хромосомы), обилия или полного отсутствия в tandemных повторах сайтов рестрикции, используемых для клонирования, потенциальной токсичности и нестабильности клонированной ДНК (Carlson, Brutlag, 1977; Osoegawa et al., 2007). Методы второго поколения, исключившие клонирование, также не вызвали большого продвижения вперед. Они не могли надежно представить центромерные области из-за невозможности получения индивидуальных прочтений нуклеотидной последовательности достаточной длины, чтобы содержащиеся в них уникальные комбинации нуклеотидов (т. н. уникальные к-меры, которые используются для соединения отдельных прочтений в более длинные контиги) позволили определить линейную последовательность тысяч копий почти одинаковых tandemных повторов (Rudd, Willard, 2004; She et al., 2004). В результате до недавнего времени все центромерные области человека были представлены в сборках в виде огромных (миллионы п. н.) пробелов, соответствующих центромерам, про

которые было известно, что они заполнены сателлитной ДНК определенного типа (Lander et al., 2001; Venter et al., 2001; International Human Genome Sequencing Consortium, 2004).

Хотя последовательности этих регионов не были систематически представлены в первоначальных геномных сборках, они не были неизвестны. Целенаправленные экспериментальные исследования центромерных последовательностей человека показали, что все нормальные (характерные для вида, а не для отдельных клинических случаев) центромерные области образованы длинными массивами альфа-сателлитной (АС) ДНК, сложенными разнообразным классом AT-богатых tandemных ДНК повторов с мономерами длиной около 170 н. п. (Manuelidis, 1976; Manuelidis, Wu, 1978). Индивидуальные мономеры обычно организованы в более крупные единицы или повторы высокого порядка (ПВП) (Willard, Waye, 1987) и образуют один или несколько гомогенных массивов в каждой центромере человека. Экспериментальные усилия по выделению, клонированию и секвенированию репрезентативных ПВП из каждой центромерной области показали, что: (1) за несколькими исключениями все большие АС массивы хромосом-специфичны, т.е. структура и последовательность их ПВП разные в разных хромосомах (обзор Willard, 1985); (2) близкие хромосом-специфичные ПВП из разных массивов объединяются в надхромосомные семейства, отражающие недавнее общее происхождение отдельных групп хромосом-специфичных вариантов (обзор Alexandrov et al., 2001); (3) хотя в одной центромере может быть несколько разных хромосом-специфичных массивов, иногда из разных надхромосомных семейств, центромерные белки связываются только с одним из них, называемым активной центромерой, в то время как остальные ПВП-массивы можно считать неактивными центромерами или псевдоцентромерами (Mahtani, Willard, 1998; Wevrick, Willard, 1989, 1991; Schueler et al., 2001). Кроме того, в геномах человека и приматов обнаружены небольшие массивы дивергентных (в отличие от гомогенных) АС повторов. Они расположены в плечах хромосом и прилегают к гомогенным массивам, образуя дискретные и хронологически упорядоченные альфа-сателлитные слои (“мертвые” центромеры), которые представляют собой остатки центромер наших предков-приматов, потерявшие гомогенность и значительно уменьшившиеся в объеме (Rudd et al., 2006; Shepelev et al., 2009; Uralsky et al., 2019; Altemose et al., 2022a). Встречаются дивергентные массивы, как содержащие ПВП, так и лишенные их (мономерные). Средняя дивергенция копий ПВП в гомогенных массивах около 1%, в дивергентных массивах более 10%. Сформированные базы данных альфа-сателлитных повторов, извлеченные из данных секвенирования всего генома позволили произвести первые оценки частоты вариаций повторов в гомогенных массивах определенных хромосом и ранние оценки различий в по-

следовательности ДНК и длине массивов между людьми в популяции (Miga et al., 2014; Langley et al., 2019; Suzuki et al., 2020). Линейное представление наблюдаемых вариантов повторов и их расчетное количество копий в геноме HuRef (Levy et al., 2007) привели к созданию “моделированных” массивов альфа-сателлитов (референс-модели) в геномной сборке человека GRCh38/hg38, выпущенной в 2013 г., где референс-модели впервые заместили центромерные пробелы (Miga et al., 2014; Uralsky et al., 2019). Хотя эти модели не реконструируют реальную последовательность, а представляют собой своего рода мозаику, собранную из кусочков AC (в данном случае пришедших из двух несколько различающихся гомологичных хромосом), они позволяли картировать короткие прочтения на ту или иную центромеру, предсказывать возможные вариации ПВП и обнаруживать новые, отсутствующие в геноме HuRef варианты (Miga et al., 2014), а также обнаруживать некартируемые на сборку прочтения (Miga et al., 2015). Модели также предлагали улучшенное картирование последовательностей, связанных с внутренними белками кинетохора (Nechemia-Arbely et al., 2017, 2019; Dumont et al., 2020). В совокупности эти ранние исследования генома человека привели к развитию нашего первого концептуального представления о геномной организации центромер и эволюции их последовательностей в сложных геномах.

Достижения в технологиях секвенирования с длинными прочтениями (третье поколение) и недавние улучшения в методах геномной сборки повторов теперь позволяют генерировать полные и точные сборки массивов центромерных повторов человека (Jain et al., 2018a; Bzikadze, Pevzner, 2020; Miga et al., 2020; Nurk et al., 2020; Logsdon et al., 2021). Этот прогресс объясняется наличием длинных прочтений (~15–20 т.п.н.) с качеством 99.9% (Pacific Biosciences HiFi) (Wenger et al., 2019), а также “сверхдлинных” прочтений, которые обычно достигают сотен т.п.н. в длину (Jain et al., 2018a) (Nanopore UL) с точностью около 95%. Параллельно мы наблюдаем огромные успехи в автоматизированной сборке сателлитных массивов и протоколах биоинформатической оценки качества сборок (Bzikadze, Pevzner, 2020; Mikheenko et al., 2020; Nurk et al., 2020) в сочетании со стандартными методами проверки точности сборки с использованием гель-электрофореза в импульсном поле (PFGE) и Саузерн-блоттинга (Miga et al., 2020; Logsdon et al., 2021). Примечательно, что первые доступные центромерные эталонные сборки были получены из “эффективно гаплоидной” линии клеток человека, выделенной из пузырного заноса (внематочной беременности) CHM13hTERT (Steinberg et al., 2014), в котором клетки имеют две идентичные пары хромосом, что значительно упрощает задачу сборки повторов по сравнению с типичными диплоидными клеточными линиями, где гомологичные хромосомы несколько различаются. Недавние первые полные сборки двух хромосом человека

от начала до конца, или “от теломеры до теломеры” (T2T) (T2T-ChrX (Miga et al., 2020) и T2T-Chr8 (Logsdon et al., 2021)), дали первую возможность оценить эти новые методы в свете ожиданий, основанных на предыдущих экспериментальных исследованиях (Ge et al., 1992; Willard 1998; Mahtani et al., 2001). Наконец, в апреле 2022 г. была опубликована первая T2T-сборка всех аутосом и X-хромосомы из CHM13hTERT (Altemose et al., 2022a), а к концу года к этому была добавлена первая T2T-сборка Y-хромосомы из диплоидной клеточной линии HG002 (https://ncbi.nlm.nih.gov/assembly/GCF_009914755.1), что в сумме представляет собой первый полный референсный геном человека с полностью собранными центромерами. Эта монументальная работа была выполнена T2T-консорциумом исследователей (<https://www.genome.gov/about-genomics/telomere-to-telomere>), который теперь приступил к T2T сборке первого диплоидного генома клеточной линии HG002. Предварительные сборки аутосом этой линии отдельно для отцовской и материнской хромосом уже доступны на сайте консорциума (<https://humanpangenome.org/hg002/>). Налажена процедура T2T-сборки диплоидных геномов, основанная на использовании точных HiFi PacBio прочтений в комбинации с длинными, но неточными Ultra-Long Oxford Nanopore прочтениями для разрешения трудных мест и использовании геномов родителей, сделанных короткими прочтениями (Illumina), для различения (фазирования) гомологов. Другие методы фазирования используют технологии Hi-C (Lieberman-Aiden et al., 2009; Jarvis et al., 2022) и strand-seq (Ghareghani et al., 2018). Единичные оставшиеся трудные места разрешаются вручную. Короткие прочтения собственного генома также используются для окончательной “полировки” собранной последовательности. Все эти процедуры (до ручной доводки) выполняются автоматизированным сборщиком Verkko (Rautiainen et al., 2022).

Кроме преодоления технологических вызовов и успешной разработки протокола T2T-сборки диплоидного генома, деятельность T2T-консорциума ознаменовалась многочисленными открытиями и находками в области биологии и геномики центромер, а также нескольких других, сложенных различными повторами и потому ранее несобранных, регионов. Самые большие из них это кластеры классических сателлитных ДНК (Hsat) в прицентромерных областях (особенно крупные в хромосомах 1, 9, 16 и Y), кластеры рибосомных генов в коротких плечах акроцентрических хромосом и сами эти короткие плечи целиком, поскольку они насыщены классическими сателлитами и сегментными дупликациями. Эти достижения мы перечислим конспективно в следующем разделе, основываясь в основном на статьях, опубликованных в специальном выпуске журнала Science за апрель 2022 г., посвященном полной расшифровке генома человека (Altemose et al., 2022a; Gershman et al., 2022; Nurk et al., 2022).

ОСНОВНЫЕ ДОСТИЖЕНИЯ T2T-КОНСОРЦИУМА

Первая полная сборка референсного генома человека. Первая версия сборки T2T-CHM13 v1.0, включающая в себя по одной копии каждой хромосомы, кроме Y, содержит приблизительно 200 Мб новых последовательностей, по сравнению с предыдущей сборкой GRCh38/hg38. В них входят 1956 генов (gene predictions), 99 из которых, по прогнозам, кодируют белки, все центромерные повторы, недавние сегментные дубликации и короткие плечи всех пяти акроцентрических хромосом целиком (Nurk et al., 2022). T2T сборка Y-хромосомы из линии HG002, которая добавлена во вторую версию сборки T2T-CHM13 v2.0, дополняет приблизительно половину этой хромосомы (30 Мб), отсутствующую в GRCh38 сборке (https://ncbi.nlm.nih.gov/assembly/GCF_009914755.1). Эта половина содержит центромеру, классические сателлитные повторы длинного плеча, а также 110 (42 белок-кодирующих) генов, не представленных в GRCh38.

Интересной иллюстрацией необходимости полной расшифровки высокоповторяющихся областей является следующая история, изложенная в работе Rhie с соавторами (Rhie et al., 2023). При секвенировании бактериальных геномов, последовательности, представляемые в базы данных, проверяются на чужеродные примеси. Частым источником загрязнения является ДНК человека, вероятно происходящая из геномов исследователей. Проверка на присутствие ДНК человека делается путем прикладывания бактериальных прочтений к человеческой геномной сборке. Поскольку в сборке Y-хромосомы человека отсутствовали огромные гетерохроматические районы, сложенные высококопийными классическими сателлитами (HSATII), эта примесь осталась незамеченной. Используя новую T2T сборку Y-хромосомы (Rhie et al., 2023), исследователи показали, что бактериальные базы данных содержат огромное количество контигов высокого уровня, построенных из этих сателлитов, и белков, транслированных из этих последовательностей.

Результаты T2T-сборки и аннотации центромер, выполненные T2T-консорциумом, в частности можно изучать в геномном браузере (https://genome.ucsc.edu/cgi-bin/hgTracks?db=hub_3671779_hs1). В нем можно рассматривать центромеры (представленные аннотационными треками, описанными в одной из работ (Altemose N, et al., 2022a)) с высоты птичьего полета, оценивая их общую архитектуру, или, спускаясь ниже, рассматривать отдельные ПВП или мономеры. Иллюстрации, приведенные в этом обзоре, по большей части представляют собой скриншоты из этого браузера или их схематические изображения. В этой же работе представлена и новая классификация AC ПВП, которую мы тут подробно описывать не будем. Отметим только, что в имени ПВП, например, S2C4H1L, S2 означает надхромосомное семейство 2,

S4 хромосому 4, H1 означает ПВП №1 в данной хромосоме, а индекс L, что это активный массив.

Активная центромера в целом гиперметилирована, а кинетохор расположен вокруг гипометилированного участка внутри нее. Активный центромерный ПВП-массив определяется эпигенетически, связыванием с основным центромерным белком CENP-A, специальным вариантом гистона H3, встречающимся только в центромерных нуклеосомах вместо H3 (Earnshaw, Rothfield, 1985; Earnshaw et al., 1986; Palmer et al., 1987; Earnshaw, Cooke, 1989; Sullivan et al., 1994). Обычно эта связь выявляется в ChIP-seq эксперименте, где хроматин расщепляется нуклеазой, и фрагменты ДНК с нуклеосомами иммунопреципитируются антителами к CENP-A, очищаются от белка и секвенируются. В более сложном CUT&RUN эксперименте (Thakur, Henikoff, 2018) используются антитела к CENP-A, конъюгированные с нуклеазой и расщепление хроматина идет только в местах посадки CENP-A. Однако активный ПВП-массив очень велик, обычно занимает 1–3 млн п.н. и содержит тысячи ПВП. Только часть его (около 500 т.п.н.) покрыта CENP-A, вовлечена в белковые структуры кинетохора и представляет собой центрохроматин, остальная часть покрыта нуклеосомами с H3 и представляет собой гетерохроматин. Анализ профилей метилирования ДНК в центромерных сборках, полученных T2T-консорциумом, показал, что активный ПВП-массив в целом гиперметилирован по сравнению с соседними гетерохроматиновыми участками, но в районах, покрытых CENP-A, метилирование значительно снижено (“центромерный провал”). Длина “провала” обычно несколько меньше, чем весь участок, покрытый CENP-A, он видимо представляет собой район с наибольшей плотностью CENP-A нуклеосом. При этом вне “провала” они разбавляются нуклеосомами с H3. “Провал” соответствует участку центромеры с самым плотно упакованным, и наименее доступным для находящихся в растворе ферментов, хроматином (Gershman et al., 2022). Можно предполагать, что эти места просто не доступны для метилаз, ответственных за гиперметилирование в активном массиве (рис. 1). Эта модель нуждается в подтверждении и детализации. Большую роль в этом может сыграть недавно предложенный и активно используемый T2T-консорциумом метод DiMeLo-seq, который позволяет картировать места связывания CENP-A (или других белков) на сверхдлинных прочтениях ДНК (Altemose et al., 2022b).

Подтверждена модель “расширяющейся центромеры” и обнаружена “послойная экспансия” внутри активных массивов (слои ПВП-гаплотипов). Ранее, основываясь на ограниченном экспериментальном материале, была предложена модель “расширяющейся центромеры” (Miga, Alexandrov, 2021), когда в центромерном районе периодически возникает новый активный ПВП-массив, а остатки старого раздвигаются к краям и уменьшаются в результате деле-

ционного процесса, в то время как новый активный массив растет в результате дупликационного процесса (амплификации ПВП). Когда этот цикл повторяется несколько раз, формируется центромера, где огромный активный массив окаймлен слоистой структурой, содержащей прежние центромеры, где более старые слои находятся снаружи и тем меньше по размеру и более разрушены мутациями и делециями, чем больше их возраст. Вся структура в целом немного напоминает годовичные кольца деревьев, или культурные слои археологических раскопок.

Одним из последних достижений (Altemose et al., 2022a) стала систематическая (в масштабах полного генома) документация характерных черт этой послойной экспансии, которые можно суммировать следующим образом:

1) В массивах АС была показана зеркальная симметрия, начиная с ПВП-гаплотипов в активных массивах центромер (см. пункт 3), расположения сестринских ПВП (см. Miga, Alexandrov, 2021) вокруг активных ПВП, расположения неактивных ПВП вокруг активных массивов, расположения массивов дивергентных ПВП вокруг гомогенного ядра центромеры, и, наконец, симметричного расположения мономерных слоев мертвых центромер вокруг гомогенного кора, образованного различными ПВП. Все эти симметрии наблюдаются далеко не на всех индивидуальных хромосомах, но тенденция четко прослеживается, несмотря на прогрессирующее уменьшение размеров доменов в результате процесса делеции, который потенциально может разрушать и маскировать этот характерный ландшафт.

2) В неактивных массивах АС наблюдаются различные возрастные градиенты распада, причем признаки распада увеличиваются с расстоянием от активного ПВП-массива, включая: градиент размера массива, обусловленный делеционным процессом; градиент дивергенции, обусловленный накоплением мутаций, предположительно ускоренным стадией гипермутативности в недавно инактивированных массивах (Shepelev et al., 2009; Uralsky et al., 2019; Miga, Alexandrov, 2021); градиент плотности инсерций транспозонов, который особенно выражен для повторяющихся элементов LINE1 (L1), поскольку они являются наиболее распространенными вставками в АС; градиент возраста элементов L1, встроенных в АС (их возраст может быть оценен независимо от возраста АС путем подсчета замен относительно предкового консенсуса); градиент других менее распространенных признаков распада, таких как инверсии и экспансии других сателлитов, которые распространены в дивергентных компартментах (мономерных и дивергентных ПВП) и редки в активных ПВП-массивах. Эти характерные признаки были замечены ранее (см. обзор Miga, Alexandrov, 2021), но лишь в нескольких хромосомах, в которых имелись более полные центромерные сборки в GRCh38/hg38. Они также были подтверждены в первых двух Т2Т сборках

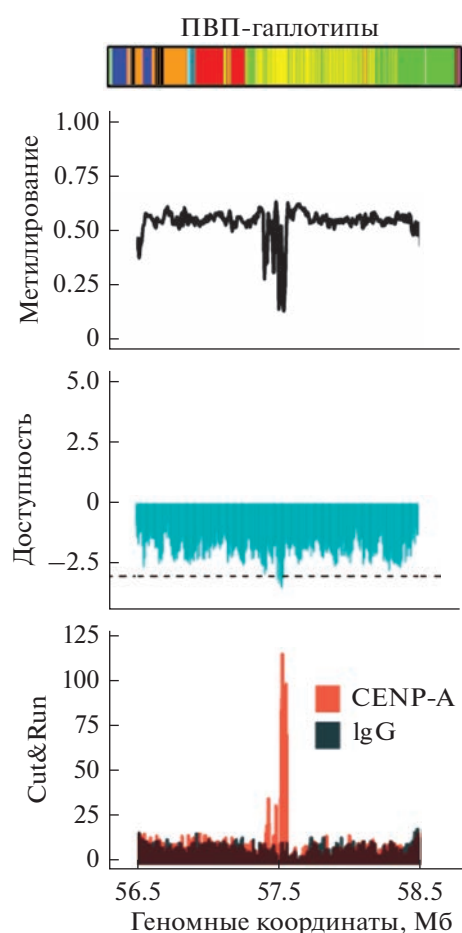


Рис. 1. Активный ПВП-массив X-хромосомы образца HG002. На верхней панели показан трек с аннотацией ПВП-гаплотипов, которые обозначены разными цветами. На второй панели показан уровень метилирования в центромерном регионе, гипометилированная область (“центромерный провал”) указывает на местоположение кинетохора. На третьей — доступность хроматина для экзогенных ферментов, измеренная методом *paHoNOMe* (Lee et al., 2020), на нижней панели — результаты CENP-A CUT&RUN эксперимента для X-хромосомы образца HG002. Красный цвет — антитела к CENP-A, черный цвет — контрольные неспецифические антитела. Представленные данные взяты из статьи Gershman et al., 2022. Видно, что гипометилированная область и область с минимальной доступностью хроматина соответствуют пику CENP-A, т.е. обозначают расположение кинетохора.

(хромосом X и 8) (Miga et al., 2020; Logsdon et al., 2021). Наконец теперь, когда центромеры всех хромосом достоверно собраны в Т2Т-СНМ13, мы смогли количественно и систематически исследовать все эти симметрии и градиенты и документировать процесс послойной экспансии, создающий характерный и сложный паттерн расширяющейся центромеры.

3) Оказалось также, что симметрии и градиенты послойного расширения можно отчасти наблюдать и внутри активных ПВП массивов центромер. Градиенты распада практически не видны, поскольку

инверсии, инсерции мобильных элементов и экспансии не-АС, хоть и встречаются, но настолько редки, что носят анекдотический характер. В то же время, градиенты размера и дивергенции и симметричное расположение внутренних слоев можно зарегистрировать. Путем сравнения отдельных ПВП друг с другом и фиксации замен (по сравнению с консенсусом), повторяющихся в одном регионе активного ПВП-массива, можно выделить комбинации мутаций, называемые ПВП-гаплотипами. Обычно гаплотип сформирован 3–10 мутациями на ПВП (считая мутации, которые встречаются только в этом гаплотипе). Построение филогенетического дерева из консенсусных ПВП всех гаплотипов позволяет поделить их на более старые (предковые) и более молодые (деривативные). ПВП одного гаплотипа образуют дискретные слои, причем слои предковых гаплотипов часто фланкируют более молодые с обеих сторон, располагаясь по краям центромеры. Кроме того, дивергенция внутри массивов в центральных ПВП-гаплотипах часто немного ниже, чем во фланкирующих массивах, что указывает на то, что экспансия центральных ПВП-гаплотипов произошла недавно. Это показывает, что модель послойного расширения распространяется и на активный массив. Кроме того, наличие идентифицируемых гаплотипических слоев дает внутреннюю топографию активного массива и позволяет изучить различия этой топографии между индивидами. Теперь можно понять: расположен ли кинетохор всегда на одном и том же слое в разных индивидуальных копиях одной и той же хромосомы, или его положение может варьировать.

Гипотеза кинетохор-селекции. Описанная выше картина расширяющейся центромеры и послойной экспансии может объясняться различными способами. Одно из возможных объяснений состоит в том, что экспансия АС происходит независимо от кинетохора, но он имеет сродство к активно растущим массивам, вызванное какими-то их свойствами, например, гомогенностью (“гипотеза независимого расширения”). Такая экспансия наблюдается в частности в виде мегабазных дупликаций в нецентромерных сателлитах, таких как массивы HSat3, которые не связаны с кинетохорами (Altemose et al., 2022a). Другая возможность состоит в том, что кинетохорные белки или другие белки, которые могут ассоциироваться с центромерами, такие как факторы загрузки гистонов, репликации, рекомбинации или репарации, играют причинную роль в экспансии конкретных вариантов ПВП (“гипотеза кинетохор-селекции” (Miga, Alexandrov, 2021)). В последнем случае возможный сценарий может быть сформулирован следующим образом. Кинетохор выбирает тот район в активном массиве (в простейшем случае тот гаплотип), к которому имеет наибольшее сродство. Оно может определяться одной или несколькими гаплотипическими мутациями, прямо или через конформацию ДНК или хроматина. При этом в месте занятом кинетохором (в простейшем случае, покрытом

СЕНР-А) дупликационный процесс (или другие формы амплификации ПВП) происходит более активно, чем снаружи, и значительно опережает делеционный процесс. В остальной части активного массива соотношение этих процессов может быть иное. В результате активный массив как бы растет своей подкинетохорной частью. Для того, чтобы оценить справедливость этой гипотезы, нужны дальнейшие исследования. Однако обнаруженные в одном из исследований (Altemose et al., 2022a) признаки недавней дупликационной активности под кинетохорами в СНМ13 как будто делают ее более правдоподобной.

НОВОЕ В ЦИТОГЕНЕТИКЕ ЦЕНТРОМЕР

Помимо перечисленных выше общих достижений и соображений, анализ первых полных центромерных сборок принес ряд новостей, модифицирующих наши представления о том, какой может и какой не может быть центромера у человека. Считалось, что активный ПВП-массив должен быть “протяженным, непрерывным и гомогенным” (Miga, Alexandrov, 2021). В частности было известно, что он может содержать АС повторы, идущие по прямой или по обратной цепи, но не может содержать переходов с прямой цепи на обратную (инверсий) (Jain et al., 2018b; Miga et al., 2020; Logsdon et al., 2021). Также было известно, что в отдельных, очень редких, случаях АС активного ПВП-массива может быть разорван инсерциями коротких или длинных мобильных элементов (LINEs или SINEs). Однако более протяженные разрывы никогда не обнаруживались. Ниже мы рассмотрим подробно только новости “цитогенетического масштаба”, т.е. модификации описанных выше представлений, которые вовлекают миллионы или сотни тысяч нуклеотидов и потенциально могут быть верифицированы с использованием микроскопических (в отличие от чисто молекулярных) методов исследования.

Расщепленные центромеры. Случаи экспансии не-АС повторов внутри АС массивов довольно часты в мертвых мономерных слоях и, как считалось раньше, не встречаются в активных центромерных ПВП-массивах. Однако Altemose с соавторами (Altemose et al., 2022a) впервые обнаружили три “расщепленные” центромеры, которые прерываются очень большими массивами не-АС повторов. При этом в центромерах хромосом 3 и 4 расщепленным оказался активный массив (рис. 2).

В центромере 3-й хромосомы активный АС массив S01/1C3H1L разделен на 3 части (0.8, 0.03 и 0.5 Мб) двумя большими доменами HSat1A (2.5 и 0.3 Мб), которые значительно превышает длину массива АС. То же самое можно наблюдать в центромере 4-й хромосомы, где домены HSat1A (1.7 и 0.1 Мб) разделяют активный АС массив S2C4H1L на 3 части (0.7, 2.7 и 0.2 Мб). Кроме того, в центромере хромосомы 5 на q-плече большой домен HSat3 (0.4 Мб) отделяет активный массив (2.5 Мб) от мо-

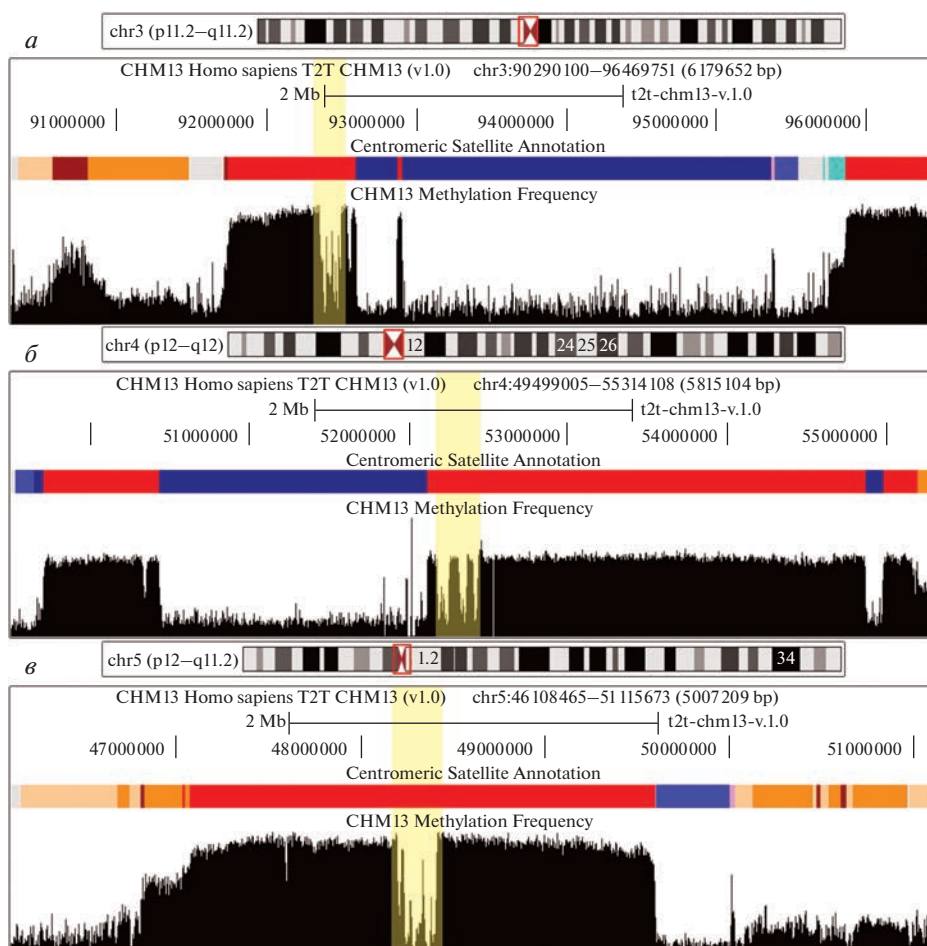


Рис. 2. Аннотация центромерных районов хромосом 3 (а), 4 (б) и 5 (в). Красным цветом на треке “Centromeric Satellite Annotation” показан активный ПВП, синим — классический сателлит, оранжевым — неактивные ПВП, коричневым — дивергентные ПВП, бежевым — мономерные массивы. Вертикальной линией желтого цвета показана область “провала” в метилировании — место расположения кинетохора. Во всех трех хромосомах видны протяженные блоки классического сателлита (синий цвет) внутри АС массивов (красный, оранжевый, коричневый, бежевый).

номерного и неактивного АС массивов (общей длиной 1 Мб). Само по себе это не является необычным, тем не менее, размер вставки классического сателлита ставит его в один ряд с вышеупомянутыми случаями. Авторы проверили ориентацию нитей расщепленных АС массивов и фланкирующих последовательностей в хромосомах 3 и 4, и обнаружили, что они совпадают, то есть вставки не могли возникнуть в результате инверсий на стыке между соседними массивами HSat1 и АС. Таким образом, вероятным объяснением является недавняя вставка и экспансия HSat1 в пределах этих двух активных массивов. Неожиданностью стало то, что вставки, которые, вероятно, сначала должны были быть маленькими, могли разрастись до таких гигантских размеров. Не ясно, за какой промежуток времени могла произойти такая амплификация. Авторы проверили “черновые версии” фазированных (с разделенными гомологами) сборок геномов 16 человек, которые получил Human Pangenome Reference Consortium

(HPRC), и во всех обнаружили расщепленные центромеры 3 и 4 хромосом. Это, однако, не исключает того, что в каких-то древних человеческих популяциях или в ископаемых останках могут обнаружиться нерасщепленные предковые активные массивы (какими они были до вставки HSat1 и экспансии). В этом случае можно будет попытаться понять, в какой период времени это произошло и определить скорость амплификации сателлитной вставки. В среднем в геноме человека частота контактов АС регионов с не-АС составляет: 0.17/Мб для ПВП, 6.48/Мб для дивергентных ПВП и 15.23/Мб для мономерных областей. Таким образом, экспансии не-АС, очевидно, являются частью общего градиента дивергенции и распада, наблюдаемого в ряду ПВП-дПВП-мономеры.

Еще одной особенностью центромер с расщепленными активными массивами является то, что дивергенция ПВП в разных разделенных сегментах может существенно различаться. Для сравнения, в неко-

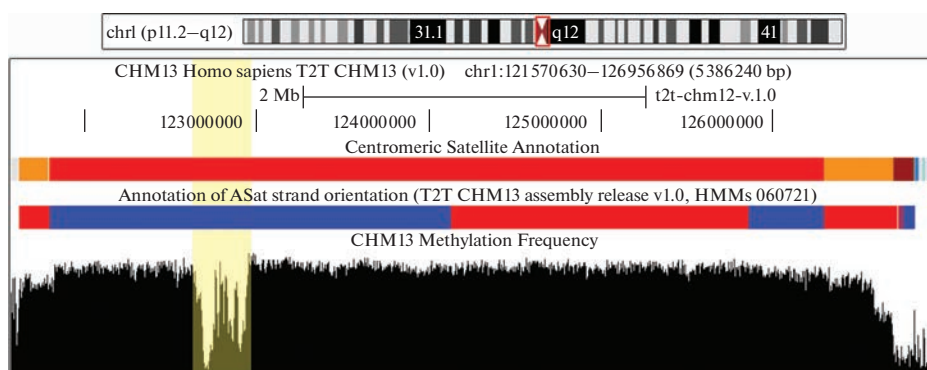


Рис. 3. Инверсия в активном ПВП массиве на 1-ой хромосоме. Красным цветом на треке Centromeric Satellite Annotation показан активный ПВП. На треке “Annotation of ASat strand orientation” синим цветом показана последовательность по прямой, красным — по обратной цепи. Вертикальной полосой желтого цвета показана область “провала” в метилировании — место расположения кинетохора.

торых активных ПВП-массивах (или их частях) мономеры имеют особенно низкую среднюю дивергенцию (0.52–0.65%, например, центромеры 6 и 13 хромосом). В то же время в расщепленной центромере 4-ой хромосомы в каждом из двух больших АС сегментов, способных разместить кинетохор, тот, в котором он действительно размещается, имеет несколько меньшую дивергенцию (0.91 против 1.36%). В центромере 3 эти величины тоже не одинаковы (0.65 против 0.86%). Это неравенство дивергенции в разделенных частях активного массива обращает на себя внимание, поскольку нечто подобное предсказывает гипотеза кинетохор-селекции (Miga, Alexandrov, 2021; Altemose et al., 2022a), которая связывает присутствие кинетохора с повышенной вероятностью дупликации (амплификации) ПВП активного массива, что должно повышать его гомогенность. В то же время предполагается, что участки активного массива перманентно отделенные от подкинетохорной области, могут превратиться в неактивные “мертвые центромеры”, которые, как считается, проходят период гипермутативности (Shepelev et al., 2009; Uralsky et al., 2019). Таким образом, различия дивергенции в разделенных сегментах активного массива входят в предсказания теории. Однако для того, чтобы оценить значимость этого совпадения, требуются дополнительные исследования.

Мегабазная полиморфная инверсия в активном домене центромеры 1. Хотя было известно, что дивергентные АС содержат множество инверсий (Rudd, Willard, 2004) и инсерций транспозонов (Kazakov, 2003), такие события в активных массивах ПВП являются неожиданными, поскольку ранее они считались однородными (Warburton, Willard, 1990, 1995).

Количественная оценка случаев инверсий нитей внутри (не на границах) массивов сателлитов выявила неожиданные аномалии. Например, была обнаружена 1.7 Мб инверсия внутри активного массива АС ПВП на 1-ой хромосоме (рис. 3), а также инверсии в

неактивных (но тоже гомогенных) массивах ПВП на хромосомах 3, 16 и 20 (Altemose et al., 2022a).

Общий подсчет точек переключения цепей в АС оценивал количество инверсий отдельно для ПВП, дПВП и мономерных областей, направление нитей было взято из трека “Annotation of ASat strand orientation”. Переключение, произошедшее прямо на границе сегмента (например, между разными массивами ПВП), не учитывалось, так как разные массивы ПВП часто идут по разным нитям. В однородных ПВП (общая длина в геноме 70 Мб) были обнаружены четыре инверсии: в ПВП S1C1/5/19H1L_1, S1C3H2, S2C16pH2-B/A и S2C20H2. Дивергентные ПВП (1.8 Мб) имели 11 переключений (7 инверсий), одна из которых включала только один мономер. Дивергентные мономерные массивы (10.7 Мб) имели 242 переключения. Таким образом, одно переключение происходило каждые 17.5, 0.16 и 0.44 Мб в однородных ПВП, дивергентных ПВП и мономерных слоях соответственно. Это иллюстрирует упомянутый выше возрастной градиент инверсий, который проявляется в том, что более молодые гомогенные ПВП содержат меньше инверсий, а более старые дивергентные ПВП и мономерные слои гораздо больше. Это один из градиентов распада, характерных для расширяющейся центромеры.

Чтобы лучше понять, являются ли сателлитные инверсии, инсерции и делеции обычными за пределами генома CHM13, авторы провели их поиск в 16 фазированных (с разделенными гомологами) черновых сборках индивидуумов из разнообразных популяций, полученных консорциумом HPRC (Miga, Wang, 2021). Он показал, что инверсия в активном АС ПВП-массиве на хромосоме 1 является полиморфной у разных особей и появляется примерно в половине индивидуальных хромосом 1 (11 из 24). Однако вставки HSat1A на хромосомах 3 и 4 присутствуют во всех проанализированных хромосомах (32 из 32 и 33 из 33 соответственно). Более того, отсутствующий в CHM13 массив HSat3B2 на хромосоме 1 целиком со-

держался в 400 Кб полиморфной делеции, которая была обнаружена в 29% (8 из 28) исследованных хромосом. В результате сделан вывод, что крупные пери/центромерные структурные перестройки, такие, как описанные выше расщепленные центромеры и мега-инверсия не являются специфичными для генома CHM13, а их присутствие либо варьирует, либо фиксировано у всех людей.

Поскольку мега-инверсия в центромере 1 является уникальной, для этого обзора мы изучили ее несколько более подробно. Из-за того, что рекомбинационный обмен между инверсией и основной частью массива, включающей подкинетохорную область, может быть затруднен, можно было бы ожидать отличий в дивергенции и последовательности ПВП между инверсией и остальным массивом. Однако таких отличий выявлено не было. Возможно, инверсия просто слишком недавняя, и с момента ее появления просто не прошло достаточное количество времени.

В связи с обнаружением этой мега-инверсии уместно задаться вопросом, не будет ли соединение двух разных гомологов в мейозе несколько нарушено, если в одном инверсия будет, а в другом нет, и не может ли такое нарушение привести к повышенной мис-сегрегации хромосомы 1?

Центромерные макро-повторы в ряде хромосом. Макро-повторы — это большие дубликации АС, которые могут довольно сильно различаться по возрасту. Недавние экспансии повторов можно определить путем построения и анализа точечных карт (dot plots) центромерных последовательностей при самовыравнивании.

Длина единицы выбирается примерно равной длине ПВП (обычно 1000–3000 н.п.). Точка ставится на графике, если два ПВП идентичны в обеих последовательностях. Главная диагональ представляет 100% идентичность, дополнительные линии, параллельные диагонали, представляют почти идентичные внутренние повторы в последовательности (недавние дубликации или экспансии). Таким образом, различные рисунки из коротких отрезков вокруг главной диагонали, которые наблюдаются на графиках, интерпретировались как признаки недавней амплификационной активности. В сборке Т2Т-CHM13 такие узоры обычно наблюдались в районах, связанных с кинетохором (в 18–19 из 23-х хромосом), но отнюдь не исключительно в этих районах (см. рис. 4а, левая панель). Если нужно было исследовать отношения последовательностей с более низкой идентичностью в АС массиве, длина единицы постепенно уменьшалась (скажем, до 500 п.н.; рис. 4б правая панель), а, например, древний макро-повтор в центромере Х-хромосомы уже не виден на точечных картах, но может быть обнаружен как повторяющаяся последовательность гаплотипов в левой половине активного массива АС (рис. 4в). Кроме того, макро-повторы иногда можно обнаружить по повторяю-

щимся паттернам в распределении структурных вариантов ПВП на StV-картах (например, древний макро-повтор в хромосоме 3).

Интересно, что в работе Altomose с соавторами (Altomose et al., 2022a) также показано, что в активном АС массиве на хромосоме 12 CENP-A обогащен только один из двух больших макро-повторов, хотя оба содержат схожие молодые ПВП-гаплотипы. Причем обогащение CENP-A наблюдали в том макро-повторе, где имела место недавняя амплификация одного из молодых гаплотипов (рис. 4а). Возможно, что макро-повторы исходно были одинаковыми (или по крайней мере более похожими), но стабильное расположение кинетохора в одном из них привело к дополнительной амплификации ПВП молодого гаплотипа, находящегося под кинетохором.

Гаплотипические центромерные эпиааллели в Х-хромосоме. Как уже говорилось выше, работы Т2Т консорциума показали, что место локализации кинетохора отмечено в ДНК активного ПВП массива “провалом” в метилировании. Авторы также изучили вариацию расположения кинетохора на активном ПВП массиве Х-хромосом разных людей (рис. 5) (Gershman et al., 2022). Для этого использовали 3 метода: (1) сравнение паттернов обогащения различных ПВП-гаплотипов в коротких прочтениях из экспериментов CENP-A NChIP-seq и CUT&RUN; (2) прямое прикладывание прочтений из этих экспериментов на полностью собранные центромеры Х-хромосомы CHM13; (3) прямое определение положения центромерных “провалов” в метилировании ДНК, которое маркирует положение кинетохора, из ONT прочтений (Simpson et al., 2017). Все три метода показали, что в разных гомологичных Х-хромосомах кинетохор может находиться в разных положениях.

Было картировано по крайней мере три таких положения: на правом фланге активного массива, на левом фланге и посередине. Расстояния между тремя точками составляют около 1 Мб. Таким образом, если взять крайний случай и предположить, что в одной клетке имеются 2 гомологичные Х-хромосомы, где кинетохоры занимают 2 фланговых положения, расстояние между ними будет составлять около 2 Мб. В этой связи приходится задаться вопросом: не будет ли в мейозе бивалент, содержащий такие хромосомы, получать вращательный момент при разделении гомологов из-за асимметричного присоединения микротрубочек веретена? Это может привести к тому, что данный бивалент примет перпендикулярное положение (по отношению к обычному). Представляется, что такая ситуация может легко вызвать нерасхождение хромосом и привести к числовым аномалиям Х-хромосом, которые, как известно, встречаются очень часто (примерно 1/500 новорожденных (Berglund et al, 2020)). Хотя эти аномалии зачастую не вызывают драматических фенотипических проявле-

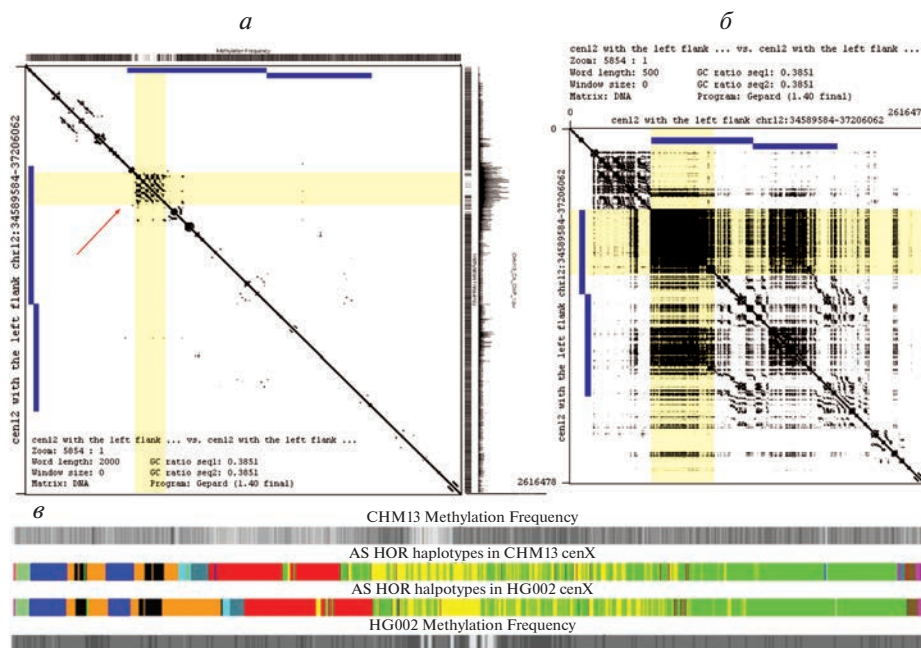


Рис. 4. Точечные карты для центромерных последовательностей 12-ой хромосомы. Слева (а) показана “жесткая” карта (точка ставится при совпадении 2000 т.н.п.). Кластер штрихов вне диагонали (указан стрелкой) показывает признаки недавней амплификационной активности, и видно, что это место совпадает с положением “центромерного провала” (показано желтой полосой) и расположением кинетохора. Справа (б) показана “мягкая” карта для этой же области (точка ставится при совпадении 500 т.н.п.), видны два макро-повтора (отмечены синими линиями), и положение кинетохора в более длинном макро-повторе. Можно предположить, что удлинение одной из копий произошло уже после дупликации из-за амплификационной активности под кинетохором. в – Древние макро-повторы в центромере X-хромосомы в CHM13 и HG002. Различные ПВП-гаплотипы S3CXH1L (DXZ1) в активном массиве обозначены разными цветами, как описано в Altomose et al., 2022. Видны 2 макро-повтора (синий–оранжевый–черный–оранжевый). Эти повторы настолько древние, что не удается подобрать условия, при которых они были бы видны на точечной карте, однако их можно видеть на карте ПВП-гаплотипов.

ний благодаря инактивации избыточных X-хромосом, имеются данные, что пациенты с такими аномалиями гораздо чаще страдают дефектами развития нервной системы и нейродегенеративными заболеваниями, а также меньше живут и быстрее стареют (Tartaglia et al., 2010; Bishop et al., 2011; Bojesen, Gravholt 2011; Skakkebaek et al., 2014). В особенностях мозаичная потеря X-хромосомы может являться скрытым фактором в ряде упомянутых патологий (Iourov et al., 2019).

ЗАКЛЮЧЕНИЕ. ПЛАСТИЧНОСТЬ ЦЕНТРОМЕР, ЧТО ОНИ МОГУТ И ЧЕГО НЕ МОГУТ ДЕЛАТЬ

Вышеизложенные новые данные о макроструктуре центромера человека несколько модифицируют традиционные представления. До сих пор считалось, что гомогенный кор центромеры, занятый активным массивом, это однородное образование, состоящее из тандемных ПВП, всегда идущих в одном направлении (по одной нити). Единственным известным нарушением были очень редкие вставки мобильных элементов. Кроме того, было известно о существовании ПВП-гаплотипов, определяемых комбинацией

нескольких мутаций, но как они распределены в центромере не было ясно, поскольку отсутствовали протяженные сборки. Наиболее логичным считалось предположение, что близко расположенные ПВП больше похожи друг на друга, чем на ПВП, расположенные далеко (Durfy, Willard, 1989).

Действительность, как это нередко случается, оказалась отчасти похожей и отчасти непохожей на прежние представления. Гомогенный кор центромеры, действительно, в среднем является гомогенным, однако никакие его параметры не контролируются строго, и повсюду можно обнаружить весьма значительные исключения. Размеры активных доменов, всегда большие, варьируют в очень широком диапазоне (Miga et al., 2014; Altomose et al., 2022a), характерная дивергенция ПВП в активном массиве в среднем обычно в пределах 1%, но на краях встречаются копии ПВП (или небольшие массивы), где дивергенция достигает 2–3% (Altomose et al., 2022a). В целом отсутствие инверсий в гомогенных массивах является правилом, однако имеются рассмотренные выше исключения. Нужно отметить, что в других клеточных линиях (индивидуумах) они могут быть другими, так что действительная частота и спектр этих событий могут быть оценены только после ана-

лиза первых десятков полных геномных сборок из примерно 350, запланированных к производству консорциумом HPRC. При этом впервые должна получить конкретное освещение проблема взаимосвязи полиморфизма прицентромерных районов хромосом, имеющая первостепенное значение в анализе причин различных заболеваний, в первую очередь нервных и нейродегенеративных, а также сердечно-сосудистых заболеваний (Fung et al., 2006; Beecham et al., 2014; Herold et al., 2016; Miga 2019).

Наличие двух расщепленных активных массивов также явилось сюрпризом. Насколько можно было судить по ограниченной выборке из 16 человек, они расщеплены у всех людей (Altemose et al., 2022a). Однако можно ожидать исключений в каких-нибудь древних вариантах хромосом 3 и 4, сохранившихся либо в современных геномах, либо в ископаемой ДНК. В этой связи следует упомянуть важную особенность центромер: мейотическая гомологичная рекомбинация в них подавлена (Mahtani, Willard, 1998). Поэтому древние варианты центромер, привнесенные в современные геномы в результате скрещивания с архаическими людьми, такими как неандертальцы и денисовцы, или случайно сохранившие предковую структуру в неизменном виде, имеют шанс дойти до нас в целости. Они не будут “измельчены” рекомбинацией на небольшие куски самого разного “структурного возраста” (определяемого по основным редко меняющимся структурным чертам). В отличие от последнего “мутационный возраст” всех последовательностей в современном геноме одинаковый, поскольку мутации происходят по всему геному более или менее равномерно. Таким образом, в каких-то человеческих линиях могут случайным образом сохраниться центромеры, не претерпевшие крупных структурных изменений в течение очень долгого времени. Такие центромеры можно назвать древними, имея в виду именно их структурный возраст. Таким образом, если древняя нерасщепленная центромера каким-то образом уцелеет при прохождении популяционных “бутылочных горлышек”, она может быть обнаружена у современного человека в целости и сохранности (за исключением отдельных случайных мутаций) и ее структурные особенности могут быть идентифицированы с таковыми в ископаемой ДНК.

Кроме того, расщепленные активные массивы и макро-повторы могут представлять интерес в плане обоснования гипотезы кинетохор-селекции и общих сценариев эволюции ПВП массивов. Например, если кинетохор находится на одном из двух почти идентичных макро-повторов, можно предполагать, что в нем будет больше признаков недавней дупликационной активности, чем в другом макро-повторе. Также интересно насколько стабильным будет такое положение? Не будет ли кинетохор располагаться на разных копиях макро-повтора у разных

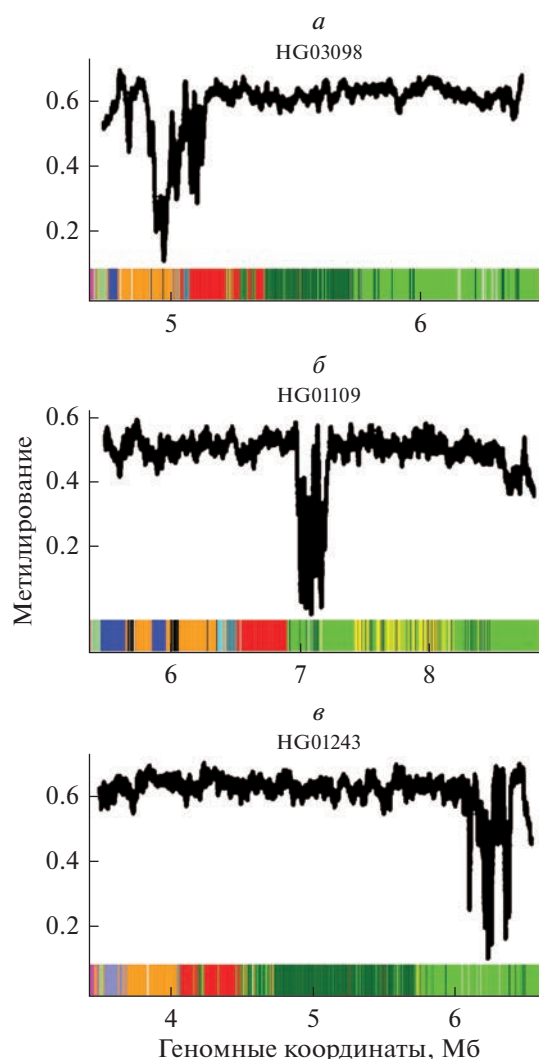


Рис. 5. Три варианта расположения кинетохора в активном ПВП-массиве X-хромосомы (гаплотипические центромерные эпипаллеты). Черной линией показан уровень метилирования в активном ПВП-массиве. Гипометилированная область указывает на место расположения кинетохора в активном ПВП массиве X-хромосомы трех индивидов (HG03098 – *а*, HG01109 – *б*, HG01243 – *в*). Для каждого образца показаны треки с аннотацией ПВП-гаплотипов, где разные гаплотипы обозначены цветами. Можно видеть, что наряду с вариантами центромер, имеющими два макро-повтора (*б*), подобными показанным на рис. 4, имеются варианты (*а* и *в*), не имеющие макро-повторов.

людей? Возможны ли случаи, когда одна часть расщепленной центромеры остается активным массивом, а другая становится мертвой центромерой, претерпевает гипермутативность и превращается в дивергентный ПВП-массив? Эти и другие интересные вопросы потребуют дальнейших исследований, однако технологическая база для них уже создана, и можно думать, что науку о центромерах ожидают захватывающие времена.

ФИНАНСИРОВАНИЕ РАБОТЫ

Работа выполнена при финансовой поддержке Российского научного фонда в рамках научного проекта № 19-75-30039.

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

В работе не было экспериментов с участием животных или людей.

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

- Alexandrov I., Kazakov A., Tumeneva I., Shepelev V., Yurov Y. 2001. Alpha-satellite DNA of primates: old and new families. *Chromosoma*. V. 110. P. 253.
<https://doi.org/10.1007/s004120100146>
- Altomose N., Logsdon G.A., Bzikadze A.V., Sidhwani P., Langley S.A., Caldas G.V., Hoyt S.J., Uralsky L., Ryabov F.D., Shew C.J., Sauria M.E.G., Borchers M., Gershman A., Mikheenko A., Shepelev V.A. et al. 2022a. Complete genomic and epigenetic maps of human centromeres. *Science*. V. 376. NO. 6588.
<https://doi.org/10.1126/science.abl4178>
- Altomose N., Maslan A., Smith O.K., Sundararajan K., Brown R.R., Mishra R., DeWeiler A.M., Neff N., Miga K.H., Straight A.F., Streets A. 2022b. DiMeLo-seq: a long-read, single-molecule method for mapping protein–DNA interactions genome wide. *Nat Methods*. V. 19. P. 711.
<https://doi.org/10.1038/s41592-022-01475-6>
- Beecham G.W., Hamilton K., Naj A.C., Martin E.R., Huentelman M., Myers A.J., Corneveaux J.J., Hardy J., Vonsattel J.-P., Yonkin S.G., Bennett D.A., De Jager P.L., Larson E.B., Crane P.K., Kamboh M.I. et al. 2014. Genome-wide association Meta-analysis of neuropathologic features of Alzheimer's disease and related dementias. *PLoS Genet*. V. 10, e1004606.
<https://doi.org/10.1371/journal.pgen.1004606>
- Berglund A., Stochholm K., Gravholt C.H. 2020. The epidemiology of sex chromosome abnormalities. *Am. J. Med. Genet*. V. 184. P. 202.
<https://doi.org/10.1002/ajmg.c.31805>
- Bishop D.V.M., Jacobs P.A., Lachlan K., Wellesley D., Barnicoat A., Boyd P.A., Fryer A., Middlemiss P., Smithson S., Metcalfe K., Shears D., Leggett V., Nation K., Scerif G. 2011. Autism, language and communication in children with sex chromosome trisomies. *Archives of Disease in Childhood*. V. 96. P. 954.
<https://doi.org/10.1136/adc.2009.179747>
- Bojesen A., Gravholt C.H. 2011. Morbidity and mortality in Klinefelter syndrome (47,XXY): Morbidity and mortality in KS. *Acta Paediatrica*. V. 100. P. 807.
<https://doi.org/10.1111/j.1651-2227.2011.02274.x>
- Bzikadze A.V., Pevzner P.A. 2020. Automated assembly of centromeres from ultra-long error-prone reads. *Nat. Biotechnol*. V. 38. P. 1309.
<https://doi.org/10.1038/s41587-020-0582-4>
- Carlson M., Brutlag D. 1977. Cloning and characterization of a complex satellite DNA from drosophila melanogaster. *Cell*. V. 11. P. 371.
[https://doi.org/10.1016/0092-8674\(77\)90054-X](https://doi.org/10.1016/0092-8674(77)90054-X)
- Cimini D., Howell B., Maddox P., Khodjakov A., Degross F., Salmon E.D. 2001. Merotelic Kinetochore Orientation Is a Major Mechanism of Aneuploidy in Mitotic Mammalian Tissue Cells. *Journal of Cell Biology*. V. 153. P. 517.
<https://doi.org/10.1083/jcb.153.3.517>
- Dumont M., Gamba R., Gestraud P., Klaasen S., Worrall J.T., De Vries S.G., Boudreau V., Salinas-Luypaert C., Maddox P.S., Lens S.M., Kops G.J., McClelland S.E., Miga K.H., Fachinetti D. 2020. Human chromosome-specific aneuploidy is influenced by DNA-dependent centromeric features. *EMBO J*. V. 39.
<https://doi.org/10.15252/embj.2019102924>
- Durfy S.J., Willard H.F. 1989. Patterns of intra- and interarray sequence variation in alpha satellite from the human X chromosome: Evidence for short-range homogenization of tandemly repeated DNA sequences. *Genomics*. V. 5. P. 810.
[https://doi.org/10.1016/0888-7543\(89\)90123-7](https://doi.org/10.1016/0888-7543(89)90123-7)
- Earnshaw W., Bordwell B., Marino C., Rothfield N. 1986. Three human chromosomal autoantigens are recognized by sera from patients with anti-centromere antibodies. *J. Clin. Invest*. V. 77. P. 426.
<https://doi.org/10.1172/JCI112320>
- Earnshaw W.C., Cooke C.A. 1989. Proteins of the inner and outer centromere of mitotic chromosomes. *Genome*. V. 31. P. 541.
<https://doi.org/10.1139/g89-103>
- Earnshaw W.C., Rothfield N. 1985. Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. *Chromosoma*. V. 91. P. 313.
<https://doi.org/10.1007/BF00328227>
- Fung H.-C., Scholz S., Matarin M., Simón-Sánchez J., Hernandez D., Britton A., Gibbs J.R., Langefeld C., Stiebert M.L., Schymick J., Okun M.S., Mandel R.J., Fernandez H.H., Foote K.D., Rodríguez R.L. et al. 2006. Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data. *The Lancet Neurology*. V. 5. P. 911.
[https://doi.org/10.1016/S1474-4422\(06\)70578-6](https://doi.org/10.1016/S1474-4422(06)70578-6)
- Ge Y., Wagner M.J., Siciliano M., Wells D.E. 1992. Sequence, higher order repeat structure, and long-range organization of alpha satellite DNA specific to human chromosome 8. *Genomics*. V. 13. P. 585.
[https://doi.org/10.1016/0888-7543\(92\)90128-F](https://doi.org/10.1016/0888-7543(92)90128-F)
- Gershman A., Sauria M.E.G., Guitart X., Vollger M.R., Hook P.W., Hoyt S.J., Jain M., Shumate A., Razaghi R., Koren S., Altomose N., Caldas G.V., Logsdon G.A., Rhie A., Eichler E.E. et al. 2022. Epigenetic patterns in a complete human genome. *Science*. V. 376. NO. 6588.
<https://doi.org/10.1126/science.abj5089>
- Ghareghani M., Porubský D., Sanders A.D., Meiers S., Eichler E.E., Korbel J.O., Marschall T. 2018. Strand-seq enables reliable separation of long reads by chromosome via expectation maximization. *Bioinformatics*. V. 34. P. i115.
<https://doi.org/10.1093/bioinformatics/bty290>
- Herold C., Hooli B.V., Mullin K., Liu T., Roehr J.T., Mattheisen M., Parrado A.R., Bertram L., Lange C., Tanzi R.E. 2016. Family-based association analyses of imputed genotypes reveal genome-wide significant association of Alzheimer's disease with OSBPL6, PTPRG, and PDCL3. *Mol. Psychiatry*.

- try. V. 21. P. 1608.
<https://doi.org/10.1038/mp.2015.218>
- International Human Genome Sequencing Consortium*. 2004. Finishing the euchromatic sequence of the human genome. *Nature*. V. 431. P. 931.
<https://doi.org/10.1038/nature03001>
- International Human Genome Sequencing Consortium*, Whitehead Institute for Biomedical Research, Center for Genome Research., *Lander E.S., Linton L.M., Birren B., Nusbaum C., Zody M.C., Baldwin J., Devon K., Dewar K., Doyle M., FitzHugh W., Funke R., Gage D. et al.* 2001. Initial sequencing and analysis of the human genome. *Nature*. V. 409. P. 860.
<https://doi.org/10.1038/35057062>
- Iourov I.Y., Vorsanova S.G., Yurov Y.B., Kutsev S.I.* 2019. Ontogenetic and pathogenetic views on somatic chromosomal mosaicism. *Genes*. V. 10. P. 379.
<https://doi.org/10.3390/genes10050379>
- Jain M., Koren S., Miga K.H., Quick J., Rand A.C., Sasani T.A., Tyson J.R., Beggs A.D., Dillthey A.T., Fiddes I.T., Malla S., Marriott H., Nieto T., O'Grady J., Olsen H.E. et al.* 2018a. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* V. 36. P. 338.
<https://doi.org/10.1038/nbt.4060>
- Jain M., Olsen H.E., Turner D.J., Stoddart D., Bulazel K.V., Paten B., Haussler D., Willard H.F., Akeson M., Miga K.H.* 2018b. Linear assembly of a human centromere on the Y chromosome. *Nat. Biotechnol.* V. 36. P. 321.
<https://doi.org/10.1038/nbt.4109>
- Jarvis E.D., Formenti G., Rhie A., Guarracino A., Yang C., Wood J., Tracey A., Thibaud-Nissen F., Vollger M.R., Porubsky D., Cheng H., Asri M., Logsdon G.A., Carnevali P., Chaisson M.J.P. et al.* 2022. Semi-automated assembly of high-quality diploid human reference genomes. *Nature*. V. 611. P. 519.
<https://doi.org/10.1038/s41586-022-05325-5>
- Kazakov A.E., Shepelev V.A., Tumenewa I.G., Alexandrov A.A., Yurov Y.B., Alexandrov I.A.* 2003. Interspersed repeats are found predominantly in the "old" α satellite families. *Genomics*. V. 82. P. 619.
[https://doi.org/10.1016/S0888-7543\(03\)00182-4](https://doi.org/10.1016/S0888-7543(03)00182-4)
- Langley S.A., Miga K.H., Karpen G.H., Langley C.H.* 2019. Haplotypes spanning centromeric regions reveal persistence of large blocks of archaic DNA. *eLife*. V. 8. e42989.
<https://doi.org/10.7554/eLife.42989>
- Lee I., Razaghi R., Gilpatrick T., Molnar M., Gershman A., Sadowski N., Sedlazeck F.J., Hansen K.D., Simpson J.T., Timp W.* 2020. Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing. *Nat. Methods*. V. 17. P. 1191.
<https://doi.org/10.1038/s41592-020-01000-7>
- Levine M.S., Holland A.J.* 2018. The impact of mitotic errors on cell proliferation and tumorigenesis. *Genes Dev*. V. 32. P. 620.
<https://doi.org/10.1101/gad.314351.118>
- Levy S., Sutton G., Ng P.C., Feuk L., Halpern A.L., Walenz B.P., Axelrod N., Huang J., Kirkness E.F., Denisov G., Lin Y., MacDonald J.R., Pang A.W.C., Shago M., Stockwell T.B. et al.* 2007. The Diploid Genome Sequence of an Individual Human. *PLoS Biol.* V. 5. P. e254.
<https://doi.org/10.1371/journal.pbio.0050254>
- Lieberman-Aiden E., van Berkum N.L., Williams L., Imakaev M., Ragoczy T., Telling A., Amit I., Lajoie B.R., Sabo P.J., Dorschner M.O., Sandstrom R., Bernstein B., Bender M.A., Groudine M., Gnirke A. et al.* 2009. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science*. V. 326. P. 289.
<https://doi.org/10.1126/science.1181369>
- Logsdon G.A., Vollger M.R., Hsieh P., Mao Y., Liskovych M.A., Koren S., Nurk S., Mercuri L., Dishuck P.C., Rhie A., de Lima L.G., Dvorkina T., Porubsky D., Harvey W.T., Mikheenko A. et al.* 2021. The structure, function and evolution of a complete human chromosome 8. *Nature*. V. 593. P. 101.
<https://doi.org/10.1038/s41586-021-03420-7>
- Mahtani M.M., Willard H.F.* 1998. Physical and Genetic Mapping of the Human X Chromosome Centromere: Repression of Recombination. *Genome Res*. V. 8. P. 100.
<https://doi.org/10.1101/gr.8.2.100>
- Manuelidis L.* 1976. Repeating restriction fragments of human DNA. *Nucleic Acids Research*. V. 3. P. 3063.
<https://doi.org/10.1093/nar/3.11.3063>
- Manuelidis L., Wu J.C.* 1978. Homology between human and simian repeated DNA. *Nature*. V. 276. P. 92.
<https://doi.org/10.1038/276092a0>
- Miga K.H.* 2019. Centromeric satellite DNAs: Hidden sequence variation in the human population. *Genes*. V. 10. P. 352.
<https://doi.org/10.3390/genes10050352>
- Miga K.H., Alexandrov I.A.* 2021. Variation and Evolution of Human Centromeres: A Field Guide and Perspective. *Annu. Rev. Genet.* V. 55. P. 583.
<https://doi.org/10.1146/annurev-genet-071719-020519>
- Miga K.H., Eisenhart C., Kent W.J.* 2015. Utilizing mapping targets of sequences underrepresented in the reference assembly to reduce false positive alignments. *Nucleic Acids Res.* V. 43. P. e133.
<https://doi.org/10.1093/nar/gkv671>
- Miga K.H., Koren S., Rhie A., Vollger M.R., Gershman A., Bzikadze A., Brooks S., Howe E., Porubsky D., Logsdon G.A., Schneider V.A., Potapova T., Wood J., Chow W., Armstrong J. et al.* 2020. Telomere-to-telomere assembly of a complete human X chromosome. *Nature*. V. 585. P. 79.
<https://doi.org/10.1038/s41586-020-2547-7>
- Miga K.H., Newton Y., Jain M., Altemose N., Willard H.F., Kent W.J.* 2014. Centromere reference models for human chromosomes X and Y satellite arrays. *Genome Res*. V. 24. P. 697.
<https://doi.org/10.1101/gr.159624.113>
- Miga K.H., Wang T.* 2021. The Need for a Human Pangenome Reference Sequence. *Annu. Rev. Genom. Hum. Genet.* V. 22. P. 81.
<https://doi.org/10.1146/annurev-genom-120120-081921>
- Mikheenko A., Bzikadze A.V., Gurevich A., Miga K.H., Pevzner P.A.* 2020. TandemTools: mapping long reads and assessing/improving assembly quality in extra-long tandem repeats. *Bioinformatics*. V. 36. P. i75.
<https://doi.org/10.1093/bioinformatics/btaa440>
- Nechemia-Arbely Y., Fachinetti D., Miga K.H., Sekulic N., Soni G.V., Kim D.H., Wong A.K., Lee A.Y., Nguyen K., Dekker C., Ren B., Black B.E., Cleveland D.W.* 2017. Human centromeric CENP-A chromatin is a homotypic, octameric nucleosome at all cell cycle points. *Journal of Cell Biology*. V. 216. P. 607.
<https://doi.org/10.1083/jcb.201608083>
- Nechemia-Arbely Y., Miga K.H., Shoshani O., Aslanian A., McMahon M.A., Lee A.Y., Fachinetti D., Yates J.R., Ren B.,*

- Cleveland D.W. 2019. DNA replication acts as an error correction mechanism to maintain centromere identity by restricting CENP-A to centromeres. *Nat. Cell Biol.* V. 21. P. 743.
https://doi.org/10.1038/s41556-019-0331-4
- Nurk S., Koren S., Rhie A., Rautiainen M., Bizikadze A.V., Mikheenko A., Vollger M.R., Altemose N., Uralsky L., Gershman A., Aganezov S., Hoyt S.J., Diekhans M., Logsdon G.A., Alonge M. et al. 2022. The complete sequence of a human genome. *Science*. V. 376. P. 44.
https://doi.org/10.1126/science.abj6987
- Nurk S., Walenz B.P., Rhie A., Vollger M.R., Logsdon G.A., Grothe R., Miga K.H., Eichler E.E., Phillippy A.M., Koren S. 2020. HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. *Genome Res.* V. 30. P. 1291.
https://doi.org/10.1101/gr.263566.120
- Osoegawa K., Vessere G.M., Li Shu C., Hoskins R.A., Abad J.P., de Pablos B., Villasante A., de Jong P.J. 2007. BAC clones generated from sheared DNA. *Genomics*. V. 89. P. 291.
https://doi.org/10.1016/j.ygeno.2006.10.002
- Palmer J.D., Nugent J.M., Herbon L.A. 1987. Unusual structure of geranium chloroplast DNA: A triple-sized inverted repeat, extensive gene duplications, multiple inversions, and two repeat families. *Proc. Natl. Acad. Sci. U.S.A.* V. 84. P. 769.
https://doi.org/10.1073/pnas.84.3.769
- Rautiainen M., Nurk S., Walenz B.P., Logsdon G.A., Porubsky D., Rhie A., Eichler E.E., Phillippy A.M., Koren S. 2022. Verkko: telomere-to-telomere assembly of diploid chromosomes (preprint). *Bioinformatics*.
https://doi.org/10.1101/2022.06.24.497523
- Rhie A., Nurk S., Cechova M., Hoyt S.J., Taylor D.J., Altemose N., Hook P.W., Koren S., Rautiainen M., Alexandrov I.A., Allen J., Asri M., Bizikadze A.V., Chen N.-C., Chin C.-S. et al. 2022. The complete sequence of a human Y chromosome (preprint). *Genomics*.
https://doi.org/10.1101/2022.12.01.518724
- Rudd M.K., Willard H.F. 2004. Analysis of the centromeric regions of the human genome assembly. *Trends in Genetics*. V. 20. P. 529.
https://doi.org/10.1016/j.tig.2004.08.008
- Rudd M.K., Wray G.A., Willard H.F. 2006. The evolutionary dynamics of α -satellite. *Genome Res.* V. 16. P. 88.
https://doi.org/10.1101/gr.3810906
- Schueler M.G., Higgins A.W., Rudd M.K., Gustashaw K., Willard H.F. 2001. Genomic and Genetic Definition of a Functional Human Centromere. *Science*. V. 294. P. 109.
https://doi.org/10.1126/science.1065042
- She X., Horvath J.E., Jiang Z., Liu G., Furey T.S., Christ L., Clark R., Graves T., Gulden C.L., Alkan C., Bailey J.A., Sahinalp C., Rocchi M., Haussler D., Wilson R.K. et al. 2004. The structure and evolution of centromeric transition regions within the human genome. *Nature*. V. 430. P. 857.
https://doi.org/10.1038/nature02806
- Shepelev V.A., Alexandrov A.A., Yurov Y.B., Alexandrov I.A. 2009. The Evolutionary Origin of Man Can Be Traced in the Layers of Defunct Ancestral Alpha Satellites Flanking the Active Centromeres of Human Chromosomes. *PLoS Genet.* V. 5, e1000641.
https://doi.org/10.1371/journal.pgen.1000641
- Simpson J.T., Workman R.E., Zuzarte P.C., David M., Dursi L.J., Timp W. 2017. Detecting DNA cytosine methylation using nanopore sequencing. *Nat Methods*. V. 14. P. 407.
https://doi.org/10.1038/nmeth.4184
- Singer M.F. 1982. Highly Repeated Sequences in Mammalian Genomes. *International Review of Cytology*. V. 76. P. 67.
https://doi.org/10.1016/S0074-7696(08)61789-1
- Skakkebaek A., Gravholt C.H., Rasmussen P.M., Bojesen A., Jensen J.S., Fedder J., Laurberg P., Hertz J.M., Østergaard J.R., Pedersen A.D., Wallentin M. 2014. Neuroanatomical correlates of Klinefelter syndrome studied in relation to the neuropsychological profile. *NeuroImage: Clinical*. V. 4. P. 1.
https://doi.org/10.1016/j.nicl.2013.10.013
- Steinberg K.M., Schneider V.A., Graves-Lindsay T.A., Fulton R.S., Agarwala R., Huddleston J., Shiryev S.A., Morgulis A., Surti U., Warren W.C., Church D.M., Eichler E.E., Wilson R.K. 2014. Single haplotype assembly of the human genome from a hydatidiform mole. *Genome Res.* V. 24. P. 2066.
https://doi.org/10.1101/gr.180893.114
- Sullivan B.A., Wolff D.J., Schwartz S. 1994. Analysis of centromeric activity in Robertsonian translocations: implications for a functional acrocentric hierarchy. *Chromosoma*. V. 103. P. 459.
https://doi.org/10.1007/BF00337384
- Suzuki Y., Myers E.W., Morishita S. 2020. Rapid and ongoing evolution of repetitive sequence structures in human centromeres. *Sci. Adv.* V. 6. eabd9230.
https://doi.org/10.1126/sciadv.abd9230
- Tartaglia N.R., Howell S., Sutherland A., Wilson R., Wilson L. 2010. A review of trisomy X (47,XXX). *Orphanet J. Rare Dis.* V. 5. P. 8.
https://doi.org/10.1186/1750-1172-5-8
- Thakur J., Henikoff S. 2018. Unexpected conformational variations of the human centromeric chromatin complex. *Genes Dev.* V. 32. P. 20.
https://doi.org/10.1101/gad.307736.117
- Uralsky L.I., Shepelev V.A., Alexandrov A.A., Yurov Y.B., Rogaev E.I., Alexandrov I.A. 2019. Classification and monomer-by-monomer annotation dataset of suprachromosomal family 1 alpha satellite higher-order repeats in hg38 human genome assembly. *Data in Brief*. V. 24. P. 103708.
https://doi.org/10.1016/j.dib.2019.103708
- Vázquez-Diez C., FitzHarris G. 2018. Causes and consequences of chromosome segregation error in preimplantation embryos. *Reproduction*. V. 155. P. R63.
https://doi.org/10.1530/REP-17-0569
- Venter J.C., Adams M.D., Myers E.W., Li P.W., Mural R.J., Sutton G.G., Smith H.O., Yandell M., Evans C.A., Holt R.A., Gocayne J.D., Amanatides P., Ballew R.M., Huson D.H., Wortman J.R. et al. 2001. The sequence of the human genome. *Science*. V. 291. P. 1304.
https://doi.org/10.1126/science.1058040
- Vorsanova S.G., Kolotii A.D., Kurinnaia O.S., Kravets V.S., Demidova I.A., Soloviev I.V., Yurov Y.B., Iourov I.Y. 2021. Turner's syndrome mosaicism in girls with neurodevelopmental disorders: a cohort study and hypothesis. *Mol. Cytogenet.* V. 14. P. 9.
https://doi.org/10.1186/s13039-021-00529-2
- Vorsanova S.G., Demidova I.A., Kolotii A.D., Kurinnaia O.S., Kravets V.S., Soloviev I.V., Yurov Y.B., Iourov I.Y. 2022. Klinefelter syndrome mosaicism in boys with neurodevel-

- omental disorders: a cohort study and an extension of the hypothesis. *Mol. Cytogenet.* V. 15. P. 8.
<https://doi.org/10.1186/s13039-022-00588-z>
- Warburton P.E., Willard H.F. 1990. Genomic analysis of sequence variation in tandemly repeated DNA. *Journal of Molecular Biology.* V. 216. P. 3.
[https://doi.org/10.1016/S0022-2836\(05\)80056-7](https://doi.org/10.1016/S0022-2836(05)80056-7)
- Warburton P.E., Willard H.F. 1995. Interhomologue sequence variation of alpha satellite DNA from human chromosome 17: Evidence for concerted evolution along haplotypic lineages. *J. Mol. Evol.* V. 41.
<https://doi.org/10.1007/BF00173182>
- Wenger A.M., Peluso P., Rowell W.J., Chang P.-C., Hall R.J., Concepcion G.T., Ebler J., Fungtammasan A., Kolesnikov A., Olson N.D., Töpfer A., Alonge M., Mahmoud M., Qian Y., Chin C.-S. et al. 2019. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* V. 37. P. 1155.
<https://doi.org/10.1038/s41587-019-0217-9>
- Wevrick R., Willard H.F. 1991. Physical map of the centromeric region of human chromosome 7: relationship between two distinct alpha satellite arrays. *Nucl. Acids Res.* V. 19. P. 2295.
<https://doi.org/10.1093/nar/19.9.2295>
- Wevrick R., Willard H.F. 1989. Long-range organization of tandem arrays of alpha satellite DNA at the centromeres of human chromosomes: high-frequency array-length polymorphism and meiotic stability. *Proc. Natl. Acad. Sci. U.S.A.* V. 86. P. 9394.
<https://doi.org/10.1073/pnas.86.23.9394>
- Willard H.F. 1985. Chromosome-specific organization of human alpha satellite DNA. *Am. J. Hum. Genet.* V. 37. P. 524.
- Willard H.F., Wayne J.S. 1987. Hierarchical order in chromosome-specific human alpha satellite DNA. *Trends in Genetics.* V. 3. P. 192.
[https://doi.org/10.1016/0168-9525\(87\)90232-0](https://doi.org/10.1016/0168-9525(87)90232-0)

New in Centromere Genomics: Lessons from the First T2T Human Genome Assembly

L. I. Uralsky^{a, b, *}, I. A. Alexandrov^{b, c, d}, F. D. Ryabov^e, A. L. Lapidus^c, and E. I. Rogaev^{b, f}

^aSirius University of Science and Technology, Sirius, 354340 Russia

^bDepartment of Genomics and Human Genetics, Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, 119991 Russia

^cCenter for Bioinformatics and Algorithmic Biotechnology, Institute of Translational Biomedicine, Saint Petersburg State University, Saint Petersburg, 199034 Russia

^dResearch Center of Biotechnology of the Russian Academy of Sciences, Moscow, 119071 Russia

^eNational Research University, Higher School of Economics, Moscow, 109028 Russia

^fUMass Chan Medical School, Department of Psychiatry, Shrewsbury, MA, 01655 USA

*e-mail: uralskiy.li@talantiuspeh.ru

With the recent development of long-read sequencing technologies, it is now possible for the first time to read a complete gapless sequence of the human genome. The result was the first T2T (telomere-to-telomere) genomic assembly, published by an international consortium of scientists in 2022. The most significant contribution of the new assembly were the centromeric regions consisting of highly repetitive satellite DNA. In this review, we will briefly list the major achievements of the T2T consortium related to centromeres and take a closer look at the unexpected findings of cytogenetic magnitude that analysis of first assembled human centromeres has brought, such as the “split” centromeres of chromosomes 3 and 4, mega-inversion in the active centromere array of chromosome 1, haplotypic epialleles in the centromere of X chromosome and the macro-repeats found in several centromeres.

Keywords: alpha satellite, centromere, HOR-haplotype, inversion, macro-repeats, centromeric epialleles, T2T-consortium