

Теоретическая оценка нуклеосомной плотности на генных последовательностях различных ортологов при эухроматической и гетерохроматической локализации

Федосеева В.Б.*

Институт молекулярной генетики, Российская академия наук, Москва, Площадь И.В. Курчатова 2, 123182, Россия

Аннотация. Проведен сравнительный анализ ортологических вариантов нескольких генов, входящих в состав гетеро- и эухроматина из разных видов дрозофилид с использованием компьютерных методов, позволяющих вычислять потенциал предпочтительной локализации нуклеосом вдоль фрагментов ДНК и определять спектр размеров областей такого предпочтения и расстояний между их центрами. Выяснено, что общий уровень потенциала нуклеосомной локализации ниже в случае эухроматина, также есть отличия в спектрах линейных размеров областей нуклеосомного предпочтения. В случае гетерохроматина гены *light*, *concertina* включают интроны, которые обогащены короткими областями выраженного нуклеосомного предпочтения. Присутствие сильных пиков нуклеосомной локализации ассоциируется с множественными повторами класса LINE и LTR в области интронов, тогда как в области эухроматина их ортологи не содержат соответствующих повторов и столь же протяженных интронов. В последнем случае пики потенциала нуклеосомной локализации для гена *light* имеют не только меньшую амплитуду, но и размыты до большей ширины. При гетерохроматиновой локализации ген *concertina*, имеющий самый высокий средний уровень сигнала нуклеосомного позиционирования, экспрессируется на эмбриональной стадии и имеет слабо выраженный эффект положения гена, а ген *light* с относительно более низким средним уровнем сигнала жизненно необходим для многих стадий развития особи и имеет сильный эффект положения.

Ключевые слова: картирование нуклеосом, ортологи, экзон-интронная структура.

ВВЕДЕНИЕ

Хромосомы высших эукариот включают как эухроматин, так и гетерохроматин, что можно отличить, прежде всего, цитологически по разнице в уровне конденсации. Также гетерохроматин ассоциируется с такими биологическими признаками, как низкая частота рекомбинации, поздняя репликация, способность модулировать генную активность [1]. Ранее сформировались представления о неоднородности прицентромерного гетерохроматина по нескольким причинам: пятнистость экспрессии при встраивании репортерного гена в его регионы [2], зависимость уровня экспрессии от окружающих последовательностей или их свойств [3]. Сосуществование активных генов (*light*, *concertina*, AGO3 и др.) и транскрипционно молчащих областей [4], множество повторов разных классов и простых повторов также свидетельствуют о сложности гетерохроматина. Характерной чертой прицентромерного хроматина является специфическая модификация нуклеосомных гистонов H3K9me2 [5–7] и

*fvb@img.ras.ru

взаимодействие нуклеосомного ансамбля с гетерохроматиновыми белками (HP1 и его вариантами), узнающими эпигенетическую метку [8, 9]. Неоднородные профили связывания белка HP1 внутри прицентромерных областей [10] и аналогичные данные по эпигенетической модификации [11] свидетельствуют о разнородности гетерохроматина в деталях. Показано также, что инициация перехода в транскрипционно неактивное состояние осуществляется через взаимодействие малых РНК, синтезированных на ретротранспозонах, с мульти-белковым комплексом (члены семейства Argonaute, HP1 белок, специфичные к гистону H3 с эпигенетической меткой, диметилтрансферазы и другие) [12, 13]. Множественность таких инициаторных сайтов обеспечена высокой плотностью транспозонных элементов, дающих начало малым РНК [14, 15]. Также в прицентромерном гетерохроматине экспериментально наблюдались кластеры нуклеосом с характерной длиной в 10 единиц, что существенно для поддержания неактивного состояния вслед за инициацией, тогда как для остальной части генома – максимум 5–6 нуклеосом в кластере [4]. Основная интрига заключается в сосуществовании областей неактивного хроматина и генов, активных, по крайней мере, на отдельных этапах онтогенеза. Так как экспериментальные данные [10, 11] не всегда имеют достаточное разрешение на фоне общей масштабной картины, для детализации свойств генов, их активности, целесообразно дополнить экспериментальные исследования теоретическим подходом, заключающимся в использовании и анализе потенциала нуклеосомного картирования. Это позволит выявить подробности и сочетать с широкомасштабной картиной в целом. В частности, ранее использование этого метода позволило выявить некоторые признаки широкомасштабной периодичной обогащенности нуклеосомными кластерами в прицентромерном гетерохроматине для двух длинных хромосом *Drosophila melanogaster* [16].

В данной работе предполагается детализировать свойства прицентромерного гетерохроматина на примере нескольких генов. Особый интерес представляют собой выявленные пары генов, которые перешли из эухроматина в гетерохроматин в процессе эволюции в течение примерно 15–20 млн. лет. Так, ген *light (lt)* из *Drosophila virilis* и *Drosophila pseudoobscura* находится в эухроматической части геномов, тогда как его ортолог из *Drosophila melanogaster* - в гетерохроматической прицентромерной области, причем ген является жизненно важным и присутствует в однокопийном варианте во всех приведенных организмах [17, 18]. Похожая картина наблюдается для гена *concertina (cta)* и его ортологов из разных дрозофилид. Все белки соответствующих генов близки по аминокислотному составу. К тому же есть в клетке модификаторы, которые влияют на экспрессию гена *light*, когда его используют для изучения эффекта положения гена [4]. В настоящий момент для ортологов этого гена для всех трех организмов известны первичные нуклеотидные последовательности, которые можно использовать для сравнения экзон-интронных структур, и нуклеотидные повторы разных классов для *D. melanogaster*. Вызывает интерес сам факт их присутствия в гетерохроматиновой части и одновременной экспрессии хотя бы на какой-нибудь стадии развития. Для возможного объяснения этого предполагается провести сравнительный анализ различных количественных характеристик потенциала нуклеосомного позиционирования для ортологов двух генов из разных дрозофилид при их эухроматической и гетерохроматической локализации, по возможности выявить связь с генной экспрессией на стадиях онтогенеза.

При кратком обзоре среди подходов предсказания нуклеосомного позиционирования можно выделить два варианта расчетов: методы, основанные на учете индуцированных изгибов в нуклеосомном фрагменте ДНК, обусловленных нуклеотидной последовательностью, и статистические расчеты. Вначале остановлюсь на первом методе. Среди множества работ по изучению индуцированных белками изгибов в ДНК, в том числе гистоновыми октамерами, можно выделить очень

детализированные [19, 20], но ограниченные по разнообразию исследованных олигонуклеотидных мотивов, и работы, в которых представлены более сглаженные оценки изгибов для обширного набора нуклеотидных мотивов, достаточные для построения моделей связывания гистоновых октамеров с нуклеотидной последовательностью. Методы расчета изгибов нуклеосомной ДНК базируются на учете пространственных углов динуклеотидной пары на основе элементов матриц для ди- или тринуклеотидов. Массивы тринуклеотидов несут больше информации, чем массив данных для динуклеотидов, так как учитывают влияние соседей на величину угла изгиба «roll». В качестве инструмента для изучения индуцированных изгибов в ДНК использовали как саму нуклеосому, так и панкреатическую нуклеазу DNase I. Этот фермент эффективно связывается с мотивами ДНК, склонными к индуцированным изгибам [21–23], и в механизмах его взаимодействия с ДНК много общих моментов с гистоновыми октамерами. Для обоих типов белков экспериментально регистрировалось присутствие положительно заряженных аминокислотных остатков в малых бороздках ДНК [20–27]. Была установлена связь между расщеплением олигонуклеотидов с помощью DNase I и их структурными свойствами (способностью к изгибам, размерами малой бороздки) [21–23, 26–28]. Вполне естественно, что количественные характеристики эффективности расщепления [29] были использованы для учета индуцированных изгибов и при формировании нуклеосом, по крайней мере, в качестве одного из вариантов. Второй вариант матрицы для тринуклеотидов основан на использовании результатов теоретического реконструирования нуклеосом на массиве нуклеосомных фрагментов ДНК, выделенных из эритроцитов цыпленка с помощью микрококковой нуклеазы [30].

Ранее в работе [31] был предложен комбинированный метод определения положения нуклеосом относительно последовательности ДНК, в этом методе используются обе матрицы иерархически выстроенных тринуклеотидных вкладов в изгиб ДНК [29, 30]. Подробности метода заключались в следующем. Вначале производился расчет изгиба как векторной суммы на основе единичных индуцированных углов «roll», соответствующих каждому тринуклеотиду, для коротких блоков, входящих в состав нуклеосомной ДНК, с учетом общепринятых параметров спирали ДНК для В-формы. Из-за отсутствия прямых данных о значениях углов «roll» для раствора в качестве приближения, как и в работе [32], при выборе значения индуцированных углов «roll» в определенном угловом диапазоне их считали пропорциональными элементам матрицы для тринуклеотидов, полученным на основе эффективности расщепления DNase I в районе тринуклеотидов. Затем последовательно через преобразования координат при переходе между соседними блоками, каждый из которых накопил свой локальный изгиб, от первого до последнего формировалась величина общего изгиба всего нуклеосомного фрагмента ДНК. Выходные данные (потенциал нуклеосомного позиционирования) – это угловое отклонение вектора нормали к плоскости последнего нуклеотидного основания относительно нормали к первому нуклеотидному основанию. Рассчитанные угловые отклонения тестировались на выбранных фрагментах с множеством нуклеосом в их контексте с фланкирующими последовательностями, опубликованными в литературе [33–44], в их числе значились фрагменты, происходящие из разных организмов. В этой же работе определялись предпочтения для каждой из матриц (в первом приближении они зависели от АТ-состава фрагмента), вероятность предсказания, ошибки I и II рода (вероятность предсказания – около 70%, обе ошибки – около 30% на уровне 50% от разницы максимального и минимального значения сигнала нуклеосомного позиционирования для всего фрагмента), также производилось сравнение с другими методами предсказания. На момент публикации статьи метод предсказания превосходил другие варианты [45–47], основанные на динуклеотидных матрицах, так же как и при

сравнении с результатами вычислений по программам, доступным для использования на момент работы с материалами.

В последующее время накопилось много данных по статистическому подходу к нуклеосомной локализации для дрожжей, интерпретации результатов их участия в генной регуляции [48–53], вкладу эпигенетических меток и вариантных гистонов [54, 55], и данные по нуклеосомной локализации для геномов мухи и человека [56, 57]. Вкратце этот подход включает обработку генома микрококковой нуклеазой, амплификацию выделенных нуклеосомных фрагментов, картирование с помощью секвенирования концов нуклеосомной ДНК. Иногда они дополняются методами иммунопреципитации с антителами для вариантных гистонов или для основных гистонов, несущих эпигенетическую метку, в ряде случаев использовали предварительное сшивание клеточной ДНК с гистонами с помощью формальдегида. Это дорогостоящие высокотехнологичные процедуры на основе самых современных высокоскоростных методов секвенирования ДНК нового поколения. Для последующего распознавания нуклеосомной локализации на основании обучающей выборки, полученной из пула нуклеосомной ДНК одного организма, вычислялись динуклеотидные частоты с учетом позиций по длине нуклеосомного фрагмента с целью построения вероятностной модели нуклеосомного предпочтения. Была создана предсказывающая программа с использованием обучающей выборки на основе нуклеосом *Saccharomyces cerevisiae* [58]. Резюмируя представление статистического подхода, можно выделить его несомненное достоинство, заключающееся в привлечении огромного объема статистических данных, но следует учесть при этом, что в системе *in vivo* в общий пул попадают нуклеосомные фрагменты со сдвинутым положением по сравнению с тем, что обусловлено нуклеотидной последовательностью. Этот сдвиг может быть вызван как транскрипционными факторами, белками-ремодуляторами, так и самим транскрипционным комплексом [59, 60]. Нужно еще отметить, что разные организмы сильно варьируют по ГЦ-составу. Так, геном дрожжей достаточно АТ-богатый и однородный по составу в этом отношении, в отличие от нуклеотидного состава геномов цыпленка, мухи, более ГЦ-богатых и отличающихся большим разбросом по этому параметру [61]. Все это ослабляет предсказательный потенциал для гетерологических систем, что отмечено в работе [62], в которой проводилось сравнение предсказаний нуклеосомного сигнала с экспериментальными данными, как для самой системы дрожжей, так и для гетерологических систем. В частности, наборы динуклеотидных предпочтений, выявленные при статистическом анализе нуклеосомных локализации, отличаются для таких организмов, как дрожжи и муха [52, 56, 58].

МЕТОДЫ

Как и в вышеупомянутом комбинированном методе определения положения нуклеосом относительно последовательности ДНК [31], в данной работе были использованы две матрицы иерархически выстроенных тринуклеотидных вкладов в изгиб ДНК [29, 30] в комбинациях, зависящих от АТ-состава текущего нуклеосомного фрагмента (145 п.н.) (для АТ-богатых фрагментов для угла «roll» - данные из работы [29], для ГЦ-богатых – данные из работы [30]). Как показано в работе [31], данный метод при сравнении с сериями экспериментальных данных приводит к наилучшему соотношению истинно предсказанных, ложно предсказанных и непредсказанных, но экспериментально выявляемых, мест нуклеосомного позиционирования (критерий минимальности ошибок I и II рода при их приблизительной равнозначности) по сравнению с другими методами.

В качестве результата в этом методе используется потенциал $P(x)$ позиционирования нуклеосом вдоль фрагмента ДНК (5'-конец нуклеосомной ДНК) в

относительных единицах, которые соответствуют угловому отклонению между векторами нормали к нуклеотидным плоскостям 1-го и последнего нуклеотида (учитывается центральный участок длиной 124 пары нуклеотидов как наиболее значимый). После усреднения $P(x)$ (ширина окна усреднения выбирается равной 200 п.н.) график приобретает вид чередующихся пиков разной ширины и высоты с участками падения этого потенциала, что наводит на мысль о возможности сведения к суперпозиции синусоидальных функций с различными длинами волн. Для выяснения возможности существования закономерности такого чередования, как и в работе [16], получаем спектр корреляционных коэффициентов, аналогов коэффициентов Фурье, числитель которых $S(\lambda)$ вычисляется как максимум по индексу k :

$$S(\lambda) = \max_k \left(\sum_i \sin \left(\frac{2\pi x_i}{\lambda} + \varphi_k \right) P(x_i) \right), \quad \varphi_k = \frac{\pi}{6} k, \quad 0 \leq k \leq 11,$$

где λ - длина волны (в дальнейшем: в тысячах пар нуклеотидов (т.п.н.)), а φ - фаза в диапазоне $[0, 2\pi]$, дискретность фазы $\pi/6$. Суммирование проводится по индексу i , соответствующему дискретности представления потенциала в вычислительной программе, а x_i - текущая координата вдоль фрагмента ДНК, в знаменателе корреляционных коэффициентов - традиционная нормировка. Чтобы не существовало проблемы в нормировке спектральных пиков по амплитуде, предлагается выбирать изучаемые фрагменты приблизительно одинаковой длины. На рисунках, содержащих спектральные характеристики, соответствующих фрагментам изучаемых ДНК из разных источников, представлены коэффициенты в зависимости от длины волны, вычисляемых в тысячах пар нуклеотидов. Каждый пик в спектре будет соответствовать расстоянию между центрами объединения нуклеосом в пространственные группы. Если такие пики встречаются в длинноволновой части спектра, то их можно рассматривать как предсказания возможности формирования хромосомного материала повышенной плотности.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

На рисунке 1 представлены потенциалы нуклеосомной локализации на ДНК двух фрагментов, включающие разные ортологи одного и того же гена *light* из 3 близкородственных организмов *D. melanogaster*, *D. virilis* и *D. pseudoobscura*, которые эволюционно разошлись 15–20 млн. лет назад, их экзон-интронная структура и известные нуклеотидные повторы для *D. melanogaster*. Как показано в работах [17, 18], первый из ортологов имеет признаки структуры гетерохроматина, тогда как второй и третий явно соответствуют эухроматину. Видно, что ярко выраженные пики соответствуют участкам интронов, для *D. melanogaster* некоторые пики соответствуют повторам. Основные отличия заключаются в том, что средний уровень потенциала позиционирования для *light* из *D. virilis*, *D. pseudoobscura* меньше на ~15% по сравнению с *light* из *D. melanogaster*, и пики заметно отличаются по высоте и по ширине, что проанализировано ниже с помощью метода спектрального анализа.

В связи с присутствием выраженных областей повышенного уровня нуклеосомной локализации в приведенном гене, особенно в интронах, для сравнения из базы данных [63] для последовательностей 2-ой и 3-ей хромосомы *D. melanogaster* были отобраны гены, содержащие протяженные интроны, большинство длиной порядка 1–3 т.п.н.. На рисунке 2 приведены типичные примеры из множества генов (CG14747, CG13211, CG13169) 2R хромосомы *D. melanogaster*, их экзон-интронная структура и профили потенциалов нуклеосомной локализации вдоль ДНК. Видно, что интронам соответствует повышенный потенциал по сравнению участками кодирования - экзонами. Линейные размеры участков повышенного потенциала (пиков) находятся в

диапазоне 500–1000 п.н. Эта ситуация вполне типична для многих других генов, включающих в себя протяженные интроны.

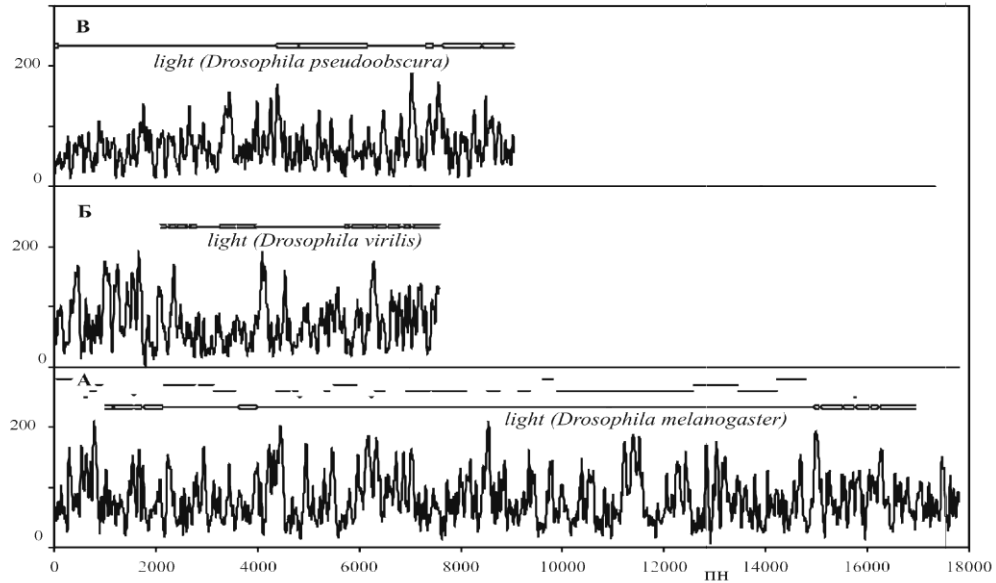


Рис. 1. Потенциал нуклеосомной локализации $P(x)$, экзон-интронные схемы ортологов гена *light* (*lt*) из *D. melanogaster* (А), *D. virilis* (Б), *D. pseudoobscura* (В). А. Внизу на рисунке изображен потенциал нуклеосомного предпочтения для фрагмента ДНК ~ 17 тысяч пар нуклеотидов (т.п.н.) длиной (ген *light* (*lt*) из *D. melanogaster*), выше схема гена представлена в виде экзонов (утолщение) и интронов (тонкая линия), на разных уровнях тонкой линией изображены повторы в соответствии с их размерами и локализацией. АТ-богатые повторы на уровне 250 относительных единиц, LINE – 260, DNA – 270, PaoLTR – 280. По оси X – текущая координата внутри фрагмента в парах нуклеотидов (п.н.), по оси Y – $P(x)$ в относительных единицах (о.е.) после усреднения в окне 200 п.н. Б. То же для гена *light* из *D. virilis* В. То же для гена *light* из *D. Pseudoobscura*.

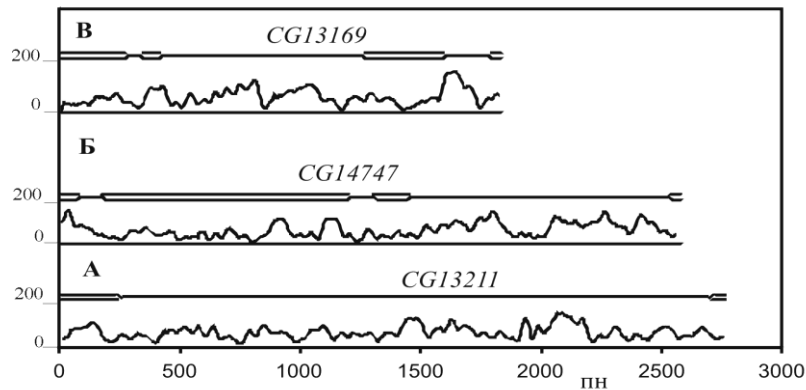


Рис. 2. Потенциал нуклеосомного позиционирования $P(x)$ для ряда генов *D. melanogaster* и их экзон-интронные схемы. Сдвоенные линии соответствуют эксонам, а одинарные – интронам. По оси X – текущая координата внутри фрагмента в парах нуклеотидов (п.н.), по оси Y – $P(x)$ в относительных единицах (о.е.) после усреднения в окне 200 п.н.

Пространственные спектры нуклеосомного потенциала представлены на рисунке 3, по оси X - длина волны в тысячах нуклеотидных пар (т.н.п.), по Y - потенциал нуклеосомного позиционирования в относительных единицах (о.е.). В рассмотрении используются фрагменты ДНК одинаковой длины (~17 т.н.п.) из разных ортологов гена *light* (А–В для *D. erecta*, *D. ananassae*, *D. melanogaster*, Д, Е – для *D. virilis* и *D. yakuba*), для фрагмента ДНК в районе гена *light* длиной 30 т.н.п. (Г для *D. melanogaster*). При сравнении спектров, соответствующих гену *light* из *D. melanogaster*, с его ортологами из *D. virilis* и *D. pseudoobscura* видно, что первый спектр для гена, последовательность которого содержит протяженный интрон 11 т.п.н., свидетельствует о множественных

пиках нуклеосомного потенциала шириной 0,5–1 т.п.н. (расстояния между ними находятся в этом же диапазоне), тогда как ген *light* из *D. pseudoobscura* и *D. virilis* вместе с его окружением имеет в спектре слабовыраженные пики с длиной волны 2 и 4 т.п.н.. Ген *light* (*D. melanogaster*), как следует из базы данных, представленной на сайте [61], в области протяженного интрона имеет повторы класса LINE, LTR и DNA, разброс которых по размерам находится в диапазоне 200–500 п.н., что и создает разнообразие пиков нуклеосомного потенциала. Каждый пик при этом не обязательно соответствует одному повтору, а может возникнуть на стыке нескольких повторов или быть несвязанным с ними совсем. Ген *light* из *D. pseudoobscura* имеет гораздо меньшие по размеру интроны, не содержащие транспозонных элементов и других повторов. В целом для всех генов кодирующей части соответствуют меньшие по величине пики нуклеосомного потенциала.

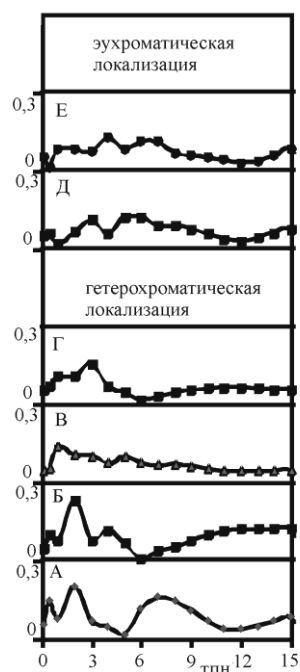


Рис. 3. Спектр линейных размеров областей предпочтительной нуклеосомной локализации для генов из разных видов. Для А–В, Д и Е длина фрагментов ~17 т.п.н. По оси X – длина волны в тысячах пар нуклеотидов (т.п.н.), по оси Y – корреляционные коэффициенты (диапазон 0,3). А. Спектр для гена *light* (*lt*) из *D. erecta*. Б. Спектр для *light* из *D. ananassae*. В. Спектр для *light* из *D. melanogaster*. Г. Спектр для *light* из *D. melanogaster* (длина фрагмента ~30 т.п.н.). Д. Спектр для *light* из *D. virilis*. Е. Спектр для *light* из *D. yakuba*.

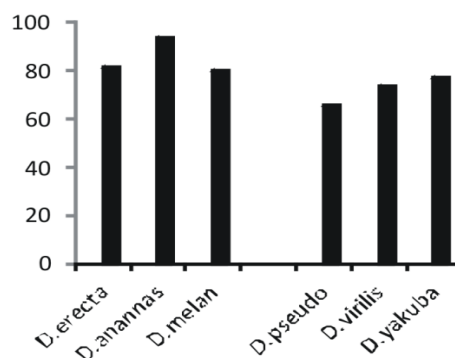


Рис. 4. Гистограмма средних значений потенциала нуклеосомной локализации $P(x)$ в относительных единицах (о.е.) для гена *light* из разных видов дрозофилид (*D. erecta*, *D. ananassae*, *D. melanogaster*, *D. pseudoobscura*, *D. virilis*, *D. yakuba*).

Из представленных на гистограмме (рисунок 4) данных видно, что для более близких к *D. melanogaster* видов *D. erecta*, *D. ananassae* средние значения потенциала позиционирования превосходят эти же значения для более разошедшихся с ними по времени видов *D. virilis*, *D. pseudoobscura*. Потенциал для *D. yakuba* занимает промежуточное значение в соответствии с временной шкалой расхождения видов, представленной в работе [18].

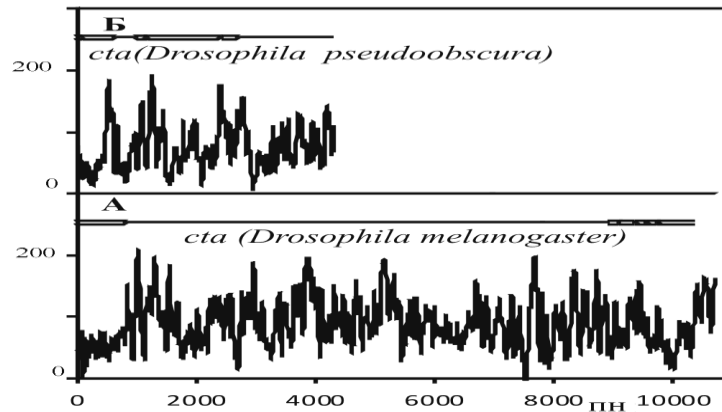


Рис. 5. Потенциал нуклеосомного позиционирования для гена *concertina* (*cta*) *D. melanogaster*, *D. pseudoobscura* и их экзон-интронные схемы. Сдвоенные линии соответствуют эксонам, а одинарные – интронам. По оси X – текущая координата внутри фрагмента в парах нуклеотидов (п.н.). По оси Y потенциал представлен в относительных единицах (о.е.) в том же масштабе, что на рисунке 1.

Для сравнения нуклеосомного потенциала для гена *light* на рисунке 5 приведены профили потенциала для гена *concertina* из *D. melanogaster* из того же гетерохроматинового контига, что и ген *light*. Ген *concertina* имеет также протяженный интрон длиной ~ 8 т.п.н., и этому соответствуют особенности спектра, а именно, наличие ярко выраженных пиков (в коротковолновой части спектра пик ~ 1 т.п.н., и широкий пик с центром в точке 5 т.п.н. в средней части спектра, сам спектр не приведен). Этот же ген из *D. pseudoobscura* имеет меньший интрон в соответствии с эухроматическим расположением. Существенно, что средний уровень нуклеосомного потенциала для гена *concertina* выше, чем для гена *light* (88 против 80 о.е.). В связи с этим обращает на себя внимание тот факт, что продукты гена *concertina* выявлены для эмбриональной стадии развития *D. melanogaster*, тогда как экспрессия гена *light* существенна для разных стадий развития.

Было проведено сравнение потенциала нуклеосомного позиционирования для генов *light* и *concertina* (*D. melanogaster*) с нуклеосомными фрагментами ДНК, выявленными экспериментально. Пики нуклеосомного сигнала совпадают по многим позициям с нуклеосомной локализацией [64], что можно считать вполне удовлетворительным, учитывая, что в экспериментальных работах представлены 2 группы данных по локализации: для нуклеосом с базовым набором гистонов и с H2A.Z гистоном, при том, что экспериментальные данные не всегда отличаются полнотой представления. Значимые совпадения наблюдаются при сравнении с картой локализации гетерохроматического белка HP1, связывающегося с нуклеосомой в гетерохроматине при наличии метки H3K9me3 [10]. При сравнении локализации нуклеосом [64] с расчетными данными, полученными при использовании программы из работы [54], также наблюдается совпадение с отдельными пиками.

ОБСУЖДЕНИЕ

При сравнении гена *light* с учетом входящих в него интронов и прилегающих к нему последовательностей в составе хроматина в двух вариантах, гетеро- и эухроматина, можно сделать вывод об определенном отличии в среднем уровне потенциала нуклеосомного позиционирования и более сильном отличии в пространственных линейных размерах фрагментов ДНК, имеющих предпочтение в нуклеосомном связывании в целом (кластерах). Так, в гетерохроматической области больше фрагментов длиной ~ 1000 п.н. по сравнению с эухроматином. Такой фрагмент ДНК может вместить максимум 6–7 нуклеосом на базе гистоновых октамеров. Биологическая функция таких нуклеосомных образований может быть как сближение каких-нибудь регуляторных последовательностей, находящихся на расстоянии ~ 1000 п.н., к примеру, энхансеров с промотором, так и формирование наднуклеосомных структур, как, например, в современной модели гетерохроматина. Замечательно, что ранее были известны экспериментальные данные о более сильной зависимости свойств, присущих гетерохроматину, от общего уровня концентрации гистонов в клетке, точнее, дозы генов гистонов [65], что находится в соответствии с более высоким уровнем сродства гистоновых октамеров к ДНК в гетерохроматине по сравнению с эухроматином. Образование фрагментов нуклеосомного предпочтения такого размера – это лишь одна краска в палитре линейных размеров нуклеосомных скоплений, как следует из сравнения с другими генами. Могут встречаться фрагменты длиной 6–8 т.п.н., имеющие нуклеосомное предпочтение по всей длине в целом. Возможно, это предпочтение есть предпосылка к поддержанию неких компактных образований с подключением белков-модификаторов, например, деацетилазы, метилтрансферазы, гетерохроматинового белка типа HP1, малых РНК, которые делают нуклеосомный кластер более компактным и инициируют переход к закрытому для транскрипции состоянию. Для сравнения, при гетерохроматиновой локализации гена *light* его средний уровень сигнала нуклеосомного позиционирования равен 80 о.е. (D.m.), при эухроматической – 73 о.е. (*D. virilis*.) и 66 о.е. (*D. pseudoobscura*); для гена *concertina* – 88 о.е. (*D. melanogaster*). Если ген *light* необходим как для нормальных уровней пигментации в ряде тканей взрослой особи и личиночной, так и жизнеспособности в целом, то функциональная активность гена *concertina* выявлена для эмбриональной стадии, для остальных же стадий развития не выявлена [66]. Так, известно, что на эмбриональной стадии происходит скачок в уровне экспрессии гистоновых генов, составляющих нуклеосому, что не противоречит экспрессии генов с высоким уровнем нуклеосомного потенциала в стадии, предшествующей скачку. Именно такая ситуация складывается для гена *concertina*. Именно, для этого гена характерен более высокий уровень сигнала, и ген активен на эмбриональном уровне [66]. Возможность экспрессии гена *light* в некоторых тканях во взрослом состоянии, локализованного в гетерохроматине, и пятнистость экспрессии в других областях эухроматина (эффект положения гена) [67] коррелирует с не слишком высоким локальным уровнем нуклеосомного потенциала. Можно говорить о двойственных свойствах гена *light* по сравнению с геном *concertina*, для которого эффект положения гена слабо выражен [67]. Это находится в соответствии с представлением о неравномерности нуклеосомного потенциала в целом, характерным для прицентромерного гетерохроматина. Как показано в работе [16], для прицентромерного гетерохроматина характерен кластерный вариант повышенной нуклеосомной плотности, ген *light*, как оказалось, локализуется в промежутке между 2 кластерами более высокого уровня сигнала, чем нуклеосомный потенциал самого гена. Для этого гена нельзя исключить варианта, вызванного возможно повышенным уровнем модифицирующих белков и белков-демодуляторов, направленных на переход в открытое состояние. Оба гена содержат повторы, включающие концевые инвертированные повторы (LTR), что

свидетельствует в пользу возможности инициации перехода в закрытое состояние [12–14]. В целом, выяснению нуклеосомных предпочтений в более широком масштабе может способствовать такой вычислительный подход, особенно в связи с наметившейся корреляцией с временным проявлением экспрессии генов на стадиях от эмбриона до зрелой особи.

СПИСОК ЛИТЕРАТУРЫ

1. Weiler K.S., Wakimoto B.T. Heterochromatin and gene expression in *Drosophila*. *Annu. Rev. Genet.* 1995. V. 29. P. 577–605.
2. Schotta G., Ebert A., Dorn R., Reuter G. Position-effect variegation and the genetic dissection of chromatin regulation in *Drosophila*. *Semin. Cell Dev. Biol.* 2003. V. 14. P. 67–75.
3. Hower M., Dimitri P., Berloco M., Wakimoto B. Cis-effects of heterochromatin on heterochromatic and euchromatic gene activity in *Drosophila melanogaster*. *Genetics*. 1995. V. 140. P. 1033–1045.
4. Sun F.-L., Cuaycong M.H., Elgin S.C.R. Long-range nucleosome ordering is associated with gene silencing in *Drosophila melanogaster* pericentric heterochromatin. *Mol. Cell. Biol.* 2001. V. 21. P. 2867–2879.
5. Noma K., Allis C.D., Grewal S.I. Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. *Science*. 2001. V. 293. P. 1150–1155.
6. Verschure P., Van der Kraan I., De Leeuw W., Van der Vlag J., Carpenter A.E., Belmont A.S., Van Drier R. *In vivo* HP1 targeting causes large-scale chromatin condensation and enhanced histone lysine methylation. *Mol. Cell. Biol.* 2005. V. 25. P. 4552–4564.
7. Stewart M.D., Li J., Wong J. Relationship between histone H3 lysine 9 methylation, transcription repression and heterochromatin protein1 recruitment. *Mol. Cell. Biol.* 2005. V. 25. P. 2525–2538.
8. Jacob S.A., Khorasanizadeh S. Structure of HP1 chromodomain bound to lysine 9-methylated histone H3 tail. *Science*. 2002. V. 295. P. 2080–2083.
9. Nielsen P.R., Nietlespach D., Mott H.R., Callaghan J., Bannister A., Kouzarides T., Murzin A.G., Mursin N.V., Laue E.D. Structure of the HP1 chromodomain bound to histone h3 methylated at lysine 9. *Nature*. 2002. V. 416. P. 103–107.
10. De Wit E., Greil F., van Steensel B. High-resolution mapping reveals links of HP1 with active and inactive chromatin components. *PLoS Genetics*. 2007. V. 3. № 3. P. e38.
11. Riddle N.C., Monoda A., Kharchenko P.V., Alekseyenko A.A., Schwartz Y.B., Tolstorukov M.Y., Gorchakov A.A., Jaffe J.D., Kennedy C., Linder-Basso D. *et al.* Plasticity in patterns of histone modifications and chromosomal proteins in *Drosophila* heterochromatin. *Genome Research*. 2011. V. 21. P. 147–163.
12. Pal-Bhadra M., Leibovitch B.A., Gandhi S.G., Chikka M.R., Bhadra U., Birchler J.A., Elgin S.C. Heterochromatic silencing and HP1 localization in *Drosophila* are dependent on the RNAi machinery. *Science*. 2004. V. 303. P. 669–672.
13. Fagegaltier D., Bouge A.-L., Berry B., Poisot E., Sismeiro O., Coppee J.-Y., Theodore L., Voinnet O., Antoniewski C. The endogenous siRNA pathway is involved in heterochromatin formation in *Drosophila*. *Proc. Natl. Acad. Sci. USA*. 2009. V. 106. P. 21258–21263.
14. Aravin A.A., Lagos-Quintana M., Yalcin A., Zavolan M., Marks D., Snyder D., Gaasterland T., Meyer J., Tuschl T. The small RNA profile during *Drosophila melanogaster* development. *Dev. Cell*. 2003. V. 5. P. 337–350.
15. Reihart B.J., Bratel D.P. Small RNAs correspond to centromere heterochromatin repeats. *Science*. 2002. V. 297. P. 1831–1838.

16. Fedoseyeva V.B., Alexandrov A.A. Large-scale periodicity of nucleosome positioning signal in pericentric regions of chromosomes (*Drosophila melanogaster*). *J. Biomol. Struct. Dynam.* 2013. doi:10.1080/0739110.2013.844081.
17. Devlin R.H., Brendan B., Wakimoto B.T. The organization and expression of the *light* gene, a heterochromatic gene of *Drosophila*. *Genetics*. 1990. V. 123. P. 129–140.
18. Yasuhara J.C., DeCrease C.H., Wakimoto B.T. Evolution of heterochromatic genes of *Drosophila*. *Proc. Natl. Acad. Sci. USA*. 2005. V. 102. P. 10958–10963.
19. Tolstorukov M.Y., Colasanti A.V., McCandlish D.M., Olson W.K., Zhurkin V.B. A novel roll and slide mechanism of DNA folding in chromatin: implication for nucleosome positioning. *J. Mol. Biol.* 2007. V. 371. P. 725–738.
20. Wang D., Ulyanov N.B., Zhurkin V.B. Sequence-dependent kink-and-slide deformations of nucleosome DNA facilitated by histone arginines bound in the minor groove. *J. Biomol. Struct. Dynam.* 2010. V. 27. P. 843–859.
21. Lahm A., Suck D. DNase I –induced DNA conformation 2A structure of the DNase I-octamer comple. *J. Mol. Biol.* 1991. V. 222. P. 645–667.
22. Suck D. DNA recognition by DNase I. *J. Mol. Recognit.* 1994. V. 7. P. 65–70.
23. Weston S.A., Lahm A., Suck D. X-ray structure of the DNase I-d(GGTATACC)₂ complex at 2.3 Å resolution. *J. Mol. Biol.* 1992. V. 226. P. 1237–1256.
24. Luger K., Mader A.W., Richmond R.K., Sargent D.F., Richmond T.J. Crystal structure of nucleosome core particle 2.8 Å resolution. *Nature*. 1997. V. 389. P. 251–260.
25. Oguel C., Foloppe N., Hartmann B. Understanding the sequence dependence of DNA groove dimensions: implications for DNA interactions. *PLoS One*. 2010. V. 5. P. e15931. doi:10.1371/journal.0015931.
26. Drew H.R., Travers A.A. DNA structural variations in the *E.coli tyrT* promoter. *Cell*. 1984. V. 37. P. 491–502.
27. Drew H.R., Travers A.A. DNA bending and its relation to nucleosome positioning. *J. Mol. Biol.* 1985. V. 186. P. 773–790.
28. Lomonosoff G. P., Butler P.J.G., Klug A. Sequence dependent variation in the conformation of DNA. *J. Mol. Biol.* 1984. V. 149. P. 745–760.
29. Bruckner I., Sanchez R., Suck D., Pongor S. Sequence –dependent bending propensity of DNA as revealed by DNase I: parameters for trinucleotides. *EMBO J.* 1995. V. 14. P. 1812–1818.
30. Satchwell S.C., Drew H.R., Travers A.A. Sequence periodicities in chicken nucleosome. *J. Mol. Biol.* 1986. V. 191. P. 659–675.
31. Fedoseyeva V.B., Alexandrov A.A. Analysis and development of the computer methods of nucleosome localization on DNA fragments with different AT-content. *J. Biomol. Struct. Dynam.* 2007. V. 24. P. 481–488.
32. Bruckner I., Sanchez R., Suck D., Pongor S. Trinucleotide models for DNA bending propensity: comparison of models based on DNase I digestion and nucleosome packaging data. *J. Biomol. Struct. Dynam.* 1995. V. 13. P. 309–317.
33. Martinez-Garcia J.F., Estruch E., Perez-Ortin J.E. Chromatin structure of the 5' flanking region of the yeast LEU2 gene. *Mol. Gen. Genet.* 1989. V. 217. P. 464–470.
34. Shimizu M., Roth S.Y., Szent-Gyorgui C., Simpson R.T. Nucleosomes are positioned with base pair precision adjacent to the alpha 2 operator in *Saccharomyces cerevisiae*. *EMBO J.* 1991. V. 10. P. 3033–3041.
35. Roberson A., Wolffe A.P., Hauserans L.J., Olins D.E. The 5S RNA gene minichromosome of *Euplotes*. *Nucl. Acids Res.* 1989. V. 17. P. 4699–4712.
36. Levy-Wilson B., Fortier C., Blackhart B.D., McCarthy. DNase I and micrococcal nuclease-hypersensitive sites in the human apolipoprotein B gene are tissue specific. *Mol. Cell Biol.* 1988. V. 8. P. 71–80.

37. Thoma F., Bergman L.W., Simpson R.T. Nuclease digestion of circular TRP1ARS1 chromatin reveals positioned nucleosomes separated by nuclease-sensitive regions. *J. Mol. Biol.* 1984. V. 177. P. 715–733.
38. Thoma F. Protein-DNA interactions and nuclease-sensitive regions determine nucleosome positions on yeast plasmid chromatin. *J. Mol. Biol.* 1986. V. 190. P. 177–190.
39. Drew H.R., Calladine C. R. Sequence-specific positioning of core histones on an 860 base-pair DNA. Experiment and theory. *J. Mol. Biol.* 1987. V. 195. P. 143–173.
40. Almer A., Horz W. Nuclease hypersensitive regions with adjacent positioned nucleosomes mark the gene boundaries of the PHO5/PHO3 locus in yeast. *EMBO J.* 1986. V. 5. P. 2681–2687.
41. Almer A., Rudolph H., Hinnen A., Horz W. Removal of positioned nucleosomes from the yeast PHO5 promoter upon PHO5 induction releases additional upstream activating DNA elements. *EMBO J.* 1986. V. 5. P. 2689–2696.
42. Capranico G., Jaxel C., Roberge M., Kohn K.W., Pommier T. Nucleosome positioning as a critical determinant for the DNA cleavage sites of mammalian DNA topoisomerase II in reconstituted simian virus 40 chromatin. *Nucl. Acids Res.* 1990. V. 18. P. 4553–4559.
43. Stein A. Unique positioning of reconstituted nucleosomes occurs in one region of simian virus 40 DNA. *J. Biol. Chem.* 1987. V. 262. P. 3872–3879.
44. Tanaka S., Livingstone-Zatchej M., Thoma F. Chromatin structure of the yeast *URA3* gene at high resolution provides insight into structure and positioning of nucleosomes in the chromosomal context. *J. Mol. Biol.* 1996. V. 257. P. 929–934.
45. Mengeritsky G., Trifonov E.N. Nucleotide sequence-directed mapping of the nucleosomes. *Nucl. Acids Res.* 1983. V. 11. P. 3833–3851.
46. Levitsky V.G., Ponomarenko M.P., Ponomarenko J.V., Frolov A.C., Kolchanov N.A. Nucleosomal DNA property database. *Bioinformatics.* 1999. V. 15. P. 582–592.
47. Levitsky V.G., Podkolodnaya O.A., Kolchanov N.A., Podkolodny N.I. Nucleosome formation potential of eukaryotic DNA: calculation and promoters analysis. *Bioinformatics.* 2001. V. 17. P. 998–1010.
48. Segal E., Fondufe-Mittendorf V., Chen L., Thasstrom A., Field Y., More I.K., Wang J.P., Widom J. A genomic code for nucleosome positioning. *Nature.* 2006. V. 442. P. 772–778.
49. Ioshikhes I.P., Albert I., Zanton S.J., Pugh B.F. Nucleosome positions predicted through comparative genomics. *Nat. Genet.* 2006. V. 38. P. 1210–1215.
50. Lee W., Tillo D., Bray N., Morse R.H., Davis R.W., Hughes T.R., Nislow C. A high-resolution atlas of nucleosome occupancy in yeast. *Nature Genetics.* 2007. V. 39. P. 1235–1244.
51. Yuan G.-C., Lui Y.-J., Dion M.F., Clack M.D., Wu L.F., Altschuler S.J., Rando O.J. Genome-scale identification of nucleosome positions in *S.cerevisiae*. *Science.* 2005. V. 309. P. 626–630.
52. Shruhl K. Naturally occurring poly(dA-dT) sequences are upstream promoter elements for constitutive transcription in yeast. *Proc Natl. Acad. Sci. USA.* 1985. V. 82. P. 8419–8423.
53. Yuan G.C., Lui J.S. Genomic sequence is highly predictive of local nucleosome depletion. *PLoS Comput. Biol.* 2008. V. 4. P. e13. doi:10.1371/journal.pcbi.0040013.
54. Barski A., Cuddapah S., Cui K., Roh T.Y., Schones D.E., Wang Z., Wei G., Chepelev I., Zhao K. High resolution profiling of histone methylations in the human genome. *Cell.* 2007. V. 129. P. 823–837.
55. Albert I., Mavrich T.N., Tomsho L.P., Qi J., Zanton S.J., Schuster S.C., Pugh B.F. Translational and notational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome. *Nature.* 2007. V. 446. P. 572–576.

56. Mavrich T.N., Jiang V., Ioshikhes I.P., Li X., Venters B.J., Zanton S.J., Tomsho L.P., Qi J., Glasen R., Schuster S.C. *et al.* Nucleosome organization in the *Drosophila* genome. *Nature*. 2008. V. 453. P. 358–362.
57. Schones D.E., Cui K., Cuddapah S., Roh T.Y., Barski A., Wang Z., Wei G., Zhao K. Dynamic regulation of nucleosomes positioning in the human genome. *Cell*. 2008. V. 132. P. 887–898.
58. Field Y., Kaplan N., Fondufe-Mittendorf Y., Moore I.K., Sharon E., Lubling Y., Widom J., Segal E. Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS Comput. Biol.* 2008. V. 4. P. e1000216. doi:10.1371/journal.pcbi.1000216.
59. Schwabish M.A., Struhl K. Evidence for eviction and rapid deposition of histones upon transcription elongation by RNA polymerase II. *Mol. Cell Biol.* 2004. V. 24. P. 10111–10117.
60. Schnitzler R. Control of nucleosome position by DNA sequence and remodeling machine. *Cell Biochem. Biophys.* 2008. V. 51. P. 6780–6788.
61. *UCSC Genome Bioinformatics Site*. URL: <http://www.genome.ucsc.edu> (дата обращения: 18.01.2012).
62. Peckham H.E., Thurman R.E., Fu Y., Stamatoyannopoulos J.A., Noble W.S., Struhl K., Weng Z. Nucleosome positioning signals in genomic DNA. *Genome Res.* 2007. V. 17. P. 1170–1177.
63. *National Center for Biotechnology Information*. URL: <http://www.ncbi.nlm.nih.gov> (дата обращения: 18.01.2012).
64. *Penn State Genome Cartography Project*. URL: <http://atlas.bx.psu.edu/project/drosophila.html> (дата обращения: 21.05.2013).
65. Khesin R.V., Leibovitch B. Influence of deficiency of the histone gene-containing 38B-40 region on X-chromosome template activity and the *white* gene position effect variegation in *Drosophila melanogaster*. *Mol. Gen. Genet.* 1978. V. 162. P. 323–328.
66. *A Database of Drosophila Genes & Genomes*. URL: <http://flybase.org> (дата обращения: 8.02.2012).
67. Wakimoto B.I., Hearn M.G. The effect of chromosome rearrangements on the expression of heterochromatic genes in chromosome 2L of *Drosophila melanogaster*. *Genetics*. 1990. V. 125. P. 141–154.

Материал поступил в редакцию 14.11.2013, опубликован 02.06.2014.