

ВЛИЯНИЕ 3Д-ОРГАНИЗАЦИИ ХРОМАТИНА НА СУПЕРЭНХАНСЕР–ПРОМОТОРНЫЕ ВЗАИМОДЕЙСТВИЯ В ЭМБРИОНАЛЬНЫХ СТВОЛОВЫХ И ЗЛОКАЧЕСТВЕННЫХ КЛЕТКАХ

© 2023 г. Ю. А. Эйдельман^{1, 2, *}, С. Г. Андреев^{1, 2, **}

¹Институт биохимической физики им. Н.М. Эмануэля Российской академии наук, Москва, 119334 Россия

²Национальный исследовательский ядерный университет “МИФИ”, Москва, 115409 Россия

*e-mail: eidel@mail.ru

**e-mail: andreev_sg@mail.ru

Поступила в редакцию 11.10.2022 г.

После доработки 05.12.2022 г.

Принята к публикации 08.12.2022 г.

Взаимодействие энхансеров и суперэнхансеров (СЭ) с промотором функционально значимо для регуляции экспрессии генов. Паттерны этих взаимодействий играют ключевую роль в регуляции таких процессов, как дифференцировка, злокачественная трансформация и др. Чтобы получить количественную информацию о взаимосвязи структуры хроматина и экспрессии генов, в настоящей работе был проведен компьютерный анализ конформаций хроматина вблизи гена плюрипотентности *Nanog* мыши, включая контакты СЭ с промотором, в нормальных эмбриональных стволовых (mESC) и в клетках лимфомы (CH12LX). С помощью биофизических методов моделирования идентифицированы параметры суперэнхансер-промоторных взаимодействий – распределения расстояний между промотором *Nanog* и СЭ, частоты контактов с одним и несколькими СЭ одновременно. В нормальных mESC частота контактов промоторов с СЭ выше, чем в злокачественных, и чаще образуются сложные контакты, с двумя и более СЭ. В злокачественных клетках обнаружена малая субпопуляция клеток, где промотор контактирует одновременно с тремя СЭ. Субпопуляции злокачественных клеток с множественными контактами промотор–СЭ могут предрасполагать к проявлению свойств стволовых клеток и гипотетически быть кандидатом на опухолевые стволовые клетки.

Ключевые слова: *Nanog*, промотор–энхансерные взаимодействия, Hi-C, полимерное моделирование, молекулярная динамика, структура хроматина.

DOI: 10.31857/S0016675823060036, **EDN:** SRUZYX

Суперэнхансерами (СЭ) принято считать кластеры энхансеров, характеризующиеся повышенным уровнем связывания транскрипционных факторов, они играют важную роль в регуляции экспрессии генов [1]. Связывание промоторов не с одним, а с несколькими энхансерами или СЭ приводит к повышению экспрессии соответствующего гена. В [1] показано, что в эмбриональных стволовых клетках повышена частота взаимодействий промотора с СЭ генов плюрипотентности *Oct4*, *Sox2* и *Nanog*. В 2000 гг. были созданы эффективные методы количественного исследования геномных контактов: “захват хромосомных конформаций” (Chromosome Conformation Capture, 3C) [2] и его производные, включая полногеномный метод Hi-C [3]. Методы основаны на лигировании ДНК вблизи контактов хромосомных локусов и последующем определении частот и расположения точек лигирования на геноме. Эти подходы стимулировали развитие структурных исследований,

взаимосвязи экспрессии генов и 3Д-контактов промоторов с энхансерами и СЭ. Однако стандартные методы захвата конформаций не дают прямой информации о 3Д-структуре генома [4, 5] и, в частности, о пространственной организации областей, где расположены промоторы и суперэнхансеры. Кроме того, они предназначены для измерения парных контактов (например, данного промотора с данным энхансером) в популяции клеток, но не могут детектировать более сложные непарные контакты. Недавно появившийся метод “картирование архитектуры генома” (Genome Architecture Mapping, GAM) позволяет детектировать и сложные контакты, но такие работы пока единичны [6]. Для восстановления информации о 3Д-структуре участков генома и контактах локусов по данным Hi-C применяются методы биоинформатики и полимерного моделирования, например [4, 5, 7, 8].

Таблица 1. Положение гена *Nanog* и суперэнхансеров (СЭ) в мышинной хромосоме 6 (геном mm10). Данные по СЭ в mESC из [13]

	Начало, млн пн	Конец, млн пн	Размер СЭ, тыс. пн
<i>Nanog</i>	122.708	122.715	
СЭ1	122.337	122.341	4
СЭ2	122.664	122.666	2
СЭ3	122.684	122.709	25
СЭ4	122.761	122.772	11

Одна из центральных проблем генетики и эпигенетики развития — как гены плюрипотентности вовлечены в регуляцию свойств нормальных клеток и могут ли они участвовать в регуляции свойств клеток со злокачественным фенотипом — остается слабоизученной. В настоящей работе мы провели сравнительный анализ структурной организации хроматина в окрестности гена плюрипотентности *Nanog*, включающей в себя промотор-суперэнхансерные контакты в эмбриональных стволовых (mESC) и в лимфоидных злокачественных (CH12LX) клетках мыши. С помощью разработанного метода полимерного моделирования были получены детальные количественные характеристики промотор—СЭ контактов, распределения расстояний между промотором и СЭ, частоты контактов с одним и несколькими СЭ одновременно. Мы демонстрируем, что в mESC средняя частота контактов промоторов с СЭ выше, чем в злокачественных, и в них намного чаще образуются сложные контакты, с двумя и более СЭ. В злокачественных CH12LX клетках расчеты обнаруживают наличие малой субпопуляции клеток, в которых промотор контактирует одновременно с тремя СЭ. Субпопуляция злокачественных клеток со структурой хроматина, характеризующаяся множественными промотор—СЭ контактами, не детектируемыми классическим методом Hi-C, может означать наличие субпопуляции опухолевых стволовых клеток (ОСК).

МАТЕРИАЛЫ И МЕТОДЫ

Метод полимерного моделирования структуры хромосом и описания карт Hi-C контактов изложен в предыдущей статье [5]. В данной работе фрагмент мышинной хромосомы 6 с геномными координатами 120–123.5 млн пн моделировался полимерной цепью размером 3.5 млн пн с субъединицами, содержащими 10 тыс. пн ДНК, диаметр субъединицы был выбран 88 нм.

Полимерное моделирование проводилось следующим образом. Конформация полимерной цепи формируется за счет потенциалов объемных взаимодействий и исключенного объема. Потен-

циалы объемных взаимодействий гетерогенны, т.е. зависят от конкретной пары локусов (i, j). Их величины определяются с помощью итеративного алгоритма по критерию наилучшего согласия расчетной и экспериментальной карт контактов. На каждой итерации для данного набора потенциалов моделируется ансамбль конформаций, определяется расчетная карта контактов, потенциал для каждой пары (i, j) меняется в зависимости от того, расчетная частота контактов больше или меньше экспериментальной.

Расчетная карта контактов вычисляется для каждой конформации: пара локусов (i, j) находится в контакте, если расстояние между ними меньше заданного, $r_{ij} < R_{\text{cont}}$; затем суммированием по всем конформациям получается средняя по ансамблю карта. Из готовых конформаций цепи 3.5 млн пн вычлениаются субцепи с ~1 млн пн окрестностью гена *Nanog*, и дальнейшие действия производятся уже на ансамбле этих субцепей.

Одной из характеристик ансамбля конформаций является функция $\Psi_{i,j}(r)$ — распределение расстояний между интересующими нас парами локусов (i — промотор, j — один из СЭ). В случае самого крупного из суперэнхансеров, СЭ3, занимающего три субъединицы (табл. 1), распределение считалось между промотором и геометрическим центром этих трех субъединиц.

Данные по экспрессии гена *Nanog* в клетках mESC и CH12LX взяты из ENCODE [9] по адресам ENCNR000CGU и ENCNR000AJV соответственно. Обработка данных по экспрессии осуществлялась с помощью HTSeq [10], и определялся уровень экспрессии в единицах FPKM (Fragments Per Kilobase Million). FPKM в клетках mESC равен 0.6, в клетках CH12LX равен нулю.

Hi-C карты рассматриваемого сегмента мышинной хромосомы 6 в разрешении 10 и 40 тыс. пн строились по базам данных из Gene Expression Omnibus: GSE35156 [11] для mESC, GSE63525 [12] для CH12LX клеток.

Исследование ансамбля конформаций на уровне 1 млн пн окрестности *Nanog* осуществлялось путем иерархического кластерного анализа [7]. Для каждой пары конформаций (p, q) вычислялось “среднеквадратичное расстояние” (root mean square distance, dRMSD) по формуле

$$\text{dRMSD}(p, q) = \sqrt{\frac{\sum_{i,j} (d_{ij}^p - d_{ij}^q)^2}{N(N-1)/2}},$$

где d_{ij}^p и d_{ij}^q — расстояния между i -м и j -м элементами в конформациях p и q соответственно [7], N — число элементов цепи. Этот показатель является количественным критерием степени сходства данных двух конформаций из ансамбля. Кластеризация осуществлялась методом невзвешенного

группового среднего. Для различного числа кластеров определялось среднее расстояние от всех конформаций до центра кластеров, к которым они принадлежат, и оптимальное число кластеров определялось как такое число, после которого среднее расстояние переставало существенно уменьшаться [7]. Для клеток mESC оно оказалось равным 7, для клеток CH12LX — 6. В дальнейшем мы рассматривали только кластеры, содержащие более 1% конформаций; таких кластеров по три для каждой линии клеток.

РЕЗУЛЬТАТЫ

Для выяснения роли пространственной структуры хромосомного локуса в окрестности гена *Nanog* в суперэнхансер-промоторных взаимодействиях нами проведено количественное описание Hi-C частот контактов и получены ансамбли конформаций 3.5-млн пн участка мышинной хромосомы 6, включающего *Nanog*, для клеток mESC и CH12LX, по 4000 конформаций для каждой линии (рис. 1). Ансамбль представляет собой набор 3Д-конформаций хроматина, которые дают информацию о флуктуациях положения локусов в единичных клетках (хромосомах), о распределении расстояний между локусами, флуктуациях числа контактов и др. При этом средние по ансамблю карты контактов описываются полимерной моделью с высокими коэффициентами корреляции. На рис. 1 представлено сравнение экспериментальных карт контактов [11, 12] с расчетными, полученными в настоящей работе. Приведены примеры 3Д-структур хроматина в окрестности гена *Nanog* для нормальных и злокачественных клеток, восстановленных на основе полимерной модели 3.5-млн пн участка хромосомы 6 клеток мыши с разрешением 10 тыс. пн. Укрупнение расчетных и экспериментальных карт контактов на рис. 1, а, д выбрано 40 тыс. пн, как в экспериментальной работе [11]. Чтобы получить представление о конформациях хроматина с более высоким разрешением, построены экспериментальные и расчетные карты с разрешением 10 тыс. пн, соответствующим размеру субъединицы полимерной цепи, в непосредственной (~1 млн пн) окрестности *Nanog* (рис. 1, в, ж). Показаны несколько конформаций участка хромосомы, на котором расположен ген *Nanog* и четыре ближайших к нему СЭ. Данные по положению СЭ в геноме взяты из [13]. Координаты гена и суперэнхансеров на хромосоме 6 приведены в табл. 1. Размеры СЭ варьируются от двух до 25 тыс. пн, т.е. они располагаются на одной-трех субъединицах полимерной цепи. Помимо четырех СЭ, на рассматриваемом участке хромосомы имеются также четыре отдельных энхансера [14], но для них расчеты контактов и расстояний не проводились.

Таблица 2. Средние расстояния между промотором гена *Nanog* и четырьмя суперэнхансерами в клетках mESC и CH12LX

Суперэнхансер	Среднее расстояние, нм		<i>p</i>
	mESC	CH12LX	
СЭ1	338	363	3e-18*
СЭ2	160	170	1e-18*
СЭ3	97	101	1e-29*
СЭ4	176	196	2e-51*

Примечание. В правой колонке — оценка значимости различий по *t*-тесту; звездочками помечены различия, значимые с $p < 0.01$.

Чтобы определить, как близость СЭ и промотора *Nanog* зависит от типа клеток, на основе полученных ансамблей конформаций для мышинных эмбриональных стволовых mESC и злокачественных CH12LX клеток были построены распределения расстояний между промотором и всеми четырьмя СЭ (рис. 2). Средние расстояния и статистическая значимость различий между двумя типами клеток приведены в табл. 2.

По данным на рис. 2 и в табл. 2 средние расстояния между промотором и всеми четырьмя СЭ различаются в двух типах клеток в пределах 10% (различия статистически значимые, по *t*-тесту $p < 1e-19$ для всех четырех СЭ). При этом на близких расстояниях, меньше 100 нм, промотор и три из четырех СЭ колокализуются чаще в mESC, чем в CH12LX (врезки на рис. 2).

Знание ансамбля 3Д-конформаций, соответствующего популяционной Hi-C карте контактов, позволило получить частоты парных контактов промотора с конкретным СЭ, а также частоты образования контактов промотора с двумя, тремя или четырьмя СЭ одновременно. В табл. 3 приведены частоты конформаций в клетках mESC и CH12LX со всеми комбинациями контактов промотора с каждым конкретным СЭ или набором СЭ. Такая информация представляет собой новые данные, не извлекаемые из Hi-C экспериментов. На рис. 3 показаны распределения по числу СЭ, с которыми одновременно контактирует промотор. Распределения для нормальных mESC и злокачественных CH12LX клеток отличаются статистически значимо (критерий Колмогорова—Смирнова, $p = 1e-58$).

Из анализа рис. 3 и табл. 3 можно вывести следующие характерные особенности организации контактов промотор—СЭ в эмбриональных стволовых и злокачественных клетках. Частота сложных контактов промотора одновременно более чем с одним СЭ чувствительна к статусу клетки. Доля промоторов, контактирующих одновременно более чем с одним СЭ, составляет ~40% в

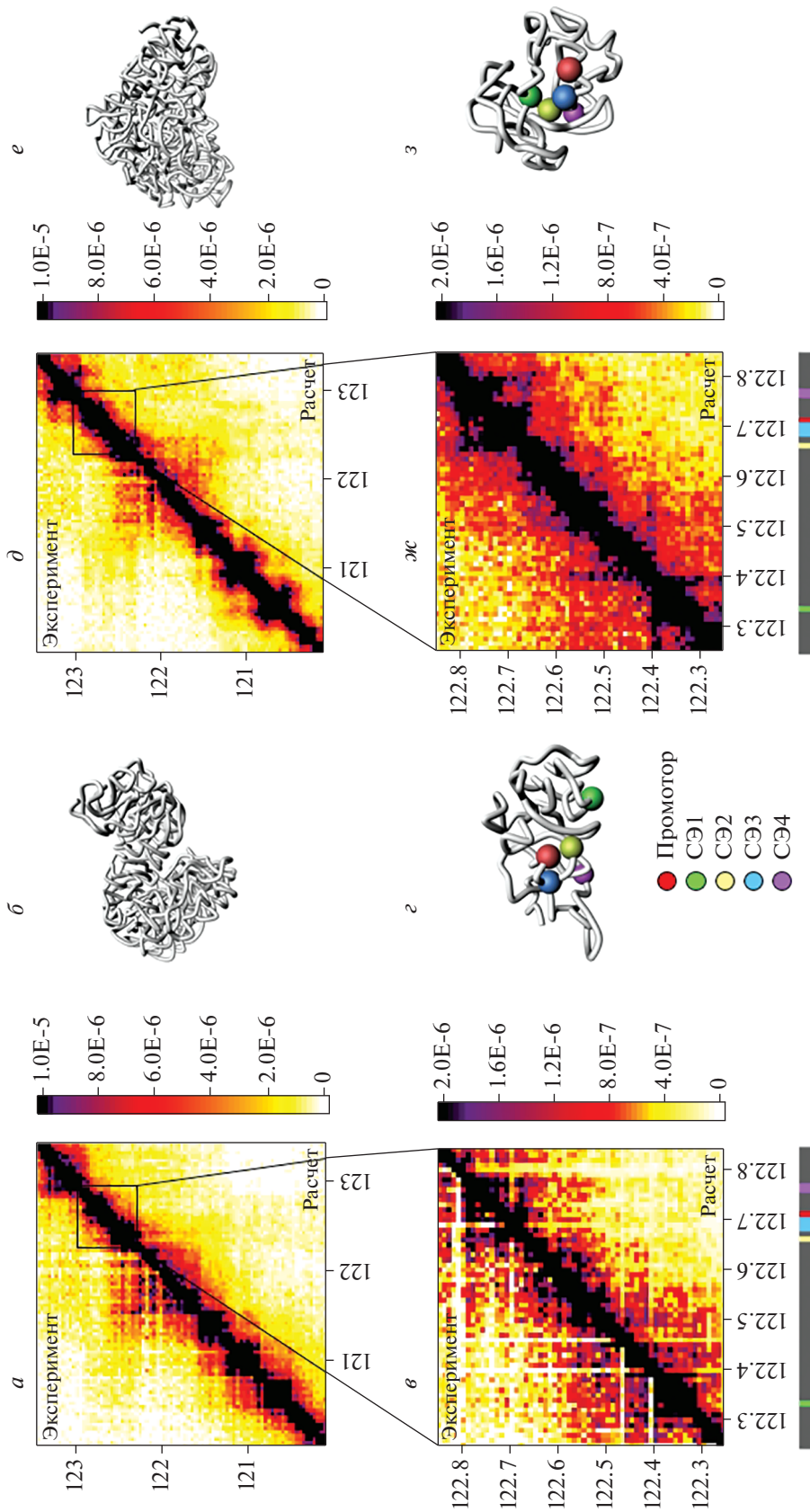


Рис. 1. Экспериментальные и расчетные карты контактов геномных локусов и 3Д-конформации 3.5-млн пн участка хроматина, содержащего ген *Nanog*, и ближайших к нему в пределах ~0.4 млн пн в обе стороны суперэнхансеров для мышиной хромосомы 6 в эмбриональных стволовых и злокачественных клетках. *а-г* — нормальные эмбриональные стволовые клетки mESC; *д-ж* — злокачественные клетки CH12LX. В-клеточная лимфома; *а, б* — карты контактов участка 120–123.5 млн пн для клеток mESC и CH12LX, расчет в сравнении с экспериментами [11] и [12] соответственно. Как расчетная, так и экспериментальная карта с разрешением 10 тыс. пн укрупнена до 40 тыс. пн. Корреляция Пирсона $R = 0.919$ и 0.955 соответственно; *б, е* — примеры 3Д-конформаций из ансамбля структур; *в, ж* — карты контактов ближайшей окрестности гена *Nanog*, где расположены суперэнхансеры, разрешение 10 тыс. пн. Экспериментальные данные [11] и [12]. Под картами приведено положение промотора *Nanog* и суперэнхансеров 1–4 (СЭ1–СЭ4); *г, з* — примеры конформаций. Положения субъединиц, содержащих промотор и СЭ, изображено шарами, размер увеличен для наглядности. Для mESC в среднем характерно больше контактов промотора с СЭ, в том числе множественные контакты в индивидуальных конформациях, чем для CH12LX.

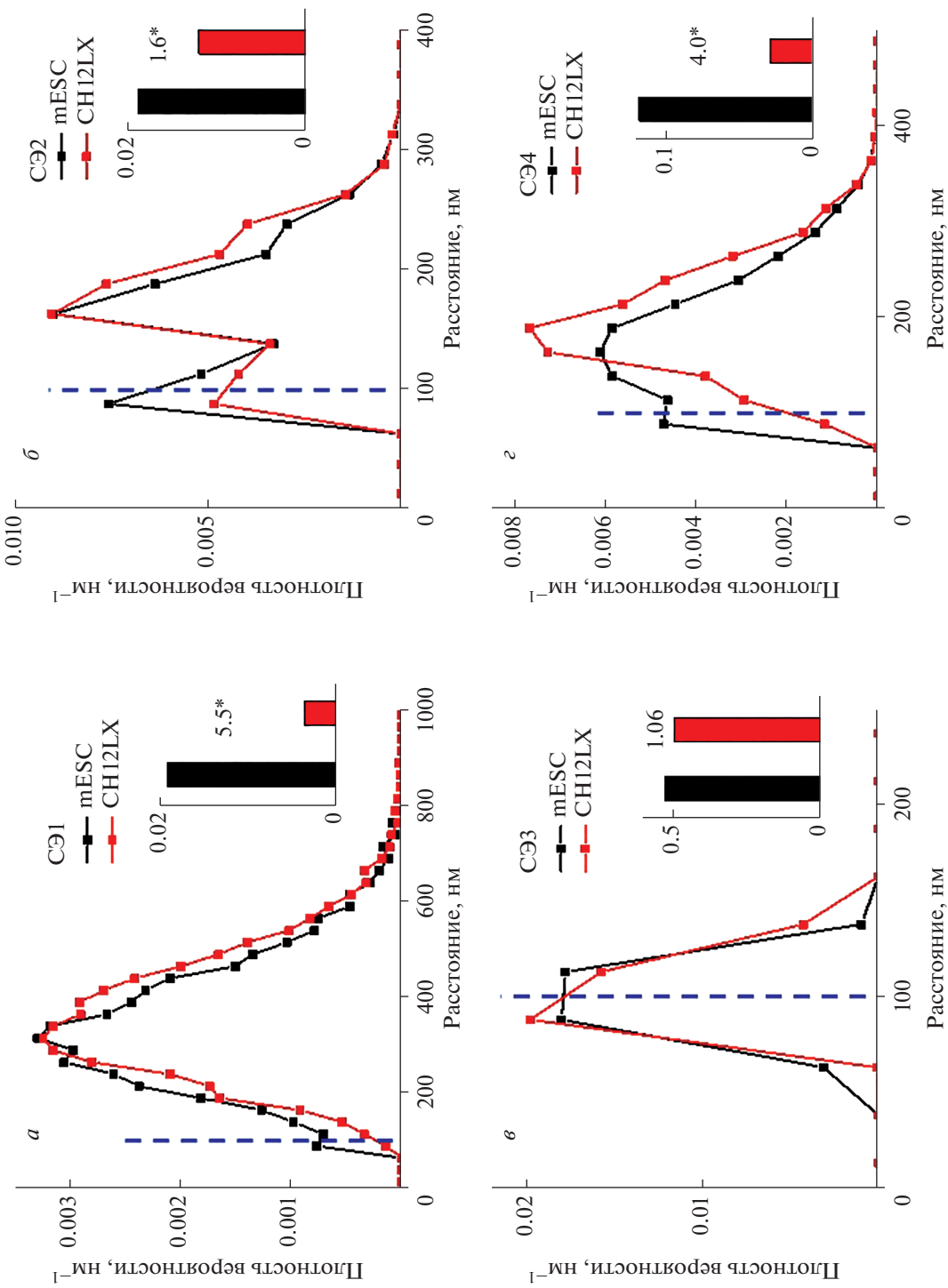


Рис. 2. Ген *Nanog* колокализован с суперэнхансерами в mESC чаще, чем в CH12LX. *а* – распределение расстояний между промотором и СЭ1; *б* – между промотором и СЭ2; *в* – между промотором и СЭ3; *г* – между промотором и СЭ4. Во врезках на всех панелях показана частота конформаций в двух типах клеток, где промотор находится в контакте с СЭ или на расстоянии меньше 100 нм (область левее синего пунктира на графиках). Числа во врезках – соотношение частот в mESC и в CH12LX; звездочками помечены статистически значимые различия по *t*-тесту ($p < 1e-4$). Распределения построены по ансамблям из 4000 конформаций.

Таблица 3. Рассчитанные частоты конформаций с контактами между промотором *Nanog* и суперэнхансерами, одним, двумя и тремя, в ансамблях из 4000 конформаций для нормальных (mESC) и злокачественных (CH12LX) клеток

	mESC	CH12LX	<i>p</i>
Нет контактов	0.072 ± 0.004	0.182 ± 0.006	<1e-5*
СЭ1	0.0035 ± 9e-4	0.0018 ± 7e-4	0.13
СЭ2	0.022 ± 0.002	0.029 ± 0.003	0.06
СЭ3	0.474 ± 0.008	0.547 ± 0.008	<1e-5*
СЭ4	0.015 ± 0.002	0.012 ± 0.002	0.20
СЭ1 + СЭ2	2.5e-4 ± 2.5e-4	2.5e-4 ± 2.5e-4	1
СЭ1 + СЭ3	0.016 ± 0.002	0.0035 ± 9e-4	<1e-5*
СЭ1 + СЭ4	5e-4 ± 4e-4	7.5e-4 ± 4e-4	0.65
СЭ2 + СЭ3	0.229 ± 0.007	0.167 ± 0.006	<1e-5*
СЭ2 + СЭ4	0.005 ± 0.001	0.0015 ± 6e-4	0.009*
СЭ3 + СЭ4	0.111 ± 0.005	0.042 ± 0.003	<1e-5*
СЭ1 + СЭ2 + СЭ3	0.006 ± 0.001	0.0025 ± 8e-4	0.016
СЭ1 + СЭ2 + СЭ4	2.5e-4 ± 2.5e-4	2.5e-4 ± 2.5e-4	1
СЭ1 + СЭ3 + СЭ4	0.0018 ± 7e-4	н/д	0.008*
СЭ2 + СЭ3 + СЭ4	0.044 ± 0.003	0.012 ± 0.002	<1e-5*
СЭ1 + СЭ2 + СЭ3 + СЭ4	7.5e-4 ± 4e-4	н/д	0.08
Всего			
0 СЭ	0.072 ± 0.004	0.182 ± 0.006	<1e-5*
1 СЭ	0.515 ± 0.008	0.589 ± 0.008	<1e-5*
2 СЭ	0.361 ± 0.008	0.215 ± 0.006	<1e-5*
3 СЭ	0.052 ± 0.003	0.015 ± 0.002	<1e-5*
4 СЭ	7.5e-4 ± 4e-4	н/д	0.08

Примечание. Приведены средние значения и стандартные отклонения. н/д — не детектировано. В правой колонке — оценка значимости различий по двухвыборочному χ -тесту; звездочками помечены различия, значимые с $p < 0.01$.

mESC, где *Nanog* экспрессируется, и ~20% (рис. 3) в CH12LX, где *Nanog* не экспрессируется. В хромосомах злокачественных клеток обнаруживается малая субпопуляция, 1.5% (табл. 3), в которой конформация хроматина такова, что промотор контактирует с тремя СЭ. Для сравнения, в mESC частота этой субпопуляции 5.2% (табл. 3 и столбец “3 СЭ” на рис. 3).

Если сравнивать две клеточные линии по специфическим множественным контактам промотора с конкретными СЭ, то в первую очередь эмбриональные стволовые клетки характеризуются большей, по сравнению со злокачественными клетками, частотой контактов с СЭ1 + СЭ3 (доля субпопуляции хромосом с таким контактом 1.58% vs 0.35%, 4.5-кратная разница), СЭ3 + СЭ4 (11.1% vs 4.2%, 2.6-кратная разница), СЭ2 + СЭ3 + СЭ4 (4.4% vs 1.2%, 3.6-кратная разница), см. табл. 3. В злокачественных клетках больше в 1.5 раза субпопуляция с двойным контактом с СЭ1 + СЭ4, хотя частоты этой субпопуляции малы в обоих типах клеток (0.075% vs 0.05%) и разница статистически не значима.

Различия в частотах контактов на средних картах (рис. 1) предполагают отличие статистических ансамблей конформаций рассматриваемых участков хромосом. Чтобы количественно охарактеризовать гетерогенность структуры хроматина в окрестности гена *Nanog* в популяции хромосом (клеток), был проведен иерархический кластерный анализ ансамблей конформаций в области ~1 млн пн в нормальных и злокачественных клетках (рис. 4).

Кластерный анализ показывает, что по клеточная вариабельность конформаций хроматина в окрестности *Nanog* проявляется в существовании трех основных классов конформаций, различающихся в нормальных и злокачественных клетках структурными свойствами — степенью конденсированности и вовлечением локусов хроматина в дальние—ближние контакты. Доминирующий кластер в обоих типах клеток (72.5% всех конформаций в mESC, 85.9% в CH12LX) содержит более компактные конформации, чем остальные два кластера, представленные структурами в большей степени деконденсированными (рис. 4, а, в). Мы

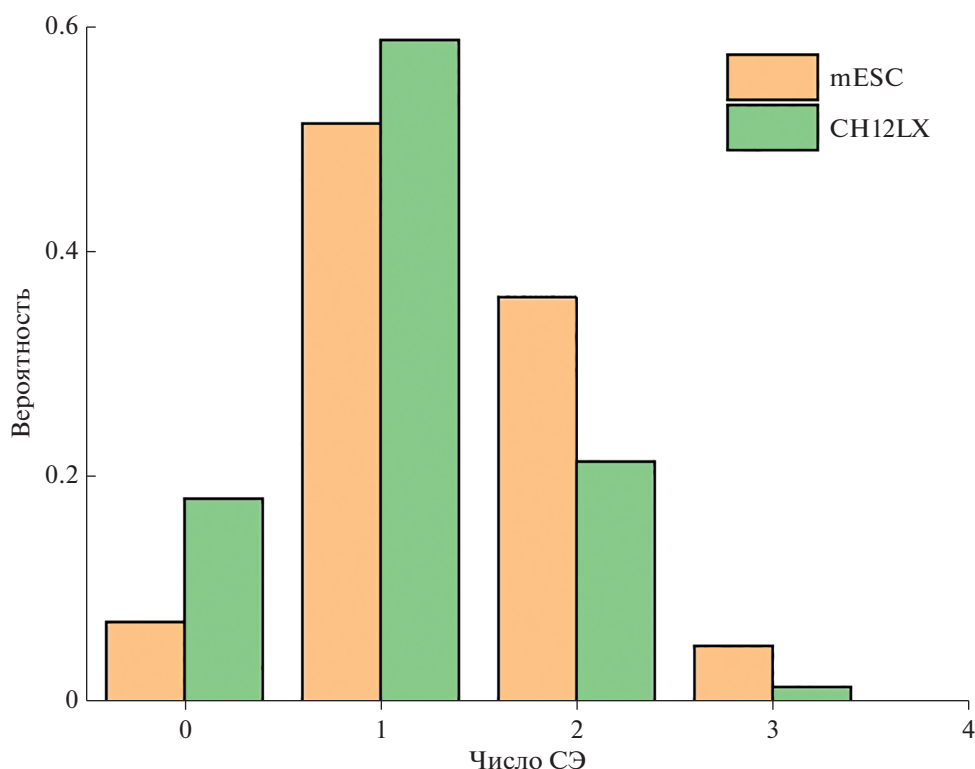


Рис. 3. Вероятность образования простых (1) и множественных (≥ 2) контактов промотора *Nanog* с СЭ в эмбриональных стволовых и злокачественных клетках.

также определили, как по этим кластерам распределены конформации с множественными (>2) контактами промотор–СЭ (рис. 4, б, г). В обоих типах клеток большинство таких конформаций принадлежит к компактным кластерам.

ОБСУЖДЕНИЕ

Суперэнхансеры как элементы регуляции экспрессии генов привлекают внимание с 2013 г., когда была введена сама концепция СЭ [1]. Будучи кластерами энхансеров они обладают повышенной способностью связывать транскрипционные факторы (Oct4, Sox2, Nanog, Klf4, Esrrb) и комплекс Mediator и тем отличаются от обычных энхансеров [14]. Критерием принадлежности локусов к обычным энхансерам или к СЭ в [14] служил уровень сигнала комплекса Mediator, конкретно Med1, хотя в этой работе конкретный порог не был указан.

Мы обнаружили различия в частотах разных комбинаций промотор–СЭ взаимодействий между mESC и злокачественными клетками в ~ 1 млн пн окрестности плюрипотентного гена *Nanog*. Частота взаимодействий промотор–СЭ определяется в модели распределением расстояний между соответствующими локусами в популяции клеток или

статистическим ансамблем конформаций хроматина. Кластерный анализ конформаций в окрестности гена *Nanog* показал больше компактных структур хроматина в злокачественных клетках по сравнению с mESC. С другой стороны, прямой корреляции между множественными контактами промотор–СЭ и компактностью структуры не обнаруживается. Несмотря на более высокую частоту множественных (≥ 3) контактов промотора *Nanog* с СЭ в mESC по сравнению со злокачественными клетками, $0.053 : 0.015$ (табл. 3), в этой окрестности гена *Nanog* доли компактных конформаций в mESC и в CH12LX соотносятся как $72.5 : 85.9$ (рис. 4).

Вероятностные характеристики взаимодействий промотор–СЭ были получены из имеющихся популяционных Hi-C данных путем генерации статистического ансамбля конформаций хроматина, на основе которого были рассчитаны поклеточные (похромосомные) характеристики, распределения расстояний и числа контактов между промоторами и СЭ. Полученные данные демонстрируют сложные взаимоотношения между экспрессией гена, частотой СЭ-промоторных контактов и компактизацией хроматина в обла-

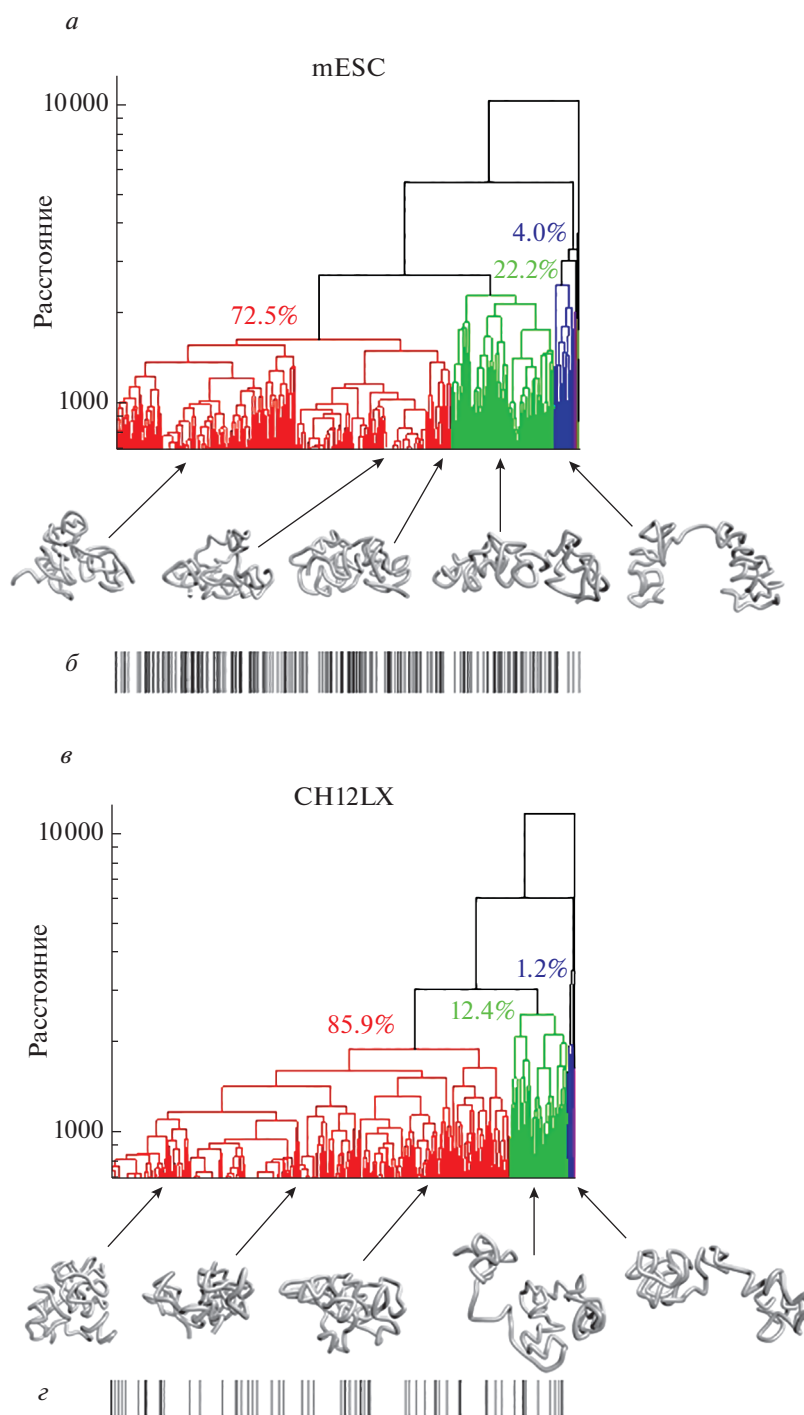


Рис. 4. Анализ поклеточной вариабельности 3Д-структуры хроматина в окрестности гена *Nanog* в клетках mESC и CH12LX. *a, б* – клетки mESC; *в, г* – клетки CH12LX; *a, в* – дендрограммы иерархической кластеризации конформаций хроматина. Найдено оптимальное число кластеров – 7 для mESC и 6 для CH12LX. Мелкие кластеры, включающие в себя менее 1% конформаций, далее не учитывали. На рисунке изображены по три основных кластера. Указан процент конформаций в них и показаны характерные конформации из разных кластеров и субкластеров; *б, г* – расположение на дендрограмме конформаций с множественными (>2) контактами промотора *Nanog* с СЭ.

стях хромосом, содержащих ген и далеко отстоящие энхансеры и СЭ.

Сравнивая с другими модельными подходами к получению информации о промотор-энхансерных и промотор-суперэнхансерных взаимодействиях, можно отметить ряд моментов, по которым наша модель отличается от других. В работе [15] представлен компьютерный подход для анализа Hi-C данных и идентификации энхансер-промоторных взаимодействий. Использовалась обратная, по сравнению с нашей, идеология: искали локусы, с повышенной частотой взаимодействующие с промоторами, и предсказывалось, что в этих локусах находятся энхансеры, которые затем проверялись на известные эпигенетические метки энхансеров, такие как ДНКазы I-гиперчувствительные сайты. Такой подход позволяет определять только энхансеры, с которыми контактируют промоторы, но не те, с которыми они не контактируют. В этом принципиальное отличие от настоящей работы, где мы находим субпопуляции хромосом с промоторами как контактирующими, так и не контактирующими с каждым из СЭ.

Взаимодействие промотора с энхансерами изучалось путем моделирования структуры участков хромосом в разных типах клеток в [16]. В фокусе внимания был ген *Pax6*, участвующий в развитии нервной ткани, и два энхансера. Авторы показали взаимосвязь между экспрессией *Pax6* (транскрипционного фактора, участвующего в развитии нервной ткани и глаз) и контактами его промотора с одним из двух энхансеров и отсутствие такой взаимосвязи с другим. Отличие подхода [16] от нашего в том, что там не рассматривались множественные контакты, более чем с одним энхансером одновременно. Структурная организация хромосом на уровне топологически ассоциированных доменов (ТАДов) в клетках mESC моделировалась в [7]. Авторы основывались на 5С-данных, теоретическое описание более полных Hi-C карт контактов, которое доказывало бы справедливость модели, отсутствует. Преимущество нашей модели реконструкции 3Д-структуры хроматина в том, что она основана на количественном описании Hi-C карт контактов с высоким коэффициентом корреляции.

В настоящей работе в обоих типах клеток обнаружена субпопуляция хромосом, в которой промотор *Nanog* взаимодействует более чем с двумя СЭ одновременно. Предсказание субпопуляции злокачественных клеток с множественными контактами промотор-СЭ ставит вопрос, не являются ли они ОСК. *Nanog* активнее экспрессируется в ряде злокачественных опухолей человека, чем в нестволовых клетках [17]. Рассчитанные

нами по данным [9] величины ФРКМ в стволовых и злокачественных клетках (0.6 и 0) предполагают отсутствие экспрессии *Nanog* в клетках CH12LX. Однако информация об уровне экспрессии [9] была получена на большой популяции без клеточного анализа и детектирования субпопуляций со стволовыми маркерами. Поэтому теоретически не исключена возможность того, что линия клеток мышинной лимфомы может содержать субпопуляции ОСК с активным геном *Nanog*. Другая возможная интерпретация субпопуляции с множественными контактами промотор-СЭ заключается в том, что предсуществующие контакты являются необходимым, но не достаточным условием экспрессии гена. Пространственная близость энхансер-промотор автоматически не индуцирует экспрессию гена. Контакты энхансер-промотор могут запускать некоторые гены для активации, но для запуска транскрипции требуются дополнительные триггеры или факторы [18]. Существует также возможность бесконтактного механизма действия энхансеров, когда транскрипция гена не требует физического контакта промотора с энхансером, или образования промотор-энхансерных петель [18]. Эти аргументы свидетельствуют в пользу того, что отсутствие прямой корреляции частоты контактов промотор-СЭ с уровнем экспрессии гена, определяемым в среднем по большой популяции клеток, отражает скорее вероятностные взаимоотношения между стохастической экспрессией генов и стохастической организацией хроматина и генома [18–20]. Представленная модель СЭ-промоторных взаимодействий допускает обобщение на случай неконтактного механизма. Для этого условие экспрессии гена через нахождение промотор-СЭ на расстоянии, не превышающем расстояние контакта, достаточно заменить на вероятность экспрессии гена, спадающую с увеличением расстояния между СЭ и промотором. Распределение расстояний между СЭ и промотором определяется 3Д-организацией хроматина и находится предложенным в работе методом.

Говоря о пределах применимости полученных результатов, следует отметить, что мы использовали данные по СЭ для mESC и предполагали, что в злокачественных мышинных клетках СЭ расположены в тех же локусах, что и в нормальных эмбриональных клетках.

Описание геномных данных Hi-C и полимерная реконструкция ансамбля конформаций хроматина в области гена *Nanog* в клетках разного статуса, нормальных и лимфомных, позволяют получить не наблюдаемую в популяционном анализе хроматиновых контактов информацию о частотах субпопуляций клеток с множественными

контактами СЭ—промотор. Продемонстрированы возможности полимерного моделирования применительно к анализу эпигенетической регуляции, СЭ—промоторных взаимодействий, на основе геномных Hi-C данных.

Расчеты были проведены с использованием вычислительных мощностей Межведомственного суперкомпьютерного центра РАН.

Настоящая статья не содержит каких-либо исследований с использованием в качестве объекта животных.

Настоящая статья не содержит каких-либо исследований с участием в качестве объекта людей.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Hnisz D., Abraham B.J., Lee T.I. et al. Super-enhancers in the control of cell identity and disease // *Cell*. 2013. V. 155. № 4. P. 934–947. <https://doi.org/10.1016/j.cell.2013.09.053>
2. Dekker J., Rippe K., Dekker M., Kleckner N. Capturing chromosome conformation // *Science*. 2002. V. 295. № 5558. P. 1306–1311. <https://doi.org/10.1126/science.1067799>
3. Lieberman-Aiden E., van Berkum N.L., Williams L. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome // *Science*. 2009. V. 326. № 5950. P. 289–293. <https://doi.org/10.1126/science.1181369>
4. Zhang B., Wolynes P.G. Topology, structures, and energy landscapes of human chromosomes // *Proc. Natl Acad. Sci. USA*. 2015. V. 112. № 19. P. 6062–6067. <https://doi.org/10.1073/pnas.1506257112>
5. Eidelman Y., Salnikov I., Slanina S., Andreev S. Chromosome folding promotes intrachromosomal aberrations under radiation- and nuclease-induced DNA breakage // *Int. J. Mol. Sci*. 2021. V. 22. № 22. P. 12186. <https://doi.org/10.3390/ijms222212186>
6. Beagrie R.A., Scialdone A., Schueler M. et al. Complex multi-enhancer contacts captured by genome architecture mapping // *Nature*. 2017. V. 543. № 7646. P. 519–524. <https://doi.org/10.1038/nature21411>
7. Giorgetti L., Galupa R., Nora E.P. et al. Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription // *Cell*. 2014. V. 157. № 4. P. 950–963. <https://doi.org/10.1016/j.cell.2014.03.025>
8. Annunziatella C., Chiariello A.M., Bianco S., Nicodemi M. Polymer models of the hierarchical folding of the Hox-B chromosomal locus // *Phys. Rev. E*. 2016. V. 94. № 4-1. P. 042402. <https://doi.org/10.1103/PhysRevE.94.042402>
9. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome // *Nature*. 2012. V. 489. № 7414. P. 57–74. <https://doi.org/10.1038/nature11247>
10. Anders S., Pyl P.T., Huber W. HTSeq – a Python framework to work with high-throughput sequencing data // *Bioinformatics*. 2015. V. 31. № 2. P. 166–169. <https://doi.org/10.1093/bioinformatics/btu638>
11. Dixon J.R., Selvaraj S., Yue F. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions // *Nature*. 2012. V. 485. № 7398. P. 376–380. <https://doi.org/10.1038/nature11082>
12. Rao S.S.P., Huntley M.H., Durand N.C. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping // *Cell*. 2014. V. 159. № 7. P. 1665–1680. <https://doi.org/10.1016/j.cell.2014.11.021>
13. Novo C.L., Javierre B.-M., Cairns J. et al. Long-range enhancer interactions are prevalent in mouse embryonic stem cells and are reorganized upon pluripotent state transition // *Cell Rep*. 2018. V. 22. № 10. P. 2615–2627. <https://doi.org/10.1016/j.celrep.2018.02.040>
14. Whyte W.A., Orlando D.A., Hnisz D. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes // *Cell*. 2013. V. 153. № 2. P. 307–319. <https://doi.org/10.1016/j.cell.2013.03.035>
15. Ron G., Globerson Y., Moran D., Kaplan T. Promoter-enhancer interactions identified from Hi-C data using probabilistic models and hierarchical topological domains // *Nat. Commun*. 2017. V. 8. № 1. P. 2237. <https://doi.org/10.1038/s41467-017-02386-3>
16. Liu L., Kim M.H., Hyeon C. Heterogeneous loop model to infer 3D chromosome structures from Hi-C // *Biophys. J*. 2019. V. 117. № 3. P. 613–625. <https://doi.org/10.1016/j.bpj.2019.06.032>
17. Wang M.-L., Chiou S.-H., Wu C.-W. Targeting cancer stem cells: emerging role of Nanog transcription factor // *Onco. Targets Ther*. 2013. V. 6. P. 1207–1220. <https://doi.org/10.2147/OTT.S38114>
18. Schoenfelder S., Fraser P. Long-range enhancer-promoter contacts in gene expression control // *Nat. Rev. Genet*. 2019. V. 20. № 8. P. 437–455. <https://doi.org/10.1038/s41576-019-0128-0>
19. Finn E.H., Misteli T. Molecular basis and biological function of variability in spatial genome organization // *Science*. 2019. V. 365. № 6457. P. eaaw9498. <https://doi.org/10.1126/science.aaw9498>
20. Sood V., Misteli T. The stochastic nature of genome organization and function // *Curr. Opin. Genet. Dev*. 2022. V. 72. P. 45–52. <https://doi.org/10.1016/j.gde.2021.10.004>

Impact of Chromatin 3D-Organization on Promoter–Superenhancer Interactions in Embryonic Stem vs Cancer Cells

Yu. A. Eidelman^{a, b, *} and S. G. Andreev^{a, b, **}

^a*Emanuel Institute of Biochemical Physics, Russian Academy of Sciences, Moscow, 119334 Russia*

^b*National Research Nuclear University “MEPHI”, Moscow, 115409 Russia*

^{*}*e-mail: eidel@mail.ru*

^{**}*e-mail: andreev_sg@mail.ru*

The interaction of enhancers and superenhancers (SE) with promoters is functionally significant for the regulation of gene expression. Pattern of these interactions plays a key role in various processes, such as differentiation, malignant transformation, etc. In order to quantify the relationship between 3D chromatin organization and promoter–SE contacts, a computational analysis of chromatin conformations near the murine *Nanog* pluripotency gene was performed for normal embryonic stem (mESC) and lymphoma (CH12LX) cells. Using biophysical modeling approach, the following parameters of the promoter–SE interactions were identified: the distribution of distances between the *Nanog* promoter and the SEs, the frequency of contacts with one and several SEs simultaneously. In normal mESC expressing *Nanog*, the frequency of contacts of promoters with SEs is higher than in cancer cells, and complex contacts with two or more SEs are more frequent. The modelling reveals a small subpopulation of cancer cells, where the promoter contacts simultaneously three SEs. The predicted subpopulation of cancer cells with multiple promoter–SE contacts may be predisposed to increased stemness and hypothetically be considered as a reservoir for generation of cancer stem cells.

Keywords: *Nanog*, promoter–enhancer interactions, Hi-C, polymer modeling, molecular dynamics, chromatin structure.