

Федеральное государственное бюджетное образовательное учреждение
высшего образования «Московский государственный университет имени
М.В.Ломоносова»

На правах рукописи

Жегалова Ирина Владимировна

Анализ трехмерной структуры хроматина эукариот

Специальность 1.5.8. —
«Математическая биология, биоинформатика»

Диссертация на соискание учёной степени
кандидата биологических наук

Научные руководители:
доктор биологических наук, профессор
Гельфанд Михаил Сергеевич
кандидат биологических наук
Храмеева Екатерина Евгеньевна

Москва — 2024

Оглавление

	Стр.
Список рисунков	5
Список таблиц	8
Список сокращений и условных обозначений	9
Введение	11
Глава 1. Обзор литературы	17
1.1 Биология <i>Dictyostelium discoideum</i>	17
1.2 Губки	23
1.3 Ферритины	25
1.4 Анеуплоидии	26
1.5 Организация хроматина	28
1.6 Данные Hi-C и методы работы с ними	34
1.6.1 Концепция бина матрицы и определение его размера	34
1.6.2 Методы нормализации данных Hi-C	37
1.6.3 Методы оценки воспроизводимости и сходства данных Hi-C	42
1.7 Алгоритмы поиска петель	44
1.7.1 Выбор распределения для поиска значимых контактов	49
1.8 Визуализация матриц контактов	51
1.9 Зависимость частоты контактов от расстояния	53
1.10 Инсуляция	53
Глава 2. Методы исследования	55
2.1 Протокол Hi-C	55
2.2 Обработка результатов секвенирования библиотек Hi-C	55
2.3 Заполнение пропущенных значений для улучшения работы алгоритмов поиска петель	56
2.4 Расчет показателей инсуляции	56
2.5 Анализ межхромосомных взаимодействий	57
2.6 Подсчет корреляций между картами Hi-C	58

2.7	Сравнительный анализ упаковки хроматина между разными стадиями развития	59
2.8	Аннотация дальних контактов	59
2.9	Расчет силы дальних взаимодействий	60
2.10	Анализ вклада длинных некодирующих РНК в формирование дальних контактов	60
2.11	Анализ экспрессии в основаниях дальних контактов	60
2.12	Анализ хроматиновых компартментов	61
2.13	Аннотация точечных взаимодействий	61
2.14	Аннотация хроматиновых фонтанов	62
2.15	Гены домашнего хозяйства	63
2.16	Анализ данных АТАС-seq	63
2.17	Аннотация вытянутых петель (e1-петель)	65
2.18	Анализ данных экспрессии	66
2.19	Поиск повторов в геноме <i>D. discoideum</i>	67
2.20	Показатель конвергентности	67
2.21	Аннотация энхансер-подобных элементов	68
2.22	Сравнительный анализ ферритинов представителей типа Porifera	69
2.23	Геномные регионы с измененной локальной упаковкой хроматина	69
2.24	Белок-кодирующие гены человека в регионах с измененной упаковкой при трисомиях	69
2.25	Расположение генов домашнего хозяйства в регионах с измененной упаковкой при трисомиях	70
Глава 3.	Результаты	71
3.1	Организация хроматина <i>Dictyostelium discoideum</i>	71
3.1.1	Выравнивание и контроль качества	71
3.1.2	Анализ межхромосомных и внутрихромосомных взаимодействий <i>D. discoideum</i>	74
3.1.3	Изучение особенностей упаковки хроматина на 0 часе развития <i>D. discoideum</i>	82
3.1.4	Дальние контакты <i>D. discoideum</i>	84

	Стр.
3.1.5 Численная оценка компартментов у <i>D. discoideum</i>	88
3.1.6 Анализ точечных цис-взаимодействий <i>D. discoideum</i>	91
3.1.7 Анализ синтенных блоков между <i>D. discoideum</i> и <i>D. purpureum</i>	96
3.1.8 Сравнительный анализ оснований последовательных и одиночных петель	97
3.1.9 Анализ асимметричных вытянутых петель	98
3.1.10 Взаимосвязь уровня транскрипции и силы петель	101
3.1.11 Анализ регуляторных последовательностей <i>D. discoideum</i>	109
3.1.12 Ключевые особенности хроматина <i>D. discoideum</i>	110
3.2 Организация хроматина <i>Halisarca dujardini</i>	112
3.2.1 Контроль качества карт Hi-C для стадий реагрегации <i>H. dujardini</i>	112
3.2.2 Анализ межхромосомных и внутрехромосомных взаимодействий <i>H. dujardini</i>	113
3.2.3 Анализ компартментов у <i>H. dujardini</i>	115
3.2.4 Анализ хроматиновых шпилек/фонтанов <i>H. dujardini</i>	116
3.2.5 Сравнительный анализ ферритинов у представителей типа Porifera	117
3.2.6 Сравнительный анализ трехмерного окружения ферритинов у губок и других представителей Metazoa	120
3.3 Организация хроматина при анеуплоидиях	124
3.3.1 Изучение межхромосомных взаимодействия на примере анеуплоидий	124
3.3.2 Изменения локальной упаковки хроматина образцов с трисомиями на уровне всего генома	128
Глава 4. Обсуждение результатов	132
Заключение	138
Список литературы	142

Список рисунков

1	Фрагмент жизненного цикла <i>D. discoideum</i>	18
2	Функциональные категории белков <i>D. discoideum</i> , участвующих в регуляции транскрипции.	21
3	Схема анализа силы межтеломерных и межцентромерных взаимодействий на карте средних межхромосомных взаимодействий.	57
4	Подбор параметров для построения карты средних межхромосомных контактов <i>D. discoideum</i>	58
5	Процедура построения средней карты межхромосомных взаимодействий для метацентрических хромосом на примере <i>H. dujardini</i>	59
6	Схема расчета силы компартментов.	62
7	Выбор оптимальных параметров запуска LASCA на карте хроматиновых контактов 0 часа развития <i>D. discoideum</i>	63
8	Контроль качества образцов ATAC-seq.	64
9	График распределения длин фрагментов экспериментов ATAC-seq.	65
10	Значения FRiP для экспериментов ATAC-seq.	66
11	Схема аннотации энхансер-подобных элементов.	68
12	Карта корреляций между биологическими репликами и часами развития.	73
13	Пример карты контактов при разном размере бина.	73
14	Полногеномная карта контактов <i>D. discoideum</i>	75
15	Структура межхромосомных контактов каждой отдельной хромосомы <i>D. discoideum</i>	76
17	Фрагмент scaling-графика для клеточного цикла и часов развития	78
18	Сравнительный анализ карт средних внутрихромосомных взаимодействий для стадий развития и стадий клеточного цикла.	80
19	Сравнительный анализ средних карт межхромосомных взаимодействий <i>D. discoideum</i> и <i>S. cerevisiae</i>	81
20	Различия частот хромосомных взаимодействий между 0 и 8 часами развития.	83

21	Графы дальних взаимодействий на 0 и 8 часах развития.	85
22	Характеристики дальних контактов.	86
23	Количество генов длинных некодирующих РНК, пересекающих основания дальних контактов на 0 и 8 часах развития.	87
24	Различия экспрессии генов длинных некодирующих РНК, лежащих в основаниях дальних контактов в сравнении со случайно выбранными генами длинных некодирующих РНК. . .	88
26	Сила компартментов хроматина <i>D. discoideum</i> на 0 часе развития.	90
27	Пример карты контактов <i>D. discoideum</i> в браузере HiGlass. . . .	91
28	График скейлинга и первой производной для четырех часов развития <i>D. discoideum</i>	92
29	Сравнительная характеристика количества и размеров последовательных и одиночных петель.	92
30	Изучение иерархии в петлях хроматина <i>D. discoideum</i>	93
31	Консервативность петель между часами развития <i>D. discoideum</i> . .	95
32	Изучение покрытия генами разных категорий вокруг петель в сравнении с контролем.	95
33	Колокализация синтенных блоков и петель <i>D. discoideum</i>	96
34	Профиль доли покрытия бинов генами, лежащими на разных цепях ДНК, и значений показателя конвергентности в бинах вокруг оснований последовательных и одиночных петель в сравнении с контролем.	98
35	Примеры симметричных и вытянутых петель в матрице хроматиновых контактов.	99
36	Усредненная карта, центрированная на вытянутых петлях, для разных часов развития.	100
37	График медианного покрытия генами плюс и минус цепи.	100
38	Графики значений сигнала RNA-seq и показателя инсуляции вокруг вытянутых и симметричных петель.	101
39	Зависимость силы петли от уровня экспрессии в основаниях. . . .	103
40	Контактные домены конвергентными парами генов.	105
41	Количество конвергентных пар генов, пересекающих основания петель и внутренние части петель.	106
42	Домен, образованный на телах конвергентных генов с разными уровнями транскрипции.	107

43	Влияние уровня транскрипции пары конвергентных генов на частоты контактов в окне 100 тыс. п.н.	108
44	Примеры аннотации энхансеров.	109
45	Анализ энхансеров <i>D. discoideum</i>	111
46	Анализ функций энхансеров <i>D. discoideum</i>	112
47	Число картированных прочтений для карт Hi-C губок.	114
48	Карта среднего хромосомного взаимодействия для карт Hi-C губок.	115
49	Компартменты в хроматине <i>H. dujardini</i>	116
50	Поиск фонтанов в хроматине губки <i>H. dujardini</i>	117
51	Средняя карта контактов вокруг хроматиновых фонтанов.	118
52	Карты хроматиновых контактов ферритинов у представителей типа Porifera.	121
53	Карта хроматиновых контактов ферритинов у <i>D. rerio</i>	122
54	Карта хроматиновых контактов ферритинов у мыши <i>Mus musculus</i> , полученная методом Micro-C.	123
55	Карта хроматиновых контактов ферритинов для ткани легкого и яичника человека <i>H. sapiens</i>	124
56	Карта межхромосомных контактов для образцов с трисомией по 13, 16, 18 хромосомам.	126
57	Абсолютные значения и доля длины хромосомы, покрытой белок-кодирующей частью генов и ламин-ассоциированными доменами в геноме человека <i>H. sapiens</i>	127
58	Различия частот межхромосомных контактов для больших и малых хромосом с трисомиями по хромосомам 13, 16, 18.	127
59	Примеры карт хроматиновых контактов для двух регионов, которые меняют свою плотность в фибробластах с трисомией по хромосоме 18 в сравнении с контрольным образцом.	128
60	Сравнительная характеристика регионов хроматина с измененной упаковкой у образцов с анеуплоидиями.	129
61	Принадлежность компактизованных и декомпактизованных регионов к эу- и гетерохроматину.	130
62	Характеристика регионов с измененной упаковкой у образцов с анеуплоидиями.	131

Список таблиц

1	Подбор значения h в зависимости от разрешения карты Hi-C . . .	44
2	Алгоритмы поиска хроматиновых петель в данных Hi-C	45
3	Подбор значений переменной для поиска дифференциально контактирующих локусов	50
4	Инструменты для визуализации данных Hi-C	52
5	Показатели качества образцов Hi-C.	72
6	Показатели цис-транс соотношений на разных стадиях развития и клеточного цикла.	74
7	GO термины генов с повышенной экспрессией на 0 часе развития	84
8	Показатели цис-транс-отношения для карт Hi-C губки	113
9	Процент аминокислотного сходства с HuHF и количество аминокислотных замен в функциональных доменах у феритинов представителей Porifera, класс Demospongiae.	119

Список сокращений и условных обозначений

5C	Chromosome Conformation Capture Carbon Copy, захват конформации хромосомы
ALT	альтернативное удлинение теломер (Alternative Lengthening of Telomeres)
ATAC-seq	секвенирование открытых хроматиновых областей (Assay for Transposase-Accessible Chromatin using sequencing)
CID	домен взаимодействия хроматина (chromatin interaction domains)
CTCF	СССТС-binding factor (СССТС-связывающий фактор)
DIRS	Повторяющаяся последовательность промежуточных элементов диктиостелиума (Dictyostelium intermediate repeat sequence)
DNA	Deoxyribonucleic acid (Дезоксирибонуклеиновая кислота)
EM	Electron Microscopy (Электронная микроскопия)
FDR	доли ложноположительных результатов (false discovery rate)
H and L	Heavy and Light chains (Тяжелая и Легкая цепи)
H3K27ac	Histone H3 lysine 27 acetylation, ацетилирование хвоста гистона H3 по лизину в позиции 27
H3K9me2/3	Histone H3 lysine 9 di- and trimethylation, ди- и триметилирование хвоста гистона H3 по лизину в позиции 9
HAATI	теломераз-независимая амплификация гетерохроматина (heterochromatin amplification-mediated and telomerase-independent)
Hi-C	High-throughput Chromosome Conformation Capture (Высокопроизводительное секвенирование взаимодействий хромосом)
ICE	нормализация контактов по методике ICE (Iterative Correction and Eigenvector decomposition)
MCM	минихромосомный поддерживающий комплекс (Minichromosome Maintenance Complex)

MicroC-XL	методика MicroC-XL для изучения трехмерной структуры хроматина (MicroC-XL method)
O/E	наблюдаемое к ожидаемому (observed over expected)
PTM	Histone Post-Translational Modifications, посттрансляционные модификации гистонов
RNA-Seq	секвенирование РНК (RNA sequencing)
SCC	скорректированные по слоям коэффициенты корреляции (Strata-adjusted correlation coefficients)
SMC	Structural Maintenance of Chromosomes (Структурные компоненты хромосомы)
SSE	сумма стандартной ошибки (sum of standard errors)
TE	transposable element, транспозоны
TPM	количество транскриптов на миллион (Transcripts Per Million)
TSS	transcription start site, сайт начала транскрипции
WAPL	Wings Apart-Like (Разгрузчик когезина)
ДНК	дезоксирибонуклеиновая кислота
КПГ	конвергентная пара генов (Convergent Gene Pairs)
ЛАД	Лamina-ассоциированный домен
РНК	рибонуклеиновая кислота
ТАД	Топологически ассоциированный домен
вв-КПГ	конвергентная пара генов, в которой оба гена с высоким уровнем транскрипции
вн-КПГ	конвергентная пара генов, в которой один ген с высоким уровнем транскрипции, а другой с низким
нн-КПГ	конвергентная пара генов, в которой оба гена с низким уровнем транскрипции
оп-КПГ	конвергентная пара генов, ассоциированная с основанием петли (Loop Anchor-associated Convergent Gene Pairs)
тРНК	Transfer RNA (Транспортная РНК)

Введение

Dictyostelium discoideum — гаплоидная амёба, живущая в лесной почве и питающаяся бактериями. При голодании этот одноклеточный организм начинает синтезировать цАМФ, что приводит к агрегации отдельных одноклеточных организмов в единый многоклеточный. Эта программа развития через 24 часа завершается формированием многоклеточного плодового тела и спор. Сходство между *D. discoideum* и клетками животных распространяется и на гены и белки, которые регулируют разнообразные консервативные процессы, такие как деление и рост клеток, хемотаксис, эндоцитоз, факультативная многоклеточность и взаимодействия между хозяином и патогеном [1–5].

Более того, *D. discoideum* является важным объектом для изучения различных заболеваний человека, и именно эксперименты с ним вследствие легкости введения мутаций в гаплоидный организм и простоты культивации помогли внести важный вклад в изучение различных нозологий и состояний, например, устойчивости к химиотерапии опухоли [6], биполярного расстройства [7], митохондриальных заболеваний [8], старения [9], болезни Альцгеймера [10], болезни Паркинсона [11] и других нейродегенеративных заболеваний [12].

Губки (тип Porifera) являются донными животными и представляют, по разным оценкам, один из или самый древний тип существующих Metazoa. Класс Demospongiae включает около 85 % всех существующих и вымерших видов губок, распространённых от тропических до северных морей [13]; губка *Halisarca dujardini*, принадлежащая к классу Demospongiae, обитает в холодных северных морях. Древность губок делает их важным объектом исследования эволюционных процессов, поскольку губки являются одной из базальных ветвей Metazoa. Несмотря на упрощённую организацию, геномы губок имеют многие общие черты с представителями Eumetazoa. Например, было показано, что геном губки *Amphimedon queenslandica* содержит некоторые гены эмбрионального развития [13; 14], изменения экспрессии по стадиям развития схожи с таковыми у представителей Eumetazoa [15; 16]. Более того, клетки губок обладают способностью переходить из одного типа в другой,

подобно трансдифференцированным и стволовым клеткам млекопитающих [17–19].

Другой уникальной особенностью губок является их способность к реакреации из диссоциированных клеток, что было впервые обнаружено Вильсоном в 1907 году [20]. Показано, что *H. dujardini* проявляет высокую способность к реакреации, поддерживаемую защитными и регенеративными механизмами [21]. Исследование *H. dujardini* предоставляет уникальную возможность расширения знаний о клеточной биологии и адаптивных стратегиях у самых ранних представителей многоклеточных организмов, а также о клеточных основах регенерации тканей.

Трисомии, т.е. наличие дополнительной копии одной из хромосом, в большинстве случаев являются летальными для эмбрионов, однако трисомии по хромосомам 13, 18 и 21 присутствуют у новорожденных с частотой от 1:5000 до 1:1000 [22; 23]; реже встречаются мозаичные (представленные не во всех клетках тела) трисомии хромосом 8, 16, 19 и 22. Фенотипические проявления включают полиорганные нарушения: задержки развития, тяжёлые пороки сердца, рецидивирующие инфекции, а также предрасположенность к аутоиммунным заболеваниям. Существует несколько гипотез о роли дополнительной копии хромосомы в возникновении подобных фенотипов. Одна из них постулирует, что существуют дозозависимые гены, дополнительная копия которых ведёт к повышенной экспрессии и объясняет фенотип. Однако более поздние наблюдения поставили под сомнение эту гипотезу, показав, что сверхэкспрессия ограниченного числа генов недостаточна для проявления всех симптомов, представленных у людей с трисомиями [24; 25]. В качестве альтернативы была предложена концепция нарушения клеточного гомеостаза [26]: фенотипические проявления трисомий обусловлены не приобретением конкретной дополнительной хромосомы, а наличием дополнительной хромосомы как таковой. Одной из характерных особенностей трисомий является пангеномная дисрегуляция транскрипции, которая может быть причиной нарушения гомеостаза.

Несмотря на то, что мы хорошо знаем, как устроены отдельные гены *D. discoideum* и как регулируется их экспрессия [27–29], мы плохо понимаем, как устроена регуляция экспрессии у *D. discoideum* на уровне всего генома, а также не имеем сведений о его трехмерной организации. Для губки *H. dujardini* отсутствует информация как о регуляции экспрессии генов, так

и о трёхмерной организации генома. Выбор организмов для исследования обусловлен тем, что губки — одни из самых простых многоклеточных, способных к реагрегации, а *D. discoideum* при голодании запускает программу агрегации одноклеточных организмов с последующей дифференциацией на клеточные типы, а потому важно установить изменения структуры хроматина и координированы ли эти изменения с экспрессией генов при переходе от одноклеточности к клеточному агрегату. С другой стороны, изучение хроматина образцов с трисомиями человека поможет пролить свет на изменение укладки хроматина при наличии дополнительной копии хромосом в сравнении с образцами нормального кариотипа, что, с одной стороны, помогает понять патогенез заболевания, а с другой стороны, также является исследованием координированности между изменением организации хроматина и экспрессии генов по всему геному.

Целью данной работы является изучить устройство хроматина у *D. discoideum*, *H. dujardini* и образцов *Homo sapiens* с трисомиями по хромосомам 13, 16 и 19 на разных уровнях организации, а также взаимосвязь изменений трёхмерной организации с изменениями транскрипционной программы.

Для достижения поставленной цели необходимо было решить следующие **задачи**:

1. Проанализировать данные Hi-C, полученные для стадий развития *D. discoideum* и стадий реагрегации *H. dujardini*.
2. Изучить организацию хроматина на уровне межхромосомных взаимодействия, компартментов и экструсионных взаимодействия для *D. discoideum* и *H. dujardini*.
3. Исследовать взаимосвязи найденных хроматиновых структур с экспрессией генов, используя данные экспрессии для *D. discoideum*.
4. Провести сравнительный анализ хроматинового окружения гена на примере генов ферритинов у *H. dujardini* и других представителей Metazoa.
5. Изучить влияние дополнительной копии хромосом на организацию ДНК в ядре на примере образцов с трисомиями у человека в сравнении с образцами нормального кариотипа.
6. Изучить взаимосвязи хроматина и систем регуляции транскрипции, используя данные эпигенетики, полученные для *D. discoideum*, а также данные о ламине для образцов с трисомиями у человека.

Научная новизна:

1. Впервые проанализировано устройство хроматина у гаплоидной амебы *D. discoideum* и морской губки *H. dujardini*.
2. Впервые обнаружены вытянутые петли в хроматине *D. discoideum* и показана их связь с уровнями транскрипции.
3. Выполнено оригинальное исследование дальних контактов (преимущественно межхромосомных) хроматина *D. discoideum* и показана их взаимосвязь с длинными некодирующими РНК, значительно меняющими свои уровни экспрессии.
4. Впервые выдвинута гипотеза о конвергентной транскрипции как механизме формирования петель у *D. discoideum*.
5. Впервые аннотированы регуляторные элементы в геноме *D. discoideum* и показано их обогащение в основаниях петель.
6. Впервые аннотированы хроматиновые фонтаны в хроматине морской губки *H. dujardini*.
7. Впервые проанализирована пространственная структура образцов с трисомиями по хромосомам 13, 16 и 18 у человека.
8. Впервые показана связь изменения частот контактов хромосом с трисомиями по хромосомам 13, 16 и 18 с содержанием ламина-ассоциированных доменов на таких хромосомах.

Практическая значимость *D. discoideum* — важный модельный организм и один из примеров независимого возникновения многоклеточности. При этом многие процессы в нем протекают похоже на высших эукариот, причем отвечают за них гомологичные гены и белки. Совместно с простотой культивирования это делает *D. discoideum* важным объектом для изучения как фундаментальных процессов, таких как хемотаксис, эндоцитоз, рост и деление, так и различных нозологий и состояний. Однако при большой популярности данного модельного организма информация о трехмерной укладке его хроматина в литературе отсутствовала.

В настоящей работе нами была исследована трехмерная структура хроматина *D. discoideum* и *H. dujardini* на разных стадиях развития. Разработанная методология может быть применена для изучения устройства хроматина других беспозвоночных. Кроме того, настоящее исследование не только восполняет пробел об устройстве хроматина двух важных организмов, но и проливает свет на разнообразие устройства хроматина организмов

в отсутствие белка CTCF. Результаты работы показали, что конвергентная пара генов может быть рассмотрена как барьер для экстрезии в других организмах.

Исследование, проведённое для образцов с трисомиями по хромосомам 13, 16 и 18 у человека, позволяет дополнить картину патогенеза анеуплоидий, добавляя наблюдения о глобальной реорганизации хромосом в ядре и изменении взаимодействия с белками ламины к имеющимся гипотезам о дозозависимых генах и нарушении гомеостаза.

Основные положения, выносимые на защиту:

1. Хроматин *D. discoideum* образует множество последовательных петель с высокой консервативностью между стадиями и низким уровнем иерархичности.
2. Основания хроматиновых петель *D. discoideum* обогащены энхансер-подобными элементами и границами синтенных блоков.
3. Хроматин губки *H. dujardini* образует фонтаны; различные хроматиновые петли и топологически ассоциированные домены отсутствуют.
4. Наличие дополнительной копии хромосом вызывает как изменения в межхромосомных взаимодействиях, так и реорганизацию локальной упаковки хроматина, причём степень этих изменений варьирует для различных хромосом и, по-видимому, связана с изменениями взаимодействий с белками ламины.

Достоверность полученных результатов основана на использовании современных и адекватных методов обработки полученных данных. Результаты находятся в соответствии с результатами, полученными другими авторами.

Апробация работы. Основные результаты работы докладывались на: Всероссийской конференции «Информационные технологии и системы» (Пермь, 2019; Москва, 2021, Москва, 2022), 10-ой Московской конференции по вычислительной молекулярной биологии МССМВ-2021 (Москва, 2021), ИТИС-2021, 13-ой Международной мультikonференции «Биоинформатика геномной регуляции и структурной/системной биологии» BGRS-2022 (Новосибирск, 2022).

Личный вклад. Автор принимал ведущее участие в анализе данных Hi-C, ATAC-seq, RNA-seq и ChIP-seq, включая разработку и реализацию

алгоритмов и скриптов для анализа данных, предварительную обработку и контроль качества, статистический анализ, визуализацию и интерпретацию результатов. Автором были аннотированы последовательные и вытянутые петли, конвергентные пары генов, энхансер-подобные элементы у *D. discoideum*, хроматиновые фонтаны *H. dujardini*, регионы с измененной упаковкой для образцов с трисомиями. Автором были исследованы особенности упаковки хроматина *D. discoideum* на вегетативной стадии. Автором было изучено хроматиновое окружение генов ферритинов у *H. dujardini* и представителей Metazoa. Автором был проведён анализ взаимосвязи изменения частот контактов между хромосомами и изменение локальной упаковки при наличии дополнительной копии хромосомы у образцов с трисомиями.

Подготовка материалов к публикации и написание статей проводилось при личном участии автора. Подготовка библиотек ДНК и РНК для экспериментов Hi-C и RNA-seq для *H. dujardini* была проведена А.В. Черкасовым и А.В. Рябовой. Подготовка библиотек ДНК и РНК для экспериментов Hi-C и RNA-seq *D. discoideum* была выполнена С.В. Ульяновым. Синхронизация клеточной культуры *D. discoideum* по клеточному циклу была проведена П.А. Васильевым. Подготовка библиотек для экспериментов Hi-C у образцов с трисомией была проведена С.В. Ульяновым, эксперимент FISH выполнен А.С. Штомпель.

Публикации. Основные результаты по теме диссертации изложены в 9 печатных изданиях, из которых 3 изданы в научных журналах, индексируемых Web of Science и Scopus, 5 — в тезисах докладов.

Объем и структура работы. Диссертация состоит из введения, 4 глав и заключения. Полный объём диссертации составляет 170 страниц, включая 62 рисунка и 9 таблиц. Список литературы содержит 301 наименование.

Глава 1. Обзор литературы

1.1 Биология *Dictyostelium discoideum*

Dictyostelium discoideum — гаплоидная амеба, населяющая лесную почву и питающаяся бактериями с помощью фагоцитоза. При голодании одноклеточные амебы вступают в жестко регулируемый цикл развития [30], который завершается формированием многоклеточного плодового тела (кратко изложено в работе [31]) (Рисунок 1).

Многоклеточность по типу агрегации и формирования плодового тела отличаются от «классической» многоклеточности, в которой клетки остаются рядом после митоза, и остаётся довольно простой. Это, как предполагается в [32], является следствием частичного или полного отсутствия положительного отбора на усложнение плодового тела, представляющего одно поколение клеток. Причины этого, по-видимому, состоят в том, что одноклеточная стадия, в которой клетки пролиферируют и соревнуются за питание, является для организма высоко эффективной и потому предпочтительной. Но при этом важно отметить, что среди агрегирующих эукариот *Dictyostelia* демонстрируют наиболее сложную форму многоклеточности [32]. Помимо спор, они могут дифференцироваться в четыре типа соматических клеток, а агрегаты могут трансформироваться в промежуточную подвижную стадию «слизняка», которая доставляет клеточную массу в оптимальное место для распространения спор.

Среди факторов, повлиявших на возникновение многоклеточности у *D. discoideum*, выделяют некодирующие РНК и РНК класса I [33]. Возраст последних насчитывает около 600 миллионов лет, поскольку они описаны для всех четырех подгрупп семейства *Dictyosteliidae*. Несмотря на древнюю историю, некодирующие РНК смогли сохранить высокое сходства GC%-состава и мотивов третичной структуры РНК.

Важно также отметить, что 80 % генов, кодирующих белки, необходимые для биохимических каскадов регуляции перехода от одноклеточности к многоклеточности, представлены и у одноклеточных представителей *Amoebozoa* [34]; однако гены, кодирующие ферменты, участвующие в синтезе

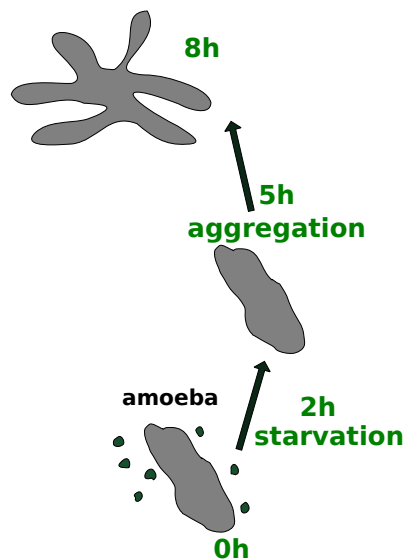


Рисунок 1 — Фрагмент жизненного цикла *D. discoideum*.

сигнальных молекул, рецепторов и секретируемых белков, часто присутствуют только в геномах представителей Dictyostelida.

Формирование плодового тела требует много энергии, и в отсутствие питательных веществ *D. discoideum* использует макроаутофагию (получение питательных веществ за счет пожирания других клеток, расщепленных фаголизосомой), а также ксенофагию (распознавание и переваривание внутриклеточных патогенов) [35]. *D. discoideum* является одним из самых известных модельных организмов для изучения различных процессов, таких как движение клеток, транспорт органелл, цитокинез и эндоцитоз, факультативная многоклеточность и взаимодействия между хозяином и патогеном, а также клеточный сигналинг [1—5].

Более того, *D. discoideum* является важным объектом для изучения различных заболеваний человека, и именно эксперименты с ним вследствие легкости введения мутаций в гаплоидный организм и простоты культивации помогли внести важный вклад в изучение различных нозологий и состояний, такие как устойчивость опухолей к химиотерапии [6], биполярное расстройств [7], митохондриальные заболевания [8], старение [9], болезнь Альцгеймера [10], болезнь Паркинсона [11] и другие нейродегенеративные заболевания [12; 36]. Рассмотрим один из примеров подробнее. Вальпроевая кислота и литий являются важными препаратами для лечения биполярного расстройства. Однако, несмотря на клиническое использование в течение значительного времени (литий использовался более века [37], а вальпрое-

вая кислота — с 1963 г. [38]), механизм действия и терапевтические мишени препаратов установлены не были. Чтобы идентифицировать сигнальные пути, на которые воздействует литий, были созданы культуры с мутациями в различных генах *D. discoideum* для идентификации генов, которые придают устойчивость к литию [39]. После того, как эффекты препарата были обнаружены и охарактеризованы у *D. discoideum*, результаты были проверены в клетках млекопитающих; исследование подтвердило, что увеличение размера нейронов ганглиев задних корешков спинного мозга крыс, вызванное тремя различными препаратами для лечения биполярного расстройства, было ингибировано выключением гена пролилолигопептидазы [40].

D. discoideum имеет очень АТ-богатый (77,75%) гаплоидный геном, организованный в 6 хромосом, который был секвенирован и аннотирован [41]. Геномы некоторых других организмов, например, *Plasmodium falciparum*, также имеют исключительно низкий GC-состав, но причины этого неизвестны. Плотность генов высокая, и 62 % генома кодируют белки. Геном *D. discoideum* богат повторами разных классов, которые составляют до 10 % генома [42], причём более половины от этого числа приходится на повторы одного из трех классов — TRE5-A, DIRS-1 и Skipper-1. Каждая хромосома содержит кластер повторов вблизи одной из теломер. Вероятно, эти последовательности и выполняют роль центромеры, так как они кластеризуются рядом с центросомой на краю ядра во время интерфазы и рассеиваются при митозе [43]. Центромерные повторы богаты транспозонами (около 86 % длины центромер), около половины из них это *Dictyostelium intermediate repeat sequence* (DIRS); в остальной части генома эти повторы встречаются реже [44]. В пользу DIRS-центромерной гипотезы говорит наличие центромерного варианта гистона CenH3 в регионе [43] и высокий уровень гистоновой метки гетерохроматина H3K9me3 [45]. Кроме того, эти повторы взаимодействуют друг с другом чаще ожидаемого, что видно на Hi-C картах [46]. Недавняя работа [47] пролила свет на устройство центросомы *D. discoideum* в сравнении с дрожжами и животными и также подтвердила акроцентрическое устройство хромосом *D. discoideum*.

Интересно также отметить, что инактивация транспозонов DIRS происходит путём взаимодействия с антисмысловыми РНК, синтезированными с участием РНК-зависимой РНК полимеразы RrpC [48]; нокаут *RrpC* приводит к потере большинства малых DIRS-1 РНК, а сами транспозоны могут быть

реактивированы. Однако при такой высокой плотности генов случайные вставки транспозонов, вероятно, приведут к поломкам белок-кодирующих генов и потому будут подвержены отрицательному отбору. Но у *D. discoideum* наблюдается высокое содержание транспозонов, что ведёт к выводу, что встраивание транспозона позволяет получить дополнительные эволюционные преимущества [49], или существуют механизмы снижения урона от транспозонов, например встраиванием транспозонов лишь в определённые участки генома, что предотвращает нарушение кодирующих последовательностей. Например, транспозон TRE5-A преимущественно встраивается в 5'-область генов тРНК, поскольку для его интеграции необходимо участие транскрипционных факторов TFIIIВ или TFIIIC [50]. В литературе известно ещё минимум пять групп транспозонов, которые независимо приобрели специфичность к 5'-области генов тРНК [51].

Классические последовательности теломер не обнаружены ни в одной из шести хромосом *D. discoideum*. Данные секвенирования генома указывают на то, что элементы, похожие на неполные копии рибосомальной ДНК (рДНК), присутствуют на концах всех хромосом [41]. В недавней работе [52] было показано, что последовательности концевой области рДНК содержат простые повторы A(G)_n, где число G является переменным, и продолжают как одноцепочечная последовательность на 3'-конце. Области, богатой гуанином, предшествуют четыре тандемных повторения двух разных мотивов ДНК. У *D. discoideum* есть ортолог транскриптазы *TERT*, *terthp*. Однако было показано, что клетки с нокаутом *terthp* жизнеспособны и могут поддерживать длины теломер, предположительно, с помощью механизмов гомологической рекомбинации, таких как альтернативное удлинение теломер (Alternative Lengthening of Telomeres, ALT) или теломераз-независимая амплификация гетерохроматина (heterochromatin amplification-mediated and telomerase-independent, НААТИ) [52]. Альтернативное удлинение теломер описано у дрожжей и опухолевых клеток [53], а внехромосомная ДНК может служить субстратом в реакциях гомологичной рекомбинации ДНК. При НААТИ теломеры заменяются повторами, чаще всего рДНК, которые располагаются в субтеломерных регионах.

Несмотря на различные косвенные свидетельства наличия ядерной ламины у Простейших, в течение многих лет у них не было обнаружено никаких генов ламины, пока в 2012 году не был идентифицирован ламиноподобный

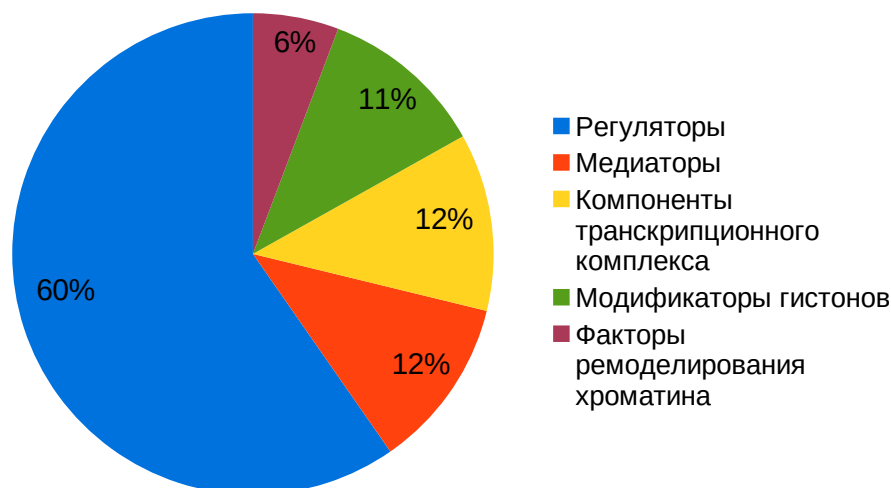


Рисунок 2 — Функциональные категории белков *D. discoideum*, участвующих в регуляции транскрипции, аннотированы по dictyBase. Адаптировано из работы [58].

белок у *D. discoideum* [54]. Вероятно, ламина-зависимое образование гетерохроматина вовлечено в дифференцировку клеток *D. discoideum* так, как это показано, например, для развития сетчатки глаза [55] и эмбриона дрозофилы [56], однако работ по этой тематике на *D. discoideum* пока нет.

У *D. discoideum* обнаружено около 240 генов (Рисунок 2), для продуктов которых показано или предсказано участие в регуляции транскрипции, ремоделировании хроматина или модификации гистонов [57]. Большинство из этих генов имеют ортологи в геноме *D. purpureum*, а многие из этих генов также сохраняют значительное аминокислотное сходство между *D. discoideum* и представителями Eumetazoa, несмотря на разделяющие их сотни миллионов лет эволюции. Помимо сходства аминокислотных последовательностей, схожие изменения уровней экспрессии регуляторов транскрипции при переходе между стадиями развития позволяют предполагать, что ортологи генов Eumetazoa обладают консервативными функциями в *D. discoideum* [58]. Многие гены, кодирующие транскрипционные факторы, также образуют регулоны (системы координированной регуляции уровней транскрипции в цикле развития), что говорит, вероятно, о наличии разветвленных систем регуляции [59].

Репертуар метильных и ацетильных посттрансляционных модификаций хвостов гистонов H3 и H4 у *D. discoideum* аналогичен таковому у млекопитающих, за исключением метилирования по позиции H3K27 [60]. Эта модификация у Metazoa часто связана с образованием факультативно-

го гетерохроматина [61], которого у *D. discoideum* не обнаружено. Однако у *D. discoideum* идентифицировали как H3K9Me₂, так и H3K9Me₃ — модификации, связанные с гетерохроматином и отсутствующие у почкующихся дрожжей [60].

Среди наиболее изученных транскрипционных факторов у *D. discoideum* — представители семейств bZIP, Myb, STAT и GATA, которые активируются под действием факторов индукции дифференциации (DIF) [62]. Эта группа факторов является модуляторами хемотаксиса и агрегации одноклеточных амёб в «слизнях», а также индукторами дифференциации клеток-предшественниц «ножки» будущего многоклеточного организма [63; 64]. Другой важный процесс — образование спор в плодовых телах — у *D. discoideum* индуцируется с помощью цАМФ, который опосредованно влияет на протеинкиназу А. Помимо этого цАМФ вовлечён в хемотаксис к месту агрегации одноклеточных организмов, если секретирован в окружающую среду [3]. Дальнейший каскад включает в себя транскрипционные факторы CudA, VzrF, SrfA и SpaA, причем последний находится выше по каскаду и способен регулировать транскрипцию остальных трех перечисленных транскрипционных факторов. Экспрессия преспоровых генов в клетках с мутированным геном SpaA была сильно снижена, а у генов, участвующих в спорообразовании, вообще отсутствовала [65]. CudA способствует экспрессии преспоровых генов в клеточном агрегате, в то время как VzrF (гомолог белка CREB) и SrfA действуют позже, индуцируя экспрессию характерных для спор генов. Мутанты *VzrF*⁻ образуют споры, окруженные стенкой, но эти споры распадаются, все еще находясь в спороносной головке. Данные RNA-seq этих мутантов показывают пониженный уровень экспрессии 205 генов у мутантов по сравнению с диким типом, многие из которых кодируют целлюлозосвязывающие и сахаросвязывающие белки [66]. Представители семейства Srf являются одними из главных регуляторов актинового цитоскелета и сократительного аппарата у высших эукариот [67]. Белки семейства Srf могут влиять на выбор одной из двух программ развития после агрегации [68]. Аммиак способствует развитию по пути прямой агрегации у *D. discoideum*, но *SrfB*⁻-мутанты демонстрируют заметную тенденцию к образованию слизней. Такие мутанты также имеют дефекты морфогенеза и образуют сферические, а не эллиптические споры с уменьшенной жизнеспособностью [69]. Потеря четвертого транскрипционного

фактора из перечисленных, StkA, заставляет клетки-предшественники споры трансдифференцироваться в клетки ножки [70].

1.2 Губки

Губки (тип Porifera) являются донными животными и, вероятно, представляют самый древний тип существующих Metazoa. Тела губок состоят из различных типов клеток, которые проходят через фазы окислительно-восстановительного и метаболического цикла. Градиенты воды, кислорода и метаболитов стимулируют непрерывные движения и трансформации клеток, например, у губок выделяют водоносную систему, которая также подвергается многочисленным перестройкам как в ходе жизненного цикла, так и в ответ на постоянные изменения в окружающей среде [71; 72].

Кроме того, клетки губок обладают способностью переходить между различными типами, подобно индуцированным стволовым клеткам млекопитающих [17—19]. Механизм включает трансдифференцировку клеток посредством мезенхимально-эпителиального перехода [17], затрагивающего преимущественно архециты и хоаноциты. Было показано, что в регенерации активно участвует апоптоз, способствуя удалению поврежденных клеток [17]. В то же время пространственно далекие от раневой поверхности клетки не меняют своего уровня пролиферации и апоптоз происходит там значительно реже. Другой связанной с этим особенностью клеток губок является их способность к реагрегации из диссоциировавших клеток, что было впервые описано Вильсоном в 1907 году [20].

Пористое тело губок, в котором происходят постоянные движения воды и метаболитов, способствует симбиотическим отношениям с множеством разных бактерий. Такое разнообразие бактериальных симбионтов губок может служить источником для поисков новых антибиотиков [73].

Губка *Halisarca dujardini*, принадлежащая к классу Demospongiae, представляет собой важный объект исследований в контексте адаптации и эволюции губок. Эксперименты показали, что *H. dujardini* также проявляет высокую регенеративную способность и способность к реагрегации [21]. Более того, известно о наличии множества бактериальных симбионтов внутри

тела *H. dujardini* [74], в то время как *Halichondria panicea* представляет собой губку с низкой микробной численностью, где среди групп доминируют альфапротеобактерии [75]. Это подчеркивает важность взаимодействия с микробиотой для *H. dujardini*.

Геном губок характеризуется высокой плотностью генов с минимальными межгенными пространствами, как и у представителей Opisthokonta. Однако регуляторные последовательности промоторов губок ближе к Eumetazoa [76]. Геном представителя класса Demospongiae, обитателя коралловых рифов Австралии, *Amphimedon queenslandica* содержит лишь отдельные гены, кодирующие белки, участвующие в эмбриональном развитии, что, по-видимому, свидетельствует либо о вторичной потере генов, либо об упрощённой программе развития общих предков представителей Metazoa [13; 14]. В течение развития у губок *A. queenslandica* наблюдаются паттерны дифференциальной экспрессии генов, схожие с таковыми у Eumetazoa [15; 16]. Транскрипционная регуляция генов губок включает дальние энхансеры, репрессивный гетерохроматин и гистоновые модификации активного хроматина [77]. Геном губок содержит гены-гомологи компонентов PRC2 комплекса дрозофилы и мотивы связывания факторов гетерохроматина, что также свидетельствует о консервативности гетерохроматина в эволюционно далеких организмах [77–81].

Многие энхансеры губок *A. queenslandica* расположены в синтенных блоках, определённых для Porifera и Metazoa [82; 83]. Это, по-видимому, свидетельствует о древности этих регуляторных элементов и их значимой роли в регуляции экспрессии генов.

Другой уровень регуляции у губок обеспечивается некодирующими РНК. Среди них выделяются длинные некодирующие РНК, многие из которых консервативны с Eumetazoa, а уровни их экспрессии у губок также демонстрируют изменения в ходе развития [15]. Кроме того, у губок выделяют несколько классов малых некодирующих РНК [84]. Особый интерес представляет класс piwi РНК, роль которых заключается в инактивации транспозонов. Для млекопитающих главным механизмом инактивации транспозонов является ДНК метилирование [85; 86], однако для дрозофилы [87; 88], моллюсков [89], так же как и для губок, разнообразие piwi РНК оказывается весьма обширным, что указывает на их ведущую роль в инактивации транспозонов.

1.3 Ферритины

Железо представляет собой важный для организма микроэлемент, поскольку он входит в состав гема и участвует в переносе кислорода. Низкие концентрации железа ведут к нарушению тканевого дыхания, а избыточно высокие обладают цитотоксичностью. Потому клеткам необходимы как системы захвата и накопления ионов железа, так и системы связывания и детоксикации избыточного количества свободных ионов.

Ферритины представляют собой группу высокоспециализированных белков, способных связывать и накапливать ионы трехвалентного железа при повышенной концентрации железа в крови [90]. В условиях дефицита железа, например, при голодании, ферритиновые комплексы могут высвободить ионы железа. Структура ферритинов млекопитающих включает тяжелые (H) и легкие (L) цепи. Субъединицы H обеспечивают быструю детоксикацию железа с помощью ферроксидазы путём окисления Fe^{2+} в Fe^{3+} [91; 92], тогда как субъединицы L способствуют минерализации и долговременному хранению железа [93; 94]. Для белков суперсемейства ферритиноподобных описано участие в различных клеточных процессах, таких как окислительные реакции, репликация ДНК, биосинтез хлорофилла, образование оболочек эндоспор, метаболизм жирных кислот и другие [95–97].

Ферритины беспозвоночных животных выполняют уникальные функции, отличные от ферритинов позвоночных. Например, ферритиноподобные белки морских беспозвоночных связаны с иммунными функциями [98], а ферритин ChF морского трубчатого червя *Chaetopterus* sp. участвует в производстве биoluminesцентной секретируемой слизи [99]. Для кристаллической структуры ферритина моллюска *Sinonovacula constricta* было предсказано наличие сайтов связывания железа и ферроксидазного центра, схожих с ферритинами млекопитающих [100].

Морские губки растут только в присутствии ионов железа [101]. Некоторые бактериальные симбионты используют рецепторы железа или ферментативные системы для извлечения железа из губок и переноса его в своё внутриклеточное пространство [102; 103]. Эти каскады могут активироваться при низких концентрациях железа в окружающей среде [104]. Грамотрицательные бактерии также участвуют в регуляции врожденного иммунитета

беспозвоночных и позвоночных животных с использованием механизмов, включающих ферритины [105; 106].

Геном *D. discoideum* не содержит последовательностей, ортологичных ферритинам Eumetazoa, однако в нем закодирован ортолог бактериального ферритина [107]. Обладая, однако, высокой способностью к фагоцитозу, *D. discoideum* обладает большим набором систем для метаболизма и использования железа для обезвреживания внутриклеточных патогенных бактерий, таких как *Legionella* и *Mycobacteria*, в связи с его высокой токсичностью в свободной форме и отсутствию у данных бактерий защитных механизмов [107].

1.4 Анеуплоидии

Анеуплоидия — это отклонение числа копий одной или нескольких хромосом. Наиболее распространенными анеуплоидиями являются трисомии (наличие одной дополнительной копии определенной хромосомы) и моносомии (потеря одной копии в паре гомологичных хромосом). Однако двойные трисомии (трисомии двух хромосом), тетрасомия, пентасомия и даже гексасомия аутомом также присутствуют при некоторых типах рака [108]. Аутомомная трисомия — наиболее распространенный тип анеуплоидии (около 60 % клинических случаев), ассоциированный с невынашиванием беременности [109; 110] и аномалиями эмбрионального развития [111]. Большинство аутомомных трисомий эмбрионально летальны [110], однако трисомии по хромосомам 13, 18 и 21 присутствуют у новорожденных с частотой от 1:5000 до 1:1000 [22; 23] и проявляются синдромами Патау, Эдвардса и Дауна соответственно; реже встречаются мозаичные (в части тканей) трисомии хромосом 8, 16, 19 и 22. Недавние исследования показали, что пятилетняя выживаемость детей с трисомией по хромосомам 13 и 18 составляет 9,7 % и 12,3 % соответственно, однако, люди с синдромом Дауна имеют среднюю продолжительность жизни около 60 лет [112]. Трисомии характеризуются некоторыми общими симптомами, такими как умственная отсталость, пороки сердца, множественные врожденные пороки развития, рецидивирующие инфекции, а также предрасположенность к острым лейкозам. Иная ситуация наблюдается с дополнительной копией одной из половых хромосом, что

обычно приводит к менее тяжелым фенотипам, характеризующимся бесплодием и замедлением роста [113]. Следует отметить, что моносомия любой аутосомной хромосомы является эмбрионально летальной, тогда как потеря половой хромосомы нет, как в случае с моносомией хромосомы X, которая фенотипически проявляется синдромом Тернера.

Патологию трисомий изначально связывали с наличием дополнительных копий генов, расположенных на дополнительной хромосоме и, как следствие, повышенной экспрессией этих генов. Например, фенотип синдрома Дауна связан с так называемой критической областью синдрома Дауна, расположенной на длинном плече chr21. Аналогично, в генетических исследованиях были найдены участки хромосомы 18, критические для проявления синдрома Эдвардса [114]. Однако более поздние наблюдения поставили под сомнение эту гипотезу, показав, что сверхэкспрессия ограниченного числа генов недостаточна для проявления всех симптомов, представленных у людей с трисомиями [24; 25]. В качестве альтернативы была предложена концепция нарушения клеточного гомеостаза [26]: эта модель постулирует, что фенотипические проявления трисомий обусловлены не приобретением конкретной дополнительной хромосомы, а наличием дополнительной хромосомы как таковой. Одной из характерных особенностей трисомий является пангеномная дисрегуляция транскрипции, которая может быть причиной нарушения гомеостаза. Сдвиги в программе экспрессии генов, в свою очередь, могут быть вызваны усилением экспрессии регуляторных белков, кодируемых дополнительной хромосомой [115]. Эти регуляторные белки могут изменять профили эпигенетических меток [116], доступность хроматина [117] и профили метилирования ДНК [118; 119] по всему геному, что в свою очередь влияет на активность промоторов генов. Другой возможной причиной нарушения регуляции транскрипции, индуцированной трисомией, является реорганизация пространственной структуры генома [120] и пространственное сближение регуляторных элементов, расположенных как на одной хромосоме, так и на разных хромосомах [121].

Исследования показали, что наличие дополнительной хромосомы вызывает изменения в структуре хроматина на нескольких уровнях. В фибробластах, несущих третью копию хромосомы 21 (Tr21), несколько других хромосом меняют свою компактность и положение внутри ядра [122]. В Tr21-индуцированных плюрипотентных стволовых клетках (iPSC) и клетках-

предшественниках нейронов (NPC), происходящих из iPSC, наблюдались глобальные изменения в структуре межхромосомных и внутрихромосомных контактов [123]. Культивированные эпителиальные клетки толстой кишки человека (HSEC), несущие дополнительную копию хромосомы 7, имеют локальные различия в профилях компартментов A/B хроматина, сопровождающиеся изменениями уровня транскрипции внутри локусов, принадлежность к компартментам которых меняется при трисомии [124]. Сообщалось также, что наличие дополнительной копии даже небольшой хромосомы влияет на расположение других хромосом внутри ядра [125], хотя такие хромосомы имеют тенденцию ко-локализоваться с двумя своими сестринскими хромосомами [126]. Открытым остается вопрос, обладают ли различные трисомии какими-то общими чертами при трехмерной реорганизации генома.

1.5 Организация хроматина

Современные теории о структуре хроматина свидетельствуют в пользу ведущей роли экструзии петель и разделения фаз хроматина [127], хотя отдельные работы говорят об этих двух системах как противодействующих и находящихся в равновесии [128; 129]. Об их независимости свидетельствует и то, что деплеция когезина приводит к ослаблению топологически ассоциированных доменов и хроматиновых петель, но не влияет на компартменты [127; 130].

Один из механизмов, вовлеченный в организацию хроматина, — это жидкость-жидкостное разделение фаз. Предполагается, что разделение фаз является результатом различий как в физико-химических свойствах, в частности, гидрофильности, так и в концентрациях белков и некодирующих РНК [131]. В результате выделяют две фазы: активный эухроматин и неактивный гетерохроматин, названные в трехмерной геномике A- и B-компартментами соответственно. На карте хроматиновых контактов компартменты характеризуются повышенной частотой взаимодействия между регионами, принадлежащими одному и тому же компартменту, и пониженной частотой взаимодействия между регионами разных компартментов.

Кроме того, компартменты высоко коррелируют со временем репликации, активный компартмент реплицируется раньше во времени, чем неактивный [132] и плотность генов в активном компартменте выше [133]. Время репликации, в свою очередь, оказывает влияние на поддержание эпигенетических меток и принадлежность к компартментам [134], усложняя понимание причинно-следственных связей. Однако существуют организмы, для которых компартменты на картах Hi-C обнаружены не были, например, *Saccharomyces cerevisiae* [135] и *Schizosaccharomyces pombe* [136].

Экструзия петель — это продолжающийся во времени АТФ-зависимый процесс протягивания ДНК через (согласно другим данным — вдоль) белковый комплекс в форме кольца, пока не произойдёт столкновение с другим белком, обладающим способностью быть барьером экструзии, или пока комплекс не отвалится или будет снят с ДНК. Экструзия петель осуществляется белками семейства SMC (structural maintenance of chromosomes), которые представляют собой моторные белки, состоящие из двух головок, способных независимо протягивать нити ДНК с затратой АТФ [137]. Экструзия петель, помимо компактизации хроматина, играет роль в пространственном сближении энхансера и промотора, в пользу чего свидетельствует то, что энхансер-промоторные взаимодействия имеют повышенную частоту контактов в виде точки на карте хроматиновых контактов [138; 139]. Кроме того, описаны также домены (chromatin interaction domains, CID), формирующиеся на телах активно экспрессируемых генов [140].

Основным типом экструзионных взаимодействий являются хроматиновые петли — взаимодействия двух регионов, линейно друг от друга удалённых, на карте контактов представляют собой точечное обогащение частоты контактов. Были выделены топологически ассоциированные домены (ТАДы), которые представляют собой глобулу ДНК, чья длина достигает несколько сотен тысяч нуклеотидов, характеризующуюся повышенной частотой контактов по сравнению с окружением [141]. Исходно считалось, что границ петель и ТАДов характеризуются высокой консервативностью: порядка 70 % границ остаются неизменными между клеточными типами [141; 142]. Однако для млекопитающих была показана высокая динамичность и вариабельность ТАДов и петель, как с помощью Hi-C единичных клеток [143], так и с помощью микроскопии живых клеток [144; 145], а исходные наблюдения оказались результатом усреднения петлевых взаимодействия

тысяч отдельных клеток. У дрозофилы, по-видимому, даже на уровне единичных клеток сохраняется большее сходство с границами, полученными из средних данных по сотням тысяч клеток (bulk) [146]. Наличие ТАДов у *S. cerevisiae* остается дискуссионным вопросом: ряд работ отмечают наличие ТАД-подобных структур на карте Hi-C размером несколько десятков тысяч п.н. [147], с помощью метода Hi-CO были обнаружены «ориентированные кластеры» [148], а Micro-C позволил разметить «самоассоциированные домены» (CID) размером до 10 тысяч п.н. [149].

Границы петель высших эукариот часто представляют собой конвергентно ориентированные белки CTCF и когезин, заякоренный на них [150]. Когезин осуществляет экструзию — протягивание ДНК и образование петель; CTCF действует как барьер для экструзии, которые обладает способностью связывать когезин, тем самым стабилизируя его на ДНК [151; 152]. Большинство ТАДов и петель исчезают при деплеции CTCF или когезина [128; 130; 150; 153; 154], а деплеция фактора разгрузки когезина с ДНК WAPL приводит к формированию более длинных петель в экспериментах Hi-C [130; 154] и значительной компактизации хроматина [155].

Недавние исследования выявили ранее неописанные структуры, такие как «вспышки», «шлейфы», «струи» и «фонтаны» (flares/hinges, plumes, jets, fountains) [156—160]. По всей видимости, эти структуры подчиняются общему механизму: загрузка когезина и последующая симметричная экструзия ДНК когезиновым комплексом до встречи с белком CTCF или другим экструзионным барьером. Кроме того, для ранних эмбрионов *Danio rerio* были описаны хроматиновые фонтаны, возникающие при загрузке когезина на активный энхансер, регулирующий экспрессию гена, связанного с эмбриональным развитием [159]. При прогрессии эмбрионального развития когезин осуществляет активную экструзию, что ведёт к возникновению привычных ТАДов и петель в эмбрионах *Caenorhabditis elegans* [158], *D. rerio*, *Oryzias latipes* и *Xenopus tropicalis* [159]. Это, по-видимому, свидетельствует о роли фонтанов в поддержании специфичной и своевременной активации генов развития, либо же фонтаны являются следствием специфических регуляторных взаимодействий, инициированных чем-то иным.

В работе [161] отмечается совпадение границ геномных перестроек с границами ТАДов, а корреляции уровней экспрессии генов внутри ТАДов выше, чем у генов из соседних ТАДов, находящихся на сравнимом рас-

стоянии друг от друга вдоль генома. Исходя из этого, предполагается, что существует эволюционное давление на поддержание границ ТАДов, поскольку эти структуры играют важную роль в регуляции экспрессии генов. Однако невыясненным остается вопрос ограниченного влияния изменения структуры ТАДов на экспрессию генов [162]. Известно, что общее влияние потери когезина и CTCF на уровни транскрипции генов довольно скромные [128; 129], и детектируется преимущественно через несколько дней после деплеции когезина и CTCF [153]. Некоторые гены вообще не меняют уровень экспрессии при структурных изменениях в хроматине [163—166], например, гены домашнего хозяйства и гены, промоторы которых обладают высоким сродством к энхансеру [167]. Однако существуют и обратные примеры: гены, промоторы которых отделены от энхансеров большим линейным расстоянием, а также гены с быстрыми изменениями уровней экспрессии. Например, гены ответа на внешние стимулы оказываются более подвержены влиянию деплеции когезина и CTCF [168]. В связи с этим авторы [162] предполагают, что ТАДы, образованные в результате экстрюзии, представляют собой «грубую структурную основу», чья специфичность регулируется транскрипционными факторами и транскрипционной машинерией, которая может взаимодействовать с факторами экстрюзии петель [169].

В рамках модели экстрюзии предполагается, что факторы экстрюзии играют центральную роль в активном формировании хроматиновых петель, доменов и фонтанов, загружаясь на ДНК и постепенно протягивая хроматин. Эксперименты подтверждают, что экстрюзия петель является распространенным явлением в клетках различных организмов, от бактерий [170] до высших эукариот [171—173]; данные визуализации одиночных молекул также поддерживают эту теорию [137]. Более того, с помощью метода визуализации одиночных молекул было описано динамическое влияние факторов экстрюзии друг на друга, даже если они находятся далеко, что ведет к изменениям размеров петель [174]. Когда факторы экстрюзии находятся в непосредственной близости, экстрюзионные комплексы могут пересекаться, образуя структуру, известную в литературе как Z-петли [174]. Эта структура представляет собой три двухцепочечные спирали ДНК, выстроенные параллельно, с одним когезиновым кольцом на каждом краю.

У эукариотических клеток выделяют три класса SMC комплексов: когезин, конденсин и SMC5/6, каждый из которых выполняют уникальную

функцию. Конденсин играет ключевую роль в уплотнении и расхождении хромосом, когезин обеспечивает когезию сестринских хроматид и организацию интерфазного генома, а SMC5/6 участвует в репарации ДНК [162].

Проведенные исследования показали, что сестринские хроматиды сцепляются во время репликации за счет когезина в S-фазе («сцепляющий» когезин). Мутантный когезин, лишенный способности к сцеплению сестринских хроматид, все равно способен формировать хроматиновые петли («экструзионный» когезин) [175], что, по-видимому, свидетельствует о том, что экструзия петель и сцепление сестринских хроматид реализуются через два независимых механизма, вовлекая разные субъединицы когезина. В недавней работе [176] было показано, что CTCF может преобразовывать когезин из сцепляющего в экструзионный. Кроме того, выявлена консервативность пептидной последовательности, взаимодействующей с консервативным сайтом связывания когезина, между белками CTCF и MCM3 [177]. Совместно эти наблюдения свидетельствуют о том, что именно этот домен белка MCM3 может участвовать в переключении экструзионного в сцепляющий [162] и обратно с помощью домена белка CTCF [176].

Существуют два основных типа выпетливания: одностороннее и двустороннее. В случае одностороннего выпетливания одна из головок остается неактивной и неподвижной, в то время как активная головка протягивает хроматин, образуя петлю, размер которой увеличивается со скоростью v . При двустороннем выпетливании обе головки активны, и размер петли растет со скоростью $2v$. В клетках, вероятно, присутствуют оба типа выпетливания, и, возможно, они взаимозаменяемы. Эксперименты на стареющих лимфоцитах млекопитающих показали, что симметричная экструзия может переходить в асимметричную, если одна из головок когезина сталкивается с препятствием, что указывает на независимый контроль каждого направления [157]. Однако в недавней работе с использованием экспериментов на единичных молекулах, была выдвинута гипотеза, что белки SMC содержат модуль переключения направления экструзии, не зависящий от столкновения с барьером [178].

Экструзия осуществляется за счёт гидролиза АТФ, причём когезин обладает двумя сайтами связывания и гидролиза АТФ, что обеспечивает энергией один цикл работы когезина [179]. Механизм перемещения молекулярного мотора по ДНК представляет из себя сгибание когезинового кольца

в локте, обеспечивающее захват новых локусов ДНК и фиксацию петли за счёт клейсиновой субъединицы. Такой цикл работы называется шагом; по оценкам, когезин выполняет около 1 шага в секунду, каждый из которых перемещает головку на приблизительно 300 нм или 1000 п.н. [179], что позволяет когезину перемещаться по ДНК со скоростью 1000 п.н./сек. Важно учитывать, что у SMC-комплекса, как и любого ДНК-связывающего белка, есть ограниченное время жизни на ДНК. По оценкам, оно достигает нескольких десятков минут [180], что является основной лимитирующей характеристикой времени жизни хроматиновых структур, образованных экстррузией.

Совместно, две описанные характеристики обуславливают производительность эктрудера (иначе — процессивность), равную времени жизни на ДНК, помноженному на скорости экстррузии [145], что отражается в максимально возможном размере петли. Но на практике этот максимальный размер редко достигается из-за ДНК-связывающих белков, способных быть барьерами экстррузии. Наиболее известным барьером экстррузии является CTCF, ДНК-связывающий белок, имеющий в своей структуре 11 цинковых пальцев, которые, по-видимому, обеспечивают его аномально долгое время связывания с ДНК. Это время достигает порядка нескольких десятков минут, что на несколько порядков больше в сравнении с несколькими секундами для обычных факторов транскрипции [181]. CTCF способен взаимодействовать с когезином, связывая и стабилизируя его на ДНК, а также предотвращая его взаимодействие с фактором разгрузки WAPL [152]. Это взаимодействие между CTCF и когезином, по-видимому, происходит только если CTCF на ДНК ориентирован N-концом в сторону эктрудерирующего когезина; C-конец CTCF способен временно останавливать экстррузию вследствие пространственного затруднения, не взаимодействуя с когезином [182]. Недавняя работа [182] выявила роль натяжения ДНК как модулятора исхода столкновения CTCF и когезина. Натяжение ДНК способно увеличить барьерную способность CTCF для экстррузии, что в крайних случаях может привести даже к сжатию петли (уменьшению её размера вследствие увеличившейся жесткости ДНК) [182].

Важно, однако, отметить, что ортологи гена CTCF найдены преимущественно у представителей Bilateria. Важным исключением здесь являются представители нематод: ген CTCF присутствует у представителей базальных

клад, но отсутствует у рода *Caenorhabditis* [183]. При этом, для организмов, в чьем геноме отсутствует ортолог CTCF, также характерны петлевые и доменные структуры. В последние годы появляется все больше данных о других типах экструзионных барьеров, таких как остановившиеся репликационные вилки [184], сцепленные сестринские хроматиды [185] и РНК-полимераза II [186], которые могут играть важную роль даже в присутствии CTCF, но являются основными игроками в его отсутствие [187].

1.6 Данные Hi-C и методы работы с ними

Одним из наиболее удобных и широко используемых методов для изучения пространственной организации ДНК в ядре является метод Hi-C [133]. Этот метод основан на лигировании пространственно приближенных участков ДНК, обработке эндонуклеазами рестрикции, а затем секвенировании полученных мест сшивки. Прочтения, полученные в результате секвенирования, картируют на геном с сохранением информации о принадлежности к одному фрагменту. Затем строится матрица контактов, которая после нормализации содержит не абсолютное количество контактов, а вероятность взаимодействия между геномными локусами [188]. С использованием такой матрицы контактов можно исследовать организацию хроматина на различных уровнях: от межхромосомных взаимодействий и компартментов до петель размером несколько тысяч пар оснований.

1.6.1 Концепция бина матрицы и определение его размера

Данные Hi-C представляют собой матрицу парных контактов между участками генома, симметричную относительно главной диагонали. Геномные локусы делятся на неперекрывающиеся окна, называемые «бинами», которые представляют собой сегменты генома одинакового размера. Эти бины объединяют частоты контактов из соответствующих рестрикционных

фрагментов. Размер бина определяет разрешение карты Hi-C и теоретически ограничен размерами использованных рестрикционных фрагментов.

Разрешение для Hi-C достигается с использованием различных рестрикционных фрагментов, например, с 6-буквенными сайтами узнавания эндонуклеазы рестрикции (HindIII), что обеспечивает разрешение в 2-4 тыс. пар нуклеотидов (п.н.) в разных геномах, или 4-буквенными сайтами узнавания эндонуклеазы рестрикции (например, DpnII), что позволяет получать карты с разрешением в 128 п.н. Однако частота контактов для рестрикционных фрагментов HindIII размером менее 200 п.н. может быть низкой из-за пространственных затруднений для эндонуклеаз рестрикции [189].

Эндонуклеаза рестрикции HindIII способна нарезать геном человека на фрагменты размером около 4 тыс. п.н., потому матрица контактов с разрешением 10 тыс. п.н. может иметь незаполненные бины даже при достаточной глубине секвенирования. Это вынуждает снижать разрешение матриц Hi-C, что приводит к потере биологической информации. Процедура снижения разрешения (увеличение размера бина) матрицы Hi-C становится возможным благодаря высокой корреляции частот взаимодействия в соседних бинах матрицы Hi-C. Такая процедура увеличивает количество картированных пар прочтений на каждый бин за счёт снижения количества бинов и присваивания каждому усредненных значений прежних соседних бинов. Эта процедура позволяет оценивать частоты взаимодействия большинства регионов друг с другом, что перевешивает потери деталей карты от снижения разрешения и усреднения значений [190].

Появление незаполненных бинов на картах Hi-C обусловлено недостаточным количеством или полным отсутствием сайтов рестрикции внутри данных бинов. Эти ограничения привели к модификациям протокола Hi-C, включающим использование других эндонуклеаз, что, по сравнению с оригинальным протоколом [191], позволяет делать более частые разрезы на ДНК. Например, в модификации протокола Hi-C 3.0 используют DpnII и DdeI [192], что позволяет разрезать хроматин чаще, чем этими эндонуклеазами по отдельности. Известен также метод Micro-C, в котором использована микрококковая нуклеаза [149; 193], которая не имеет специфических сайтов рестрикции, но способна нарезать хроматин на монопонуклеомосные фрагменты. Такие модификации протоколов позволяет снизить размер фраг-

ментов и достичь более высокого разрешения при достаточной глубине секвенирования [192; 194—196].

Важно отметить, что различные структуры хроматина рекомендуют искать при различных разрешениях. Например, поиск петель обычно проводится при разрешении 1–10 тыс. п.н., тогда как для топологически ассоциированных доменов (ТАДы) и компартментов рекомендуют разрешения 50–100 тыс. п.н. и 100–1000 тыс. п.н. соответственно [197]. Однако оптимальное разрешение для конкретного исследования остается нерешенным вопросом, и выбор может зависеть от качества экспериментальных данных. Данные Micro-C часто используются для изучения хроматиновых петель, поскольку, в сравнении с Hi-C, у данных Micro-C выше соотношение сигнала к шуму на близких расстояниях. Но известно также, что Micro-C требует тщательного подбора концентрации микрококковой нуклеазы, чтобы обеспечить достаточное переваривание как эухроматина, так и гетерохроматина, иначе поиск компартментов этим методом может значительно ухудшиться [198]. Поэтому отдельные работы рекомендуют для численной оценки компартментализации использовать классический Hi-C [192]. Разрешение матрицы Hi-C также может быть установлено в зависимости от расстояния между интересующими локусами, что создает гибкий подход к оптимизации выбора разрешения в зависимости от конкретных исследовательских задач [199].

В работе [200] проведена оценка корреляции между репликами Hi-C, полученными с использованием одних и тех же или различных эндонуклеаз рестрикции, при нормализации различными алгоритмами. Как и ожидалось, корреляции реплик, полученных с использованием различных эндонуклеаз рестрикции, были ниже, независимо от применяемого алгоритма нормализации, по сравнению с картами, полученными с использованием одних и тех же эндонуклеаз рестрикции. Воспроизводимость быстро снижается с увеличением разрешения (от 100 тыс. п.н. до 20 тыс. п.н.) и расстояния (от 2,5 до 10 млн. п.н.) между взаимодействующими локусами. Для надежного обнаружения дальних взаимодействий требуется высокая глубина секвенирования, превосходящая доступные данные на момент написания статьи [200].

Предполагается, что более высокое разрешение может быть достигнуто вблизи диагонали матрицы, где глубина секвенирования может создать компромисс между высоким разрешением и геномным расстоянием. Эта идея легла в основу подходов динамического бинирования, которые адаптиру-

ют размер бина локально в зависимости от количества сайтов рестрикции, обеспечивая равномерное покрытие по всему геному и локально повышенное разрешение [201]. Инструмент HIPPIE и алгоритм Binless также отказываются от фиксированного размера бина, работая на уровне рестриционных фрагментов [202; 203]. Алгоритм 2D fused lasso, используемый в Binless, гарантирует оптимальный размер ячейки в каждой позиции, избегая усреднения признаков и избыточного увеличения шума [203] (подробнее об алгоритме Binless в разделе 1.6.2).

1.6.2 Методы нормализации данных Hi-C

Hi-C, аналогично любому другому методу, подвержен ошибкам, вызванным преимущественно различной плотностью сайтов рестрикции или ошибками секвенирования [204]. Для оценки последней принято использовать эффективную длину генома — общую длину геномных регионов, расположенных в пределах определенного расстояния (обычно 500 п.н.) от сайта рестрикции [189]. Исследования показали [204], что именно эффективная длина является важнейшим фактором ошибок, а GC-состав занимает второе место, поскольку отрицательно коррелирует с количеством контактов в картах, полученных с помощью эндонуклеазы рестрикции HindIII, и положительно коррелирует с количеством контактов в картах, полученных с помощью эндонуклеазы рестрикции NcoI.

Для нормализации данных Hi-C были предложены различные подходы [205], которые в целом можно разделить на [206]:

1. Подходы нормирования неявных смещений основаны на теоретическом предположении об одинаковой видимости всех локусов независимо от их линейного расстояния друг от друга. Алгоритмы вычисляют смещения, которые следует вычесть для восстановления нормализованных матриц Hi-C. Такие методы не исправляют ошибки, связанные с расстоянием или выбором размера продуктов лигирования во время подготовки библиотеки. Неявные подходы включают в себя методы балансировки матриц, пространственное преобразование и локальную регрессию с ограниченным

количеством параметров и небольшим объемом дополнительной информации.

2. Подходы нормирования явных смещений предполагают, что все факторы, влияющие на данные Hi-C, известны и могут быть предоставлены в качестве входных данных для инструментов нормализации. Явные подходы обычно используют вероятностные модели, и их реализация зависит от различной дополнительной информации, такой как сайты рестрикции, GC-состав генома и показатель картируемости (mappability), который показывает, сколько раз прочтение, исходно полученное с этого региона, может быть уникально картировано в геноме.

Важно отметить, что результаты применения любых подходов часто зависят не только от качества данных, но и от глубины секвенирования, определяющей разрешение матрицы Hi-C. Независимо от используемого метода важно визуально оценить данные до и после нормализации, чтобы определить успешность процедуры [206]. Успешная фильтрация и коррекция смещения сгладят матрицу контактов таким образом, что не останется бинов с явно высокими значениями частот контактов, по сравнению в остальными бинами.

На данный момент наиболее популярным методом нормализации неявных смещений является алгоритм итеративной коррекции [207], который обеспечивает равномерный профиль покрытия и одинаковую видимость каждого региона. Процедура основана на предположении, что, поскольку мы беспристрастно изучаем все пространство взаимодействия, каждый фрагмент или бин должен наблюдаться примерно одинаковое количество раз в эксперименте (интерпретируется как сумма полногеномной строки или столбца в матрице взаимодействия). Отдельные работы [208] отмечают различие в количестве контактов для эу- и гетерохроматина, что, по мнению авторов работы, может свидетельствовать о биологической значимости этих различий.

Алгоритм итеративной коррекции (или балансировки) выполняет следующие шаги до сходимости: сначала каждая строка делится на свое среднее значение, затем каждый столбец делится на его среднее значение. Итоговая частота контактов E_{ij} для каждой пары областей (i, j) может быть рассчитана как:

$$E_{ij} = B_i B_j T_{ij}, \quad (1.1)$$

где B_i и B_j — веса, а T_{ij} — искомая матрица относительных вероятностей контакта.

По умолчанию строки и столбцы матрицы нормализуются так, чтобы сумма весов равнялась единице. Регионам низкого качества или некартируемым областям присваиваются нулевые веса, что ведет к появлению «белых полос» или «пустых бинов» из-за невозможности доверять сигналу, полученному в таких регионах. Для нормирования полученных матриц контактов использовалась команда `balance_cooler` из библиотеки `cooler`.

Геномные локусы, которые линейно удалены друг от друга на ДНК, демонстрируют меньшую частоту взаимодействий на картах Hi-C по сравнению с близкими локусами. Для коррекции этого эффекта применяют нормализацию на линейное расстояние между контактирующими локусами. Один из подходов, известный как нормализация на ожидаемое и представленный в работе [133], заключается в делении частоты контактов в матрице Hi-C на среднюю частоту контактов по всему геному для локусов на данном линейном расстоянии:

$$x'_{i,j} = \frac{x_{i,j}}{x_{L_{i,j}}}, \quad (1.2)$$

где $x_{i,j}$ — частота контактов, i — номер строки, j — номер столбца, $L_{i,j}$ — расстояние между локусами.

Другой метод включает в себя z-трансформацию матрицы контактов. В этом случае взаимодействия нормализуются на линейные расстояния с использованием скорректированного z-показателя, который учитывает среднее количество прочтений Hi-C для всех взаимодействий на заданном расстоянии d и соответствующее стандартное отклонение. Таким образом, z-показатель взаимодействия между ячейками i и j контактной матрицы Hi-C (z_{ij}) задается следующим уравнением:

$$z_{ij} = \frac{y_{ij} - m(d)}{\sigma(d)}, \quad (1.3)$$

где y_{ij} — количество взаимодействий между ячейками i и j , $m(d)$ — среднее количество взаимодействий для расстояния $d = |j - i|$, а $\sigma(d)$ — оценка стандартного отклонения среднего.

К подходам нормирования явных смещений относится, например, `HiCNorm` [204]. Этот метод требует входных данных о GC-составе и длине фрагмента, а затем использует распределение Пуассона или отрицательное биномиальное распределение для нормализации на эффект множественного картирования. В работе [200] показано, что поправка только на эффективную длину может одновременно корректировать сдвиги, вызванные GC-составом и эффектом множественного картирования.

Скорректированная частота контактов y_{ij} между элементами матрицы Hi-C для бинов i и j выражается следующим образом:

$$y_{ij} = \frac{x_{ij}}{e_i \cdot e_j \cdot N}, \quad (1.4)$$

где x_{ij} — исходное количество взаимодействий между бинами i и j , e_i и e_j — эффективные длины для бинов i и j , а N — общее количество пар парно-концевых прочтений. Для каждого бина и каждого разрешения вычисляются эффективные длины, как описано в [204].

Важно также отметить применение отрицательного биномиального распределения при нормализации данных Hi-C. Это распределение успешно использовалось в анализе дифференциальной экспрессии [209] и подтвердило свою применимость к данным Hi-C [204]. Существуют подходы, основанные на вероятностном моделировании и устранении смещений [189], а также основанные на пуассоновском или отрицательном биномиальном распределении [204]. Эти методы подходят для клеток с нормальным кареотипом. Однако существуют методы для коррекции смещений у клеток с хромосомными аномалиями. Например, метод итеративной коррекции смещения с поправкой на хромосомы `saICB` [210] применяет единую коррекцию числа копий для всей хромосомы, исключая случаи множественной частичной анеуплоидии и региональных вариаций числа копий. Стоит также отметить метод `OneD` [211], который корректирует локальные хромосомные аномалии в экспериментах Hi-C. Этот метод моделирует влияние известных смещений с использованием обобщенной аддитивной модели и отрицательного биномиального распределения. В работе [212] был предложен алгоритм `HiCorr` для коррекции известных источников ошибок (картируемость, GC-состав) и неизвестных источников ошибок (неравная видимость контактов двух локусов, находящихся на различном расстоянии) в картах Hi-C.

Каждый рестрикционный фрагмент имеет два конца, которые могут образовывать лигирующие соединения с другими фрагментами. Два конца одного и того же фрагмента могут иметь разные значения локальной картируемости и GC-состава. В связи с этим в работе [212] предложено анализировать два конца фрагмента независимо. HiCorr оценивает ожидаемые значения для каждой пары фрагментов и использует наблюдаемые отношения к ожидаемым для определения взаимодействий хроматина. Если два конца i и j принадлежат одному и тому же рестрикционному фрагменту, эти концы будут иметь одинаковые параметры длины и расстояния, но разные параметры GC-состава и картируемости. Целью этого шага нормализации является оценка $\mu_{i,j}$ — ожидаемого количества контактов между двумя концами i и j , которая корректирует известные и неизвестные источники смещения в Hi-C:

$$\mu_{i,j} = m_i m_j F_{i,j}^{gc} L_{i,j} V_i V_j, \quad (1.5)$$

где m_i и m_j — картируемость конца i и j соответственно, $F_{i,j}^{gc}$ — поправочный коэффициент для смещения GC-состава, $L_{i,j}$ — ожидаемая частота цис-контактов между концом i и j , если бы оба конца были на 100 % картируемы. Явная коррекция факторов m_i , m_j , $F_{i,j}^{gc}$, $L_{i,j}$ аналогична описанной ранее [213]. В формуле вводятся два дополнительных фактора, V_i и V_j , для равной видимости двух концов, что неявно корректирует неизвестные источники смещения. Для вычисления видимости используются только транс-контакты, так как нормализация по цис-контактам связана с риском чрезмерной коррекции из-за многих цис-прочтений, происходящих из петель хроматина. После коррекции явных смещений предполагается, что неявные факторы видимости V_i и V_j являются дополнительными независимыми источниками смещения, не зависящими друг от друга. Биологически V_i и V_j можно интерпретировать как концентрацию двух концов в протоколе Hi-C. Например, сайты HindIII в открытом хроматине с большей вероятностью будут расщеплены эндонуклеазой рестрикции.

1.6.3 Методы оценки воспроизводимости и сходства данных Hi-C

В ранних работах, посвященных Hi-C для измерения сходства между экспериментами Hi-C, использовались коэффициенты корреляции Пирсона и Спирмена [204; 214]. Однако стандартные подходы расчета корреляций неявно рассматривают все элементы матрицы Hi-C как независимые измерения, не учитывая пространственные паттерны в карте контактов Hi-C, такие как компартменты или ТADы, а также убывающую зависимость частоты контактов от линейного расстояния между локусами. Потому для двух похожих карт можно получить низкие значения корреляции Спирмена из-за стохастической вариации силы точечных взаимодействий и наоборот, возможно получение высоких корреляций между неродственными образцами [215].

Чтобы преодолеть эту проблему было разработано несколько методов для оценки воспроизводимости анализа Hi-C в популяционных данных, включая HiCRep [215], HiC-spector [216], Genome-DISCO [217] и QuASAR [218]. Эти методы были специально разработаны с учетом биологически значимой структуры в наборе данных Hi-C при одновременном уменьшении неважных для глобальной оценки точечного сходства. Они дают более точную оценку воспроизводимости, чем корреляции Пирсона или Спирмена для реальных и смоделированных наборов данных. Их также можно использовать для оценки попарного сходства между различными образцами Hi-C.

HiCRep [215] учитывает пространственные структуры данных Hi-C при оценке сходства. Он корректирует эффект геномного расстояния и решает проблему разреженности матриц контактов посредством стратификации и сглаживания соответственно.

Подсчет коэффициента корреляции с поправкой на страты (SCC), т.е. несколько диагоналей матрицы Hi-C, соответствующие взаимодействиям между локусами в узком диапазоне расстояний [215], зарекомендовал себя как надежный метод оценки сходства между картами Hi-C, поскольку, в отличие от применявшихся ранее коэффициентов корреляции Спирмена и Пирсона, SCC учитывает доменные структуры и влияние расстояния между локусами на их «видимость» (см. раздел 1.6.3).

Метод состоит из двух шагов. На первом этапе HiCRep сглаживает каждую карту контактов Hi-C, используя двумерный средний фильтр, который заменяет число прочтений каждого контакта на карте контактов средним числом всех контактов в его окрестности. Процедура сглаживания необходима, чтобы уменьшить локальный шум в карте контактов и сделать доменные структуры более заметными. В более поздних работах предложены другие процедуры сглаживания матриц Hi-C, что позволяет добиться улучшения воспроизводимости между биологическими репликами [200; 203]. На втором этапе HiCRep разделяет сглаженную матрицу контактов Hi-C на страты. Затем HiCRep вычисляет корреляцию Пирсона для сглаженных контактов в каждой страте. Такая процедура позволяет избежать смещения, вносимого расстоянием между локусами, в оценку сходства данных региона между разными картами. Затем происходит расчет коэффициента корреляции с поправкой на страту (SCC), который представляет собой средневзвешенное значение корреляции ρ_s , характерной для страты, по всем стратам:

$$\rho_s = \sum_k w_k \rho_k, \quad (1.6)$$

где ρ_k — коэффициент корреляции в каждой страте, w_k — веса, рассчитываемые на основе статистики Кохрана-Мантеля-Хензеля (СМН). SCC находится в диапазоне $[-1, 1]$ и интерпретируется аналогично стандартному коэффициенту корреляции.

HiCRep имеет несколько настраиваемых пользователем параметров: *resol* — разрешение данных Hi-C, т.е. размер бина, например, 40 тыс. п.н.; *lbr* и *ubr* — нижняя и верхняя границы геномного расстояния между локусами взаимодействия. Используя эти границы, пользователи могут ограничить оценку воспроизводимости взаимодействиями в пределах заданного диапазона геномного расстояния, *h* — размер окрестности, управляющий уровнем сглаживания, причем большее значение указывает на более высокий уровень сглаживания. Когда $h = 0$, сглаживание не применяется. Оптимальное значение *h* зависит от разрешения.

Авторами работы [215] рассчитаны ориентировочные значения *h*, представленные в Таблице 1, на основе эталонного набора данных с помощью функции `htrain` для различных разрешений. Пользователи могут использовать эти значения для своего анализа, если обучение невозможно. Опти-

мальное значение при выбранном разрешении можно получить, запустив функцию `htrain` на паре достаточно глубоко секвенированных образцов при этом разрешении.

Таблица 1 — Подбор значения h в зависимости от разрешения карты Hi-C [215].

Разрешение	Значение h
10 тыс. п.н.	20
25 тыс. п.н.	10
40 тыс. п.н.	5
100 тыс. п.н.	3
500 тыс. п.н.	1 или 2
1 млн. п.н.	0 или 1

1.7 Алгоритмы поиска петель

Существуют два класса алгоритмов поиска петель, ищущих либо глобальное обогащение частот контактов среди всех бинов матрицы Hi-C, либо локальное обогащение частот контактов над окружающим сигналом в окне. Алгоритмы, ищущие глобальное обогащение, часто основываются на статистических или вероятностных моделях и допускают, что сигнал Hi-C подчиняется распределению (часто отрицательному биномиальному). К алгоритмам, основанным на поиске глобального обогащения, относятся `Fit-Hi-C` [219], `GOTHIC` [220] и `HOMER` [221]. Алгоритмы, осуществляющие поиск локального обогащения, преимущественно реализуют процедуру, похожую на поиск пиков в ChIP-seq или ATAC-seq. К алгоритмам, основанным на поиске локального обогащения, относятся `diffHiC` [222], `SIP` [223] и `HiCCUPS` [224]. Однако наборы петель, выявленные алгоритмами разных классов, лишь частично перекрываются [197]. В работе [212] выдвигается гипотеза, что в основном это связано с помехами от сильного фонового сигнала на малом расстоянии, что выражается в виде сильного сигнала вдоль диагонали в необработанных контактных матрицах. Методы глобального

обогащения очень чувствительны к нормализации данных Hi-C, поскольку недостаточная или чрезмерная коррекция систематических ошибок Hi-C приведет к большому количеству ложноположительных или ложноотрицательных результатов соответственно. С другой стороны, альтернативное локальное обогащение лучше идентифицирует вершины дискретных пиков при низком окружающем сигнале, но теряет свою эффективность при высоком фоновом сигнале, например, на близком от диагонали расстоянии.

В работе [212] также высказывается предположение, что, по-видимому, недостаточная глубина секвенирования является основной причиной не воспроизведенных петель между репликами, что влияет на показатели работы различных алгоритмов. Поэтому авторы предлагают проводить процедуру поиска петель после сливания нескольких биологических реплик для получения наибольшей из возможных глубины прочтения. Алгоритмы поиска петель показывают совершенно разные результаты на образцах разных типов, при этом невозможно выделение четкого золотого стандарта [197].

В Таблице 2 приведены некоторые популярные алгоритмы поиска петель и дана их краткая характеристика.

Таблица 2 — Алгоритмы поиска хроматиновых петель в данных Hi-C. Адаптировано из [197] и дополнено.

Метод	Язык	Алгоритм
Fit-Hi-C	Python	NB + глобальное обогащение
GOTHIC	R	NB + глобальное обогащение
HOMER	Perl, R	NB + глобальное обогащение
HIPPIE	Python, Perl, R	Глобальное обогащение на уровне фрагментов
diffHiC	R, Python	Локальное обогащение над соседними бинами
HiCCUPS	Java	Локальное обогащение над соседними бинами, FDR
LASCA	Python	Глобальное обогащение + распределение Вейбулла
covNorm	R	Глобальное обогащение + NB
chromosight	python	компьютерное зрение

Метод	Язык	Алгоритм
SIP	Java	анализ изображений

Шесть инструментов для поиска петель хроматина при анализе данных Hi-C были рассмотрены в работе [197]. Важно отметить, что в этой работе алгоритмы были протестированы на картах с низким разрешением (от 40 КБ до 1 МБ), созданных с помощью экспериментальной процедуры Hi-C первого поколения. Более того, если в алгоритм поиска петли не был включен шаг выравнивания, был использован Bowtie для выравнивания прочтений целиком, что плохо подходит для прочтений Hi-C, большая часть которых является химерными [225]. В части инструментов, среди которых HIPPIE, HiCCUPS и diffHiC, используется химерное выравнивание, которое позволяет отображать также считывания, охватывающие соединение лигирования, что закономерно дает больший процент картирования: в среднем на 18,4 % (химерный STAR в HIPPIE), 27,4 % (химерный BWA в HiCCUPS) и 40,1 % (химерный Bowtie в diffHiC) больше прочтений, чем Bowtie. Важно также отметить, что разница в долях выравнивания между химерным выравниванием (части прочтения могут быть картированы на некотором расстоянии друг от друга, аналогично процедуре картирования RNA-seq) и выравниванием прочтений целиком становилась более очевидной по мере увеличения длины считывания, в диапазоне от 30,9 % (при 36 п.н.) до 55,4 % (при 101 п.н.) дополнительно выровненных прочтений.

Поэтому рассмотрим общие принципы работы подобных инструментов, дополнив и расширив список недавно опубликованными работами. Fit-Hi-C [219], GOTHiC [220], covNorn [226] и HOMER [221] поиск значимых взаимодействий происходит с использованием отрицательного биномиального распределения для оценки ожидаемых вероятностей контактов в зависимости от расстояния и p -значений, с коррекцией на множественные сравнения. В GOTHiC встроена нормализация с помощью квадратного корня покрытия, а HOMER проводит фильтрацию на уровне прочтений и типов их пар, а также неявную нормализацию, основанную на покрытии региона и расстоянии между локусами. Fit-Hi-C и covNorm анализирует значимость только цис- контактов, в то время как GOTHiC и HOMER работают также

с транс- контактами. Все четыре инструмента ищут петли путем анализа глобального обогащения [197].

Отдельно стоит упомянуть инструмент `HIPPIE` [202], который для поиска петель реализует подход, аналогичный представленному в [213]. Здесь, как и в алгоритмах, описанных выше, используется отрицательное биномиальное распределение, но добавлена нормализация на GC-состав, картируемость, длину фрагмента и расстояние между локусами, а параметр избыточной дисперсии является фиксированным и получен из [213]. Помимо поиска петель, `HIPPIE` обладает функционалом для химерного выравнивания, фильтрации на уровне прочтений, пар прочтений и фрагментов, а также явной нормализации без применения бинирования. Результатом является набор взаимодействий на основе фрагментов рестрикции (цис- и транс-) с соответствующим p -значениями, по которым и происходит отбор значимых взаимодействий. `HiCCUPS` [224] и `diffHiC` [222] также имеют инструменты для предобработки данных, реализуя выравнивание, чтение и фильтрацию на уровне пар прочтений. Значимые взаимодействия (цис- и транс- для `diffHiC`, только цис- для `HiCCUPS`) идентифицируются с использованием локального обогащения, т.е. поиска пар локусов, которые имеют значительно больше прочтений, чем их соседи. Важно отметить, что `diffHiC` работает с значениями обогащения для каждого взаимодействия, которые рассчитываются как логарифмированное отношение числа пар прочтений целевой пары бинов и наиболее обогащенной из соседних областей. Фильтрация значимых обогащений производится путем установления пороговых значений для обогащения, количества вспомогательных прочтений и расстояния между локусами, но при работе с отдельными картами статистический тест не выполняется, и значение значимости не возвращается. `HiCCUPS` же обнаруживает пиксели, обогащенные по отношению к соседним областям с учетом ширины пика и размера окна, и возвращает центроиды кластеров обогащенных пиков, отсеченные с помощью p -значений с поправкой Бенджамини-Хохберга.

Важно отметить, что все описанные выше инструменты были разработаны для поиска петель в картах Hi-C млекопитающих, причем на не очень высоком разрешении (10 тыс. п.н.); интерес к устройству хроматиновых петель в других организмах, а также увеличивающееся разрешение карт привело к необходимости создания новых инструментов.

Одним из таких инструментов является инструмент SIP [223]. Анализ, проведенный авторами, показывает, что этот инструмент устойчив к шуму и глубине секвенирования данных и может использоваться для обнаружения петель, которые ранее не обнаруживались в клетках человека, а также петель в других организмах, например, *C. elegans*. В основе работы инструмента лежит следующий принцип: геном анализируется с помощью скользящих окон с использованием алгоритмов обработки изображений для выявления потенциальных петель, затем используется расчет значений, нормализованных на ожидаемое (observed-over-expected, ooe), по Формуле, аналогичной 1.2 с последующей фильтрацией экстремальных значений на первых двух диагоналях:

$$x'_{i,j} = \frac{x_{i,j} - \text{avg}(\sum_{L_{|i-j|}} x)}{\text{avg}(\sum_{L_{|i-j|}} x)}, \quad (1.7)$$

где $x_{i,j}$ — частота контактов (i — номер строки, j — номер столбца), $L_{i,j}$ — расстояние для этого взаимодействия. Полученные значения используются для создания изображений, которые подаются в программу ImageJ и подвергаются там дальнейшей обработке:

1. Размытие по Гауссу для избежания признания отдельных пикселей с экстремальными значениями сигнала петлями и повышение контрастности (для более явного выделения петли);
2. Метод белого цилиндра (white top-hat) представляет собой преобразование значений интенсивности в оттенки серого для каждого изображения Hi-C, в результате чего яркие структуры имеют повышенный контраст по сравнению с фоном. В результате происходит гомогенизации фона, что упрощает для алгоритма обнаружение ярких структур;
3. Минимально-максимальный фильтр (Minimum and Maximum Filter) для удаления изолированных пикселей и дальнейшей гомогенизации фона.

Эти шаги обеспечивают получение скорректированного изображения, которое используется алгоритмом обнаружения локальных максимумов из ImageJ для обнаружения предварительного списка петель-кандидатов.

Петли-кандидаты затем должны пройти через несколько фильтров, которые используют исходный нормированный по расстоянию сигнал в данных Hi-C. Во-первых, удаляются пиксели рядом с областями, содержащими

повторы. Во-вторых, чтобы удалить отдельные пиксели, представляющие собой аномальные выбросы, введена проверка, что центральный пиксель петли имеет самые высокие значения сигнала, а соседние пиксели имеют спадающий сигнал по мере отдаления от центра петли. Наконец, петли-кандидаты фильтруются на основе эмпирического значения доли ложноположительных результатов (false discovery rate, FDR), рассчитанного как обогащение петель по сравнению со случайными участками на тех же расстояниях.

1.7.1 Выбор распределения для поиска значимых контактов

Рассмотрим подробнее использование отрицательного биномиального распределения для определения значимости контактов локусов хроматина на примере работы [212]. Было показано, что количество прочтений Hi-C $X_{i,j}$ между двумя фрагментами i и j можно смоделировать с помощью отрицательного биномиального распределения [213]:

$$X_{i,j} \sim NB(r_{i,j} = \frac{\mu_{i,j}}{\beta - 1}, \rho = \frac{\beta - 1}{\beta}), \quad (1.8)$$

где $\mu_{i,j}$ — среднее значение распределения, $\beta \cdot \mu_{i,j}$ — дисперсия распределения, а β — постоянное число, которое представляет собой значение наклона между дисперсией и средним значением, полученным в результате линейного регрессионного анализа.

Значение β необходимо подбирать для каждого набора данных отдельно. Таким образом, мы можем вычислить p -значение, используя отрицательное биномиальное распределение для любой пары фрагментов, что будет отражать значимость рассчитанного обогащения (процедура нормализации карты контактов описана в разделе 1.6.2):

$$p_{i,j} = P(X_{i,j} \geq x_{i,j} | \mu_{i,j}, \beta) \quad (1.9)$$

Однако, если использовать только p -значения для поиска петель, распределение расстояний между локусами, образующими значимые взаимодействия, сместится в сторону ближнего диапазона. Причиной этому является высокая плотность данных в ближнем диапазоне, и пиксель может достигать

статистической значимости даже при умеренном обогащении. Потому было предложено проводить поиск петель, используя коэффициенты обогащения. Однако использование порога для признания коэффициента обогащения значимым дает смещение в сторону дальних взаимодействий, потому что, когда значение $\mu_{i,j}$ очень мало, что случается часто из-за низкой плотности данных, соотношение $\frac{x_{i,j}}{\mu_{i,j}}$ может быть очень большим, но не иметь при этом статистической значимости. Потому был разработан подход использования d , фиктивной переменной:

$$e_{i,j} = \frac{x_{i,j} + d}{\mu_{i,j} + d} \quad (1.10)$$

Если значение d слишком мало, то карты контактов имеют много пикселей с высоким обогащением, которые на деле являются зашумленными пикселями на большом расстоянии. Эта проблема была решена путем корректировки значений фиктивной переменной. Для любого пикселя отношение уменьшается с увеличением d , но p -значение при этом не меняется. Поэтому, если использовать отсечение как по p -значениям, так и по порогу коэффициента обогащения, при достаточно высоком значении фиктивной переменной петель будет найдено меньше.

Авторы подобрали минимальное значение фиктивной переменной, чтобы каждый пиксель, прошедший двойное отсечение, имел p -значение $< 0,001$ для некоторых из проанализированных ими типов данных:

Таблица 3 — Подбор значений переменной для поиска дифференциально контактирующих локусов [212].

Данные	Значение d
H1 hESC	6
IMR90, зародышевая зона плода, подкорковая пластинка плода и кора БП взрослого человека	10
GM12878	7
hiPSC, hNPC, hNeuron и кора БП плода	13

В работе [227] было показано, что распределение Вейбулла в сравнении с нормальным, лог-нормальным и гамма-распределением подходит

для описания распределения частоты контактов в Hi-C. Другое исследование применимости этого распределения к Hi-C было проведено в работе [228], где отмечают, что распределение Вейбулла хорошо описывают частоты контактов на разных расстояниях. В этой же работе был предложен алгоритм поиска петель, основанный на распределении Вейбулла, которому подчиняются частоты контактов каждой диагонали. Процедура поиска петель состоит из подбора модели, основанной на распределении Вейбулла, для каждой диагонали матрицы Hi-C, и расчета p -значений с коррекцией Бенджамина–Хохберга (q -значения) для каждого пикселя как вероятности нахождения соответствующего пикселя из модели с такой же или большей интенсивностью. Смежные значимые пиксели группируются в кластеры; для каждого кластера определяется центр, и его координаты извлекаются и рассматриваются как координаты петли. Более того, пользователям предлагаются дополнительные фильтры, например, коррекция на ожидаемые значения каждой диагонали, интенсивности сигнала и обогащению сигнала по сравнению со случайными сигналами, расположенными на том же расстоянии от центральной диагонали. Анализ, проведенный в самой работе, показал, что использование этих фильтров на данных Hi-C млекопитающих повысит точность аннотации петель, тогда как использование фильтров с геномом дрожжей, по-видимому, менее полезно. Успешная работа алгоритма была показана на *H. sapiens* и *C. elegans*, причем для человека были аннотированы также CTCF-независимые петли. Выдача алгоритма демонстрирует хорошее пересечение с набором петель, размеченных HiCCUPS (65%) и MUSSTACHE, а также идентифицирует гораздо больше пикселей на энхансерах и промоторах, что хорошо совпадает с результатами визуальной оценки.

1.8 Визуализация матриц контактов

К настоящему времени реализованы различные методы для визуализации данных, к ним относятся, например, представления контактных матриц с помощью тепловых карт, дуговые представления дистальных контактов и

подходы на основе сетевых графов. Таблица 4 отражает текущие наиболее употребляемые инструменты для визуализации данных Hi-C:

Таблица 4 — Инструменты для визуализации данных Hi-C [229].

Инструмент	Язык	Ссылка	Матрица контак- тов	Интерактивный интер- фейс	Создание аннотации
HiGlass	JS, python	[230]	классическая	+	+
HiCExplorer	python	[231]	повернутая	-	+
HiCBrowser	python	[232]	повернутая	+	+
cooltools	python	[233]	классическая	-	-

Среди этих решений тепловые карты стали стандартом де-факто для визуализации карт контактов Hi-C. При использовании тепловых карт данные Hi-C можно визуализировать в виде квадратной сетки частот контактов, сопоставленных с цветовым градиентом, а взаимодействующие геномные координаты отображаются на осях x и y . При представлении внутривромосомных (цис-) взаимодействий это представление тепловой карты приводит к получению симметричного графика вокруг диагонали, т.е. содержит избыточную информацию (классическое отображение). Общепринятая альтернатива, позволяющая избежать такой избыточности, основана на отображении только верхнего треугольника контактной матрицы путем разрезания по главной диагонали и поворота графика таким образом, чтобы ось x содержала геномные координаты, а ось y — расстояние между геномными локусами.

Поскольку данные Hi-C имеют широкий динамический диапазон, для отображения используется логарифмическая шкала, что позволяет одновременно визуализировать объекты вблизи и вдали от диагонали при корректном подборе цветовой гаммы. Касательно цветовой шкалы, создатели библиотеки `cooltools` рекомендуют использовать непрерывную цветовую шкалу «fall» (вдохновленную `colorbrewer`) или другую, представленную в библиотеке `cooltools.lib.plotting`, каждая из которых предлагает широкий динамический диапазон, необходимый для визуализации матриц Hi-C, что позволяет отображать объекты более четко, чем классические красно-белые цветовые шкалы.

1.9 Зависимость частоты контактов от расстояния

На картах Hi-C частота контактов значительно снижается с увеличивающимся геномным расстоянием. Для оценки этой зависимости часто используют график скейлинга, который представляет собой зависимость средней частоты контактов от линейного расстояния между локусами. Согласно опубликованным моделям [234], хроматин большинства организмов представляет собой фрактальную глобулу, что описывается уравнением зависимости частоты контактов от расстояния:

$$P_c(s) = s^{-3/2}, \quad (1.11)$$

где s — линейное расстояние между локусами.

В области анализа Hi-C эту линейную зависимость часто называют «ожидаемой», потому что ожидается определенная средняя частота контактов при данном линейном расстоянии; или *scaling*, что заимствовано из литературы по физике полимеров. Скорость снижения частоты контактов при увеличивающемся геномном расстоянии отражает полимерную природу хромосом и может указывать на глобальные модели укладки генома. Например, отклонения от зависимости, описанной в (1.11) могут свидетельствовать о наличии петель по всему геному или о теломер-теломерных взаимодействиях. Более того, подобный график используется как один из этапов контроля качества данных, а также при сравнении глобальных изменений укладки хроматина между двумя и более стадиями или состояниями. Для облегчения интерпретации также принято строить график первой производной частоты контактов для поиска отклонений от ожидаемого. Обе оси преобразованы в логарифмические координаты.

1.10 Инсуляция

Метрика «показатель инсуляции» была впервые предложена в работе [235]. Чтобы рассчитать показатель инсуляции (или «изоляции») для каждого

бина данных Hi-C, необходимо рассчитать среднее количество взаимодействий, которые произошли в каждом бине. Это можно визуализировать, продвигая квадрат заданного размера (равен размеру окна) вдоль диагонали матрицы и агрегируя весь сигнал внутри квадрата. Затем рассчитывается средний сигнал в каждом из квадратов, и присваивается бину на диагонали. Процедура повторяется для всех бинов на диагонали.

Формула для вычисления показателя инсуляции бина i записывается как

$$IS = \frac{\sum_{m \in M, n \in N} C(m, n)}{k}, \quad (1.12)$$

где M — множество бинов слева, а N — справа от бина i на расстоянии, не превышающем L ; k — мощность множеств M и N ; $C(m, n)$ — количество хроматиновых контактов между парой бинов m и n .

Важно также отметить, что для любых бинов в пределах размера окна от начала или конца матрицы показатель инсуляции не присваивается, поскольку квадрат подсчета показателя инсуляции выходит за границы матрицы. Затем показатель инсуляции нормализуется относительно среднего значения по каждой хромосоме путем вычисления отношения \log_2 -показателей инсуляции каждого бина и среднего значения всех показателей инсуляции. Долины или минимумы вдоль нормализованного вектора показателей инсуляции представляют собой участки уменьшенных частот взаимодействий на Hi-C, которые видны по всему бину. Эти позиции часто называют границами, и расчет минимумов в значениях профилей инсуляции лежит в основе некоторых алгоритмов детекции ТАДов и петель.

Глава 2. Методы исследования

2.1 Протокол Hi-C

Hi-C проводили по протоколу 2.0 [194]: хроматин разрезали рестриктазой DpnII, концы ДНК заполняли с помощью фрагмента Клёнова с использованием биотинилированных нуклеотидов, лигировали *in situ*, после чего выделяли ДНК, фрагментировали её ультразвуком, обогащали фракцию продуктов лигирования с использованием магнитных частиц, конъюгированных со стрептавидином. Полученные фрагменты ДНК использовали для приготовления библиотеки для секвенирования с адаптерами Illumina TruSeq в двух независимых биологических повторностях. Библиотеки секвенировали на платформе Illumina HiSeq 4000 с глубиной чтения около 100 млн парных ридов на библиотеку.

2.2 Обработка результатов секвенирования библиотек Hi-C

Парные прочтения Hi-C обрабатывались с помощью конвейера `distiller-nf` версии 0.3.3 с параметрами по умолчанию. Кратко, прочтения были картированы на референсный геном *D. discoideum* версии 2.7 с помощью `bwa mem` (0.7.17-r1188), затем обработаны в пары выравненных прочтений с помощью `pairtools` (0.3.1-dev.1). Только те прочтения, которые были картированы с MAPQ не менее 30, считались потенциальными парами контактов. Результаты множественного картирования, односторонне картированные чтения и потенциальные дубликаты ПЦР были отфильтрованы. Полученные карты контактов были разбиты на бины размером 100 п.н. с помощью `cooler` (0.8.7). Затем разрешение было установлено на 2000 п.н. Карты были сэмплированы до одинакового количества выровненных пар прочтений (~46 млн.) с помощью модуля `sample` инструмента `cooltools` и сбалансированы при помощи `cooler balance` [207].

Как было показано ранее в работе [212], увеличение глубины секвенирования ведёт к более аккуратному детектированию петель и других структур хроматина, поэтому биологические реплики были слиты.

2.3 Заполнение пропущенных значений для улучшения работы алгоритмов поиска петель

Чтобы получить лучшие результаты при запуске алгоритмов поиска петель, было применено сглаживание с помощью алгоритма `adaptive_coarsegrain` из библиотеки `cooltools.lib.numutils` (0.4.1), который адаптивно сглаживает скорректированные матрицы на основе количества контактов в ненормализованных матрицах. Пустые бины были заполнены функциями `interp_nan` из библиотеки `cooltools.lib.numutils` (0.4.1), поскольку алгоритм `chromosight`, использованный для поиска петель, не допускает нахождения петель рядом с пустыми бинами, а алгоритм `LASCA` находит избыточное количество ложных вхождений вокруг пустых бинов.

2.4 Расчет показателей инсуляции

Для анализа показателей инсуляции на картах Hi-C *D. discoideum* были использованы несглаженные и незаполненные данные Hi-C для избежания получения неточных результатов. Был применен алгоритм `insulation` из библиотеки `cooltools` с разной комбинацией размеров окон (10 тыс. п.н., 20 тыс. п.н., 50 тыс. п.н. и 100 тыс. п.н.) и размеров бина (1.5 тыс. п.н., 2 тыс. п.н., 4 тыс. п.н.), оптимальной была признана комбинация окна 20 тыс. п.н. и размера бина равная 2 тыс. п.н.

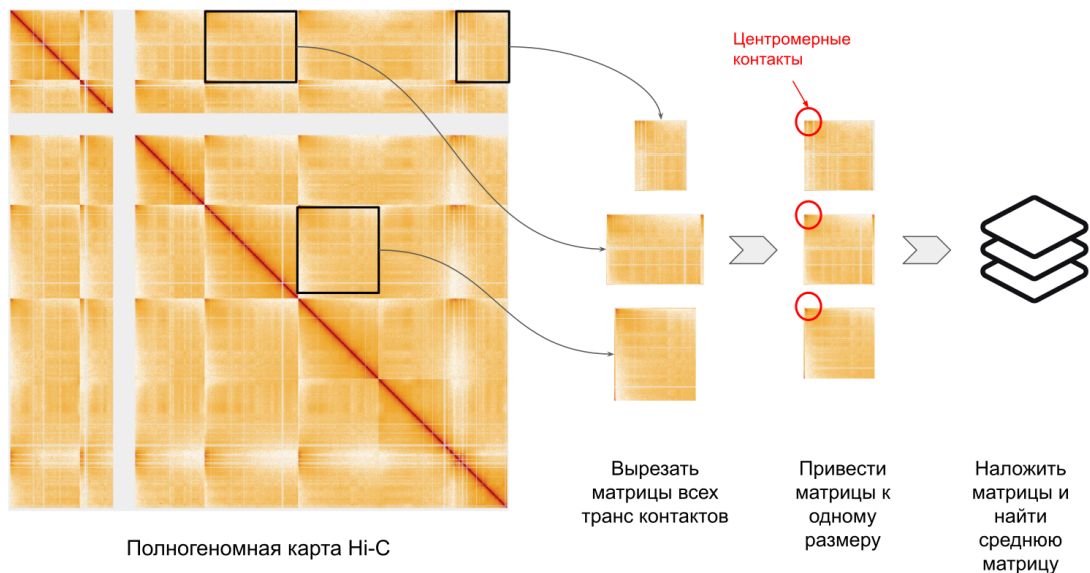


Рисунок 3 — Схема анализа силы межтеломерных и межцентромерных взаимодействий на карте средних межхромосомных взаимодействий.

2.5 Анализ межхромосомных взаимодействий

Средняя карта межхромосомных взаимодействия представляет средние частоты межхромосомных взаимодействий и выделяет хроматиновые контакты, обогащенные из-за структуры Рабля хромосом организма. Для создания графика отфильтрованные карты Hi-C с разрешением 100 КБ были подвергнуты итеративной процедуре коррекции инструментом `cooler` (v0.8.6) с помощью команды `cooler balance --trans-only --min_nnz 300`. Подбор параметров (Рисунок 4) осуществлялся сравнительный анализом графика среднего межхромосомного взаимодействия с разными параметрами. Матрицы межхромосомных контактов из полученных карт извлекали и поворачивали так, чтобы центромера оказалась вблизи левого верхнего угла (Рисунок 3). Затем каждая повернутая матрица была нормализована по ее среднему значению (значения NaN игнорируются) и масштабирована до общего размера 30x30 с помощью функции `cooltools.lib.numutils.zoom_array()` из пакета `cooltools` (v0.4.1) с параметрами `same_sum=False, order=1`. Полученные матрицы транс-взаимодействий накладывались друг на друга, вычислялось среднее значение для каждого пикселя (значения NaN игнорировались), и окончательная матрица преобразовывалась с помощью логарифма с основанием 2.

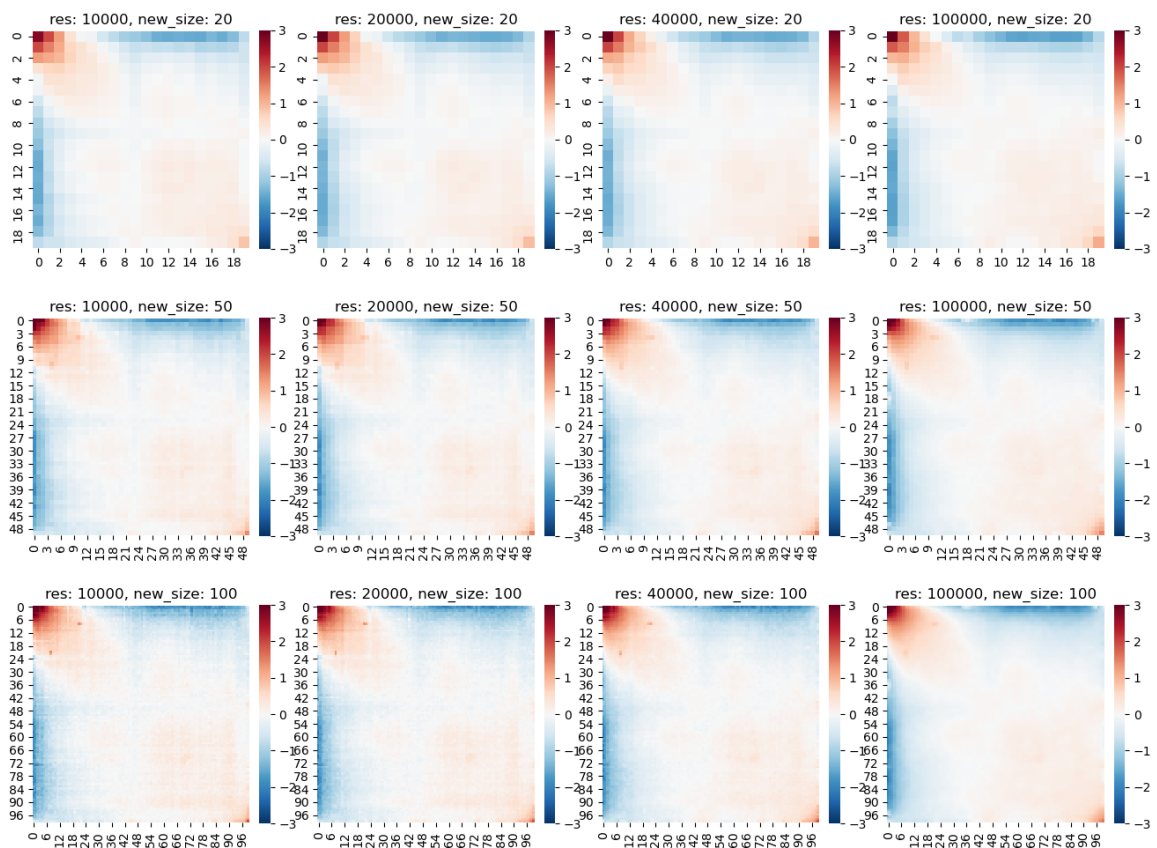


Рисунок 4 — Подбор параметров для построения карты средних межхромосомных контактов *D. discoideum*. Столбцы: разрешения карт 10 тыс. п.н., 20 тыс. п.н., 40 тыс. п.н., 100 тыс. п.н. Строки: размере финальной карты, 20x20 бинов, 50x50 бинов, 100x100 бинов.

Карты межхромосомных контактов для *S. cerevisiae* и *H. dujardini* были дополнительно «разрезаны» по центромерам (Рисунок 5) и ориентированы так, чтобы в левом верхнем углу были межцентромерные взаимодействия, а в нижнем правом углу — межтеломерные.

2.6 Подсчет корреляций между картами Hi-C

Подсчет корреляций с поправкой на страту (stratum-adjusted correlation, SCC) между картами Hi-C был рассчитан с помощью инструмента HiCRep. Детали и обоснование метода описаны в разделе 1.6.3.

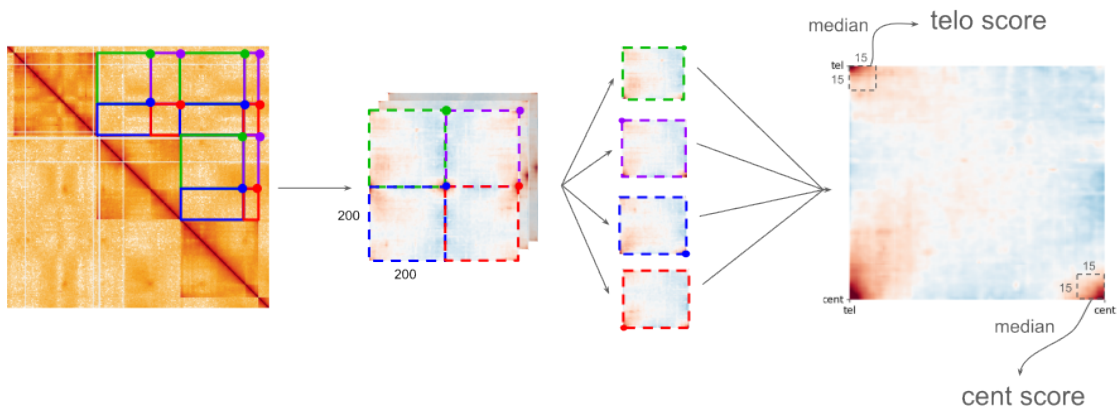


Рисунок 5 — Процедура построения средней карты межхромосомных взаимодействий для метацентрических хромосом на примере *H. dujardini*.

2.7 Сравнительный анализ упаковки хроматина между разными стадиями развития

Матрицы частот контактов для каждой хромосомы на 0 часе развития были нормализованы на среднюю частоту контактов на данном линейном расстоянии (матрица *observed-over-expected*) и приведены к одному размеру. Матрицы для каждой хромосомы на 8 часе развития были преобразованы аналогично. Затем полученные матрицы 0 часа развития были поделены на матрицу 8 часа развития для соответствующей хромосомы. Полученные матрицы отношений частот контактов были усреднены и отрисованы в логарифмической шкале.

2.8 Аннотация дальних контактов

Для изучения дальних контактов хроматина *D. discoideum* была проведена ручная аннотация на разрешении 4 тыс. п.н. с использованием HiGlass [230]. Некоторые контакты имели протяженные границы, т.е. длиной более 1 бина.

2.9 Расчет силы дальних взаимодействий

Сила петель была рассчитана как сумма частот контактов в окне 3x3 на разрешении 20 тыс. п.н., преобразованных с помощью формулы:

$$S(x) = \log_{10}(x), \quad (2.1)$$

где S — сила петли, x — петля.

2.10 Анализ вклада длинных некодирующих РНК в формирование дальних контактов

Для изучения связи длинных некодирующих РНК и оснований дальних контактов была использована разметка длинных некодирующих РНК, аннотированных в работе [236], и подсчитали, сколько их них пересекаются с основаниями дальних контактов. Для проверки статистической значимости гипотезы о перепредставленности длинных некодирующих РНК в основаниях дальних контактов был проведен пермутационный тест, который заключается в выборе случайных последовательностей тех же медианных длин и того же количества 1000 раз, что и исходная выборка, а затем подсчета количества пересекающихся длинных некодирующих РНК. P -значение рассчитывается как доля экспериментов, в которых искомое значение было больше (или соответственно меньше, в зависимости от проверяемой гипотезы), чем значение в реальном эксперименте.

2.11 Анализ экспрессии в основаниях дальних контактов

Для анализа экспрессии длинных некодирующих РНК в основаниях дальних контактов были использована разметка длинных некодирующих РНК из работы [236] и рассчитана их экспрессия в ТРМ, основанных на

полученных в этой работе данных. В качестве контроля были взяты случайные длинные некодирующие РНК в том же количестве. Процедура была повторена 1000 раз, для каждого набора подсчитана медиана ТРМ значений экспрессии выбранных генов. Было также подсчитано изменение значений экспрессии как разница ТРМ значений между 0 и 8 часами развития.

2.12 Анализ хроматиновых компартментов

Для анализа разделения регионов, принадлежащих разным компартментам был использован показатель силы компартментов, который рассчитывается как отношение сумм частот взаимодействий между регионами, принадлежащими одному компартменту, к сумме частот взаимодействий между регионами, принадлежащими разным компартментам. По оси x откладывается размер окна, для которого была подсчитана сила компартментов (Рисунок 6).

2.13 Аннотация точечных взаимодействий

Для аннотации точечных взаимодействий (петель) были использованы инструменты LASCA [238] и chromosight [239]. Для запуска LASCA были протестированы несколько наборов параметров (Рисунок 7) и оптимальными выбраны следующие параметры: `dist=20`, `size=3`, `intensity q-value=0.2`. Инструмент chromosight был запущен со следующими параметрами: `--min-dist 6000 --max-dist 45000 --pearson 0.2`. Финальной разметкой были выбраны петли, размеченные с помощью обоих инструментов одновременно; такой подход позволил добиться оптимального сигнала средней петли и снижения ложных находок.

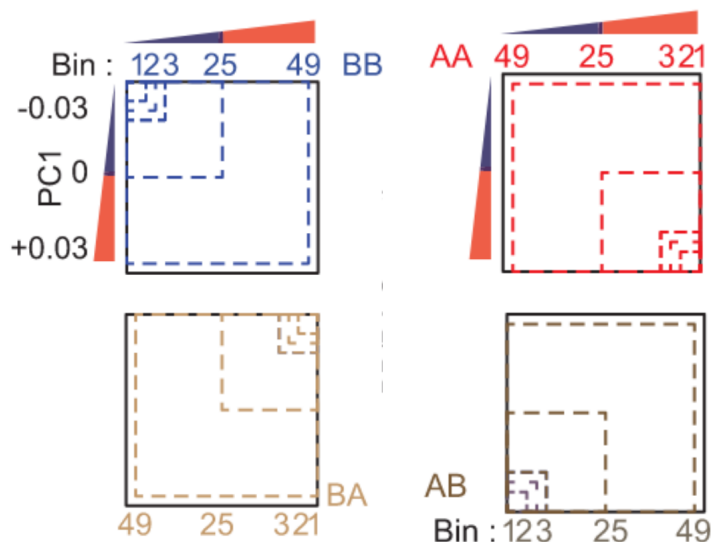


Рисунок 6 — Схема расчета силы компартментов. Верхняя панель: расчет частот контактов внутри неактивного компартмента (слева) и внутри активного компартмента (справа). Нижняя панель : расчет частот контактов между бинами, принадлежащим разным компартментам. Пунктиром обозначен размер окна, для которого рассчитана частота контактов. Выше (для верхней панели) или ниже (для нижней панели) обозначены размеры окон, выраженные количеством бинов из седловой диаграммы. Иллюстрация адаптирована из работы [237].

2.14 Аннотация хроматиновых фонтанов

Как и `chromosight` [240], `fontanka` использует маску, заданную пользователем (Рисунок 50а) для оценки сходства участков хроматина заданного пользователем размера («окно») и заданной маски — сила фонтана. но в `fontanka` добавлены дополнительные опции фильтрации, например, индекс Шарра, что позволяет получить набор наиболее достоверных участков.

Сила фонтана (`fountain score`) рассчитана как коэффициент корреляции между маской и участком матрицы Hi-C.

Для того, чтобы получить индекса Шарра, был рассчитан градиент с использованием оператора Шарра для каждого окна с помощью функции `signal.convolve2d` из пакета `scipy`, а затем взят максимум от абсолютных значений градиента в каждом окне независимо.

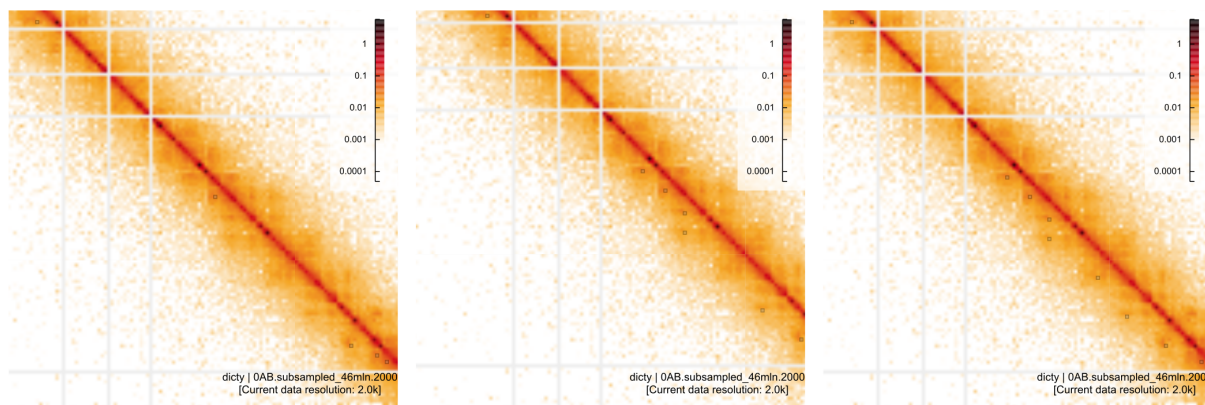


Рисунок 7 — Выбор оптимальных параметров запуска LASCА на карте хроматиновых контактов 0 часа развития *D. discoideum*. Левая панель: q -значение равное 0,05, центральная панель: q -значение равное 0,1, правая панель: q -значение равное 0,2. Черными квадратами обозначена разметка петель при заданных параметрах.

2.15 Гены домашнего хозяйства

Гены домашнего хозяйства были определены с использованием данных из работы [241] в соответствии с процедурой, описанной в работе [242] с некоторой модификацией, специфичной для *D. discoideum*: (i) экспрессия наблюдалась на всех стадиях развития; (ii) низкая дисперсия во времени: стандартное отклонение $\log_2(RPKM) < 1$; (iii) отсутствие исключительных проявлений в любой момент времени; то есть ни одно значение логарифмического выражения не отличалось от среднего значения экспрессии более чем в четыре раза. Финальный список содержал 1138 генов.

2.16 Анализ данных АТАС-seq

Сырые данные АТАС-seq были взяты из работы [243] (GSE137604). Данные обрабатывались с использованием конвейера nf-core для АТАС-seq (v.1.2.1) с параметрами `--narrow_peak --macs_gsize 3.16e7 --macs_fdr 0.1 --clip_r1 9 --clip_r2 2 --three_prime_clip_r1 2 --three_prime_clip_r2 9` для стадий **Vegetative** и **Streaming** в трех био-

логических репликах, что приблизительно соответствует стадиям 0 и 5 часов развития наших данных. Для всех образцов насчитывается около 2 млн картированных парных прочтений (Рисунок 8а). Пики, полученные с помощью `macs2`, кластеризуются по стадиям развития (Рисунок 8б). Высокое качество полученных данных подтверждено и графиком распределения размера фрагментов (Рисунок 9), поскольку на всех образцах есть пик около 50—100 п.н. (безнуклеосомная фракция) и пик около 150—200 п.н. (мононуклеосомная фракция). На основании этих результатов для дальнейшего анализа были использованы слитые файлы поскольку известно, что увеличение глубины секвенирования улучшает детекцию пиков. Улучшение детекции пиков на слитых данных подтверждается полученными значениями доли прочтений, пересекающих пики (fraction of reads in peaks, FRiP) для реплик отдельно и для слитых файлов (Рисунки 10а и 10б), где для слитых файлов наблюдается повышение значений FRiP при общем хорошем качестве данных.

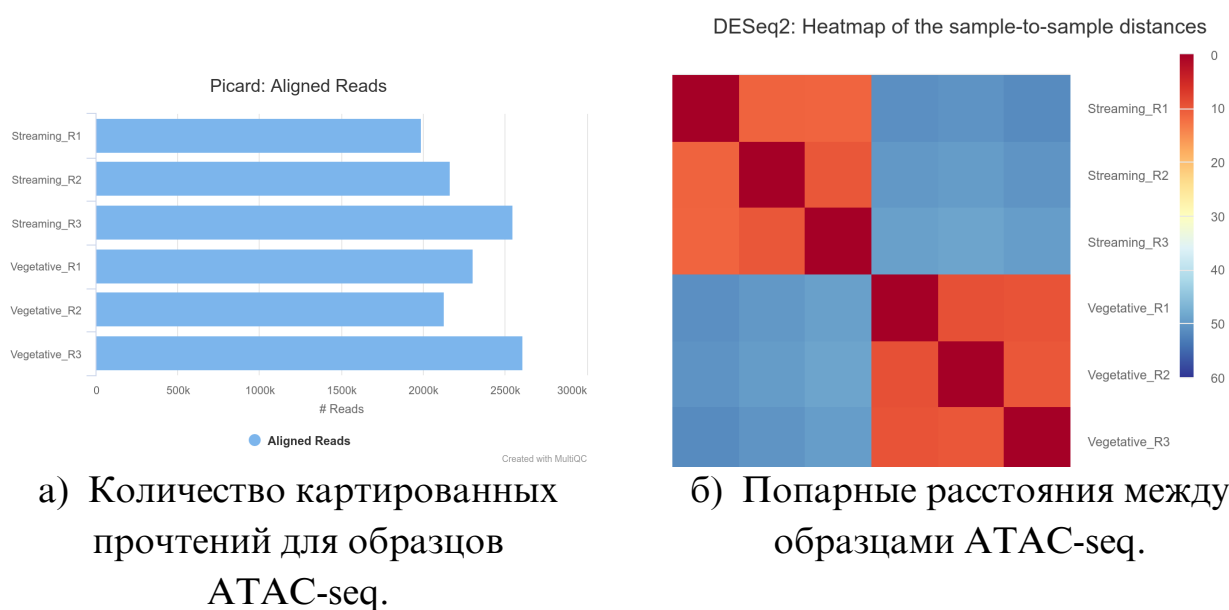


Рисунок 8 — Контроль качества образцов АТАС-seq. Образцы: три биологические реплики стадии Streaming, соответствующей 5 часам развития, и три биологические реплики стадии Vegetative, соответствующей 0 часу развития.

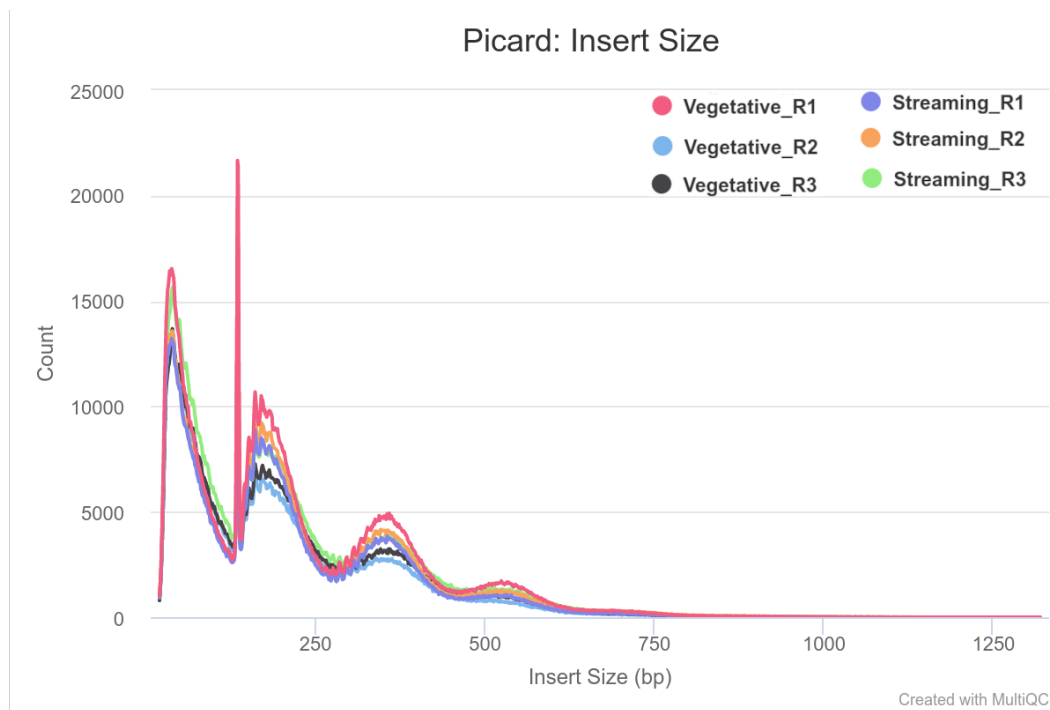


Рисунок 9 — График распределения длин фрагментов для биологических реплик экспериментов ATAC-seq. Образцы: три биологические реплики стадии Streaming, соответствующей 5 часам развития, и три биологические реплики стадии Vegetative, соответствующей 0 часу развития. Пик количества фрагментов размером около 50-100 п.н. представляет собой фрагменты открытого хроматина без нуклеосом, пик количества фрагментов размером около 150-200 п.н. представляет мононуклеосомную фракцию. Последующие пики представляют собой фракцию фрагментов, содержащих 2 и 3 нуклеосомы соответственно. Ось x — размер фрагмента, ось y — количество фрагментов данной длины.

2.17 Аннотация вытянутых петель (el-петель)

Вытянутые петли были выделены из существующей аннотации петель на основе разработанного нами алгоритма: (1) вычислить медиану пяти средних пикселей вдоль 3'-основания (3'_track) и вдоль 5'-основания (5'_track); (2) вычислить fold-change как $\log_{10}(5'_track/3'_track)$; (3) определить «удлиненную вдоль 5'-основания» группу как группу ниже 20-перцентиля значения fold-change; (4) определить группу «вытянутых вдоль 3'-основания» как груп-

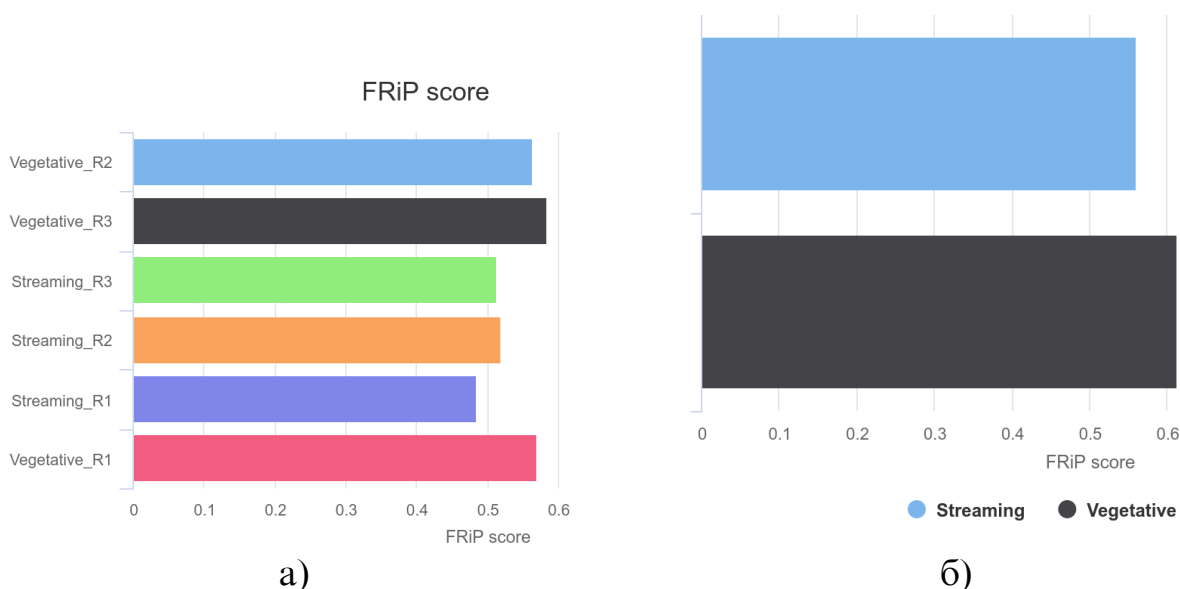


Рисунок 10 — Значения FRIp для экспериментов ATAC-seq. (а) Значения FRIp для биологических реплик экспериментов ATAC-seq. (б) Значения FRIp для файлов слитых реплик ATAC-seq. Образцы: три биологические реплики стадии Streaming, соответствующей 5 часам развития, и три биологические реплики стадии Vegetative, соответствующей 0 часу развития.

пу выше 80-перцентиля fold-change; (5) все точки, попадающие в диапазон от 20 до 80 перцентилей, называются симметричными петлями.

2.18 Анализ данных экспрессии

Прочтения RNA-Seq, полученные коллегами из Института биологии гена РАН, были картированы на геном с помощью инструмента `hisat2`, использовались параметры по умолчанию, кроме максимальной длины интрона (3100), далее была проведена фильтрация уникальных картирований с помощью `samtools` (`mapq > 10`), вычисление покрытия — с помощью `bedtools`. Кроме того, для прочтений, картированных на ген, были вычислены значения Transcripts Per Million (TPM), которые позволяют проводить сравнения уровней экспрессии внутри одного образца:

$$TPM = \frac{N_i L_i 10^6}{\sum (\frac{N_1}{L_1} + \frac{N_2}{L_2} + \dots + \frac{N_n}{L_n})}, \quad (2.2)$$

где N_i — число прочтений, картированных на ген i , L_i — длина гена i , n — общее число генов.

2.19 Поиск повторов в геноме *D. discoideum*

Для поиска повторов в геноме *D. discoideum* была использована программа RepeatMasker. Поиск осуществлялся по всем эукариотическим последовательностям в базе данных RMBlast (параметр `-species eukaryota`).

2.20 Показатель конвергентности

Для анализа направления генов *D. discoideum* была использована метрика «показатель конвергентности» [136]. Сначала были рассчитаны веса каждого бина:

$$G_i = \begin{cases} 1, & N_+ > N_- \\ -1, & N_+ < N_- \\ 0, & N_+ = N_- \end{cases} \quad (2.3)$$

где N_+ — количество генов, лежащих на плюс-цепи, чей TSS расположен в бине i , N_- — количество генов, лежащих на минус-цепи, чей TSS расположен в бине i

Затем, показатели ориентации генов W_i были рассчитаны как:

$$W(i) = S_i \frac{(|i| - w)}{w}, \quad (2.4)$$

где i — порядковый номер бина генома в окне, w — количество бинов в окне, $S = 1$, если $j > 0$, $S = -1$, если $j < 0$.

Наконец, показатель конвергентности был рассчитан как сумма взвешенных показателей ориентации генов G_i :

$$L_b(k) = \sum_{j=-w}^{j=w} (W(j)G_b(k + j)), \quad (2.5)$$

где k — абсолютный номер бина генома, рассчитанный по порядку от 5'-конца хромосомы 1, b — размер бина, j — порядковый номер бина генома в окне, w — количество бинов в окне, G_b — показатель ориентации генов.

2.21 Аннотация энхансер-подобных элементов

Данные, использованные в этом анализе, были получены из работы [244]. Модификации гистонов, в частности данные ChIP-seq H3K4me1 и H3K27ac, были предварительно обработаны с помощью конвейера `nf-core/chipseq v1.2.2`. Аналогичным образом данные ATAC-seq предварительно были обработаны с использованием конвейера `nf-core/atacseq v2.0`. Если полученные пики H3K3me1 и H3K27ac пересекались, использовалось их объединение. Затем полученные пики пересекались с пиками ATAC-seq (Рисунок 11).

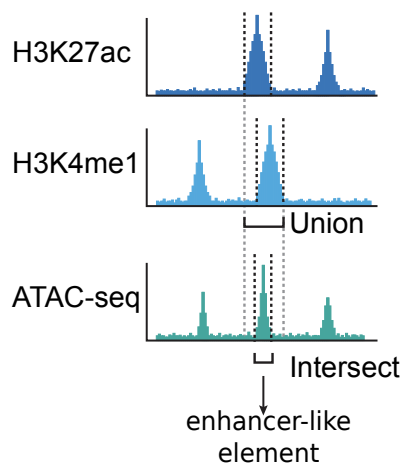


Рисунок 11 — Схема аннотации энхансер-подобных элементов.

2.22 Сравнительный анализ ферритинов представителей типа Porifera

Для поиска ферритинов в геномах представителей типа Porifera были использованы аминокислотная последовательность ферритина человека (HuHF, PDB:4OYN). С помощью `exonerate` были идентифицированы гены-ортологи для HuHF.

Для белков ферритинов из литературы [245; 246] и базы UniProt мы получили информацию о трех функциональных доменах и позициях аминокислот, входящих в состав этих доменов у HuHF: ферроксидазный центр (E28, H66, E108, Q142), ионный канал (D127, E130, E136) и центр минерализации (nucleation site, E61, E64, E67).

Проценты идентичности и сходства аминокислотных последовательностей были рассчитаны с помощью алгоритма `needle` из пакета EMBOSS.

Для идентифицированных генов ферритинов наличие железосвязывающей способности было предсказано с помощью веб-сервера SIRE [247].

2.23 Геномные регионы с измененной локальной упаковкой хроматина

Для оценки изменений локальной упаковки хроматина при анеуплоидиях по сравнению с нормальным кариотипом было использовано соотношение R между частотой контактов с регионами, расположенными на линейном расстоянии не более 1 млн. п.н., и суммой всех цис-контактов. Для регионов, в которых относительно нормального кариотипа при анеуплоидиях хроматин стал менее или более плотный были выбраны пороги, соответствующие 5

2.24 Белок-кодирующие гены человека в регионах с измененной упаковкой при трисомиях

Чтобы проверить, обогащены ли кластеры UP и DOWN генами, кодирующими белки, мы загрузили аннотацию гена из Gencode v43 для сборки

генома hg38. Мы рассчитали количество пересечений между кластерами и генами, используя функцию пересечения из `pybedtools v0.9.0`. Чтобы проверить статистическую значимость такого числа, мы использовали пермутационный тест с 1000 итерациями.

2.25 Расположение генов домашнего хозяйства в регионах с измененной упаковкой при трисомиях

Мы использовали ранее опубликованный список генов домашнего хозяйства [248] и использовали имена генов для получения координат из Gencode v43 для сборки hg38, поскольку исходный список был составлен для сборки hg19. Обогащение генов домашнего хозяйства в кластерах UP и DOWN рассчитывали так же, как и для генов, кодирующих белки.

Глава 3. Результаты

3.1 Организация хроматина *Dictyostelium discoideum*

3.1.1 Выравнивание и контроль качества

Чтобы получить полную картину изменений конформации хроматина, экспрессии и регуляции генов, коллеги из Института биологии гена РАН провели эксперименты Hi-C [191] и RNA-seq для четырех стадий развития *D. discoideum* (штамм AX4) через 0, 2, 5 и 8 часов после индукции голодания (Рисунок 2). Дополнительно были проведены исследования структуры хроматина на различных стадиях клеточного цикла: вход в митоз (0 часов после синхронизации), G1-фаза (1 час после синхронизации), ранняя S/G2 (3 часа после синхронизации) и рассинхронизированные постмитотические клетки (6 часов после синхронизации). Стадия 6 часов после синхронизации соответствует 0 часу развития.

Для каждой точки развития были получены карты Hi-C в двух биологических репликах в соответствии с ранее опубликованным протоколом [194] (см. раздел 2.1). Библиотеки секвенировали на платформе Illumina HiSeq 4000 с глубиной прочтений около 100 млн парных прочтений на библиотеку. Полученные парные прочтения Hi-C обрабатывались А. Галицыной от сырых прочтений до этапа получения карт в формате mcool включительно с помощью конвейера `distiller` с параметрами по умолчанию. Для выравнивания была использована сборка генома *D. discoideum* версии 2.7. Прочтения с показателями MAPQ менее 30, результаты множественного картирования, пары, пары, закартированные на геном только одним прочтением, и потенциальные дубликаты ПЦР были отфильтрованы. Важнейшие из показателей качества представлены в Таблице 5.

Таблица 5 — Показатели качества образцов Hi-C.

Имя образ-ца	Число пар прочтений, млн		% пар с одним закартированным прочтением	% пар, где оба прочтения закартированы		% дубликатов	% цис-пар
	млн	% некартированных		% пар с одним закартированным прочтением	% пар с двумя закартированными прочтениями		
0A	75,8	36,2 %	18,4 %	45,4 %	9,2 %	82,1%	
0B	96,0	35,9 %	18,5 %	45,6 %	11,2 %	81,5 %	
2A	84,1	38,9 %	13,9 %	47,1 %	11,8 %	84,8 %	
2B	78,6	38,8 %	14,3 %	46,8 %	13,0 %	84,6 %	
5A	96,2	41,2 %	13,5 %	45,3 %	13,3 %	86,3 %	
5B	80,5	40,2 %	13,8 %	46,0 %	13,6 %	85,9 %	
8A	78,3	35,3 %	16,5 %	48,2 %	14,0 %	82,7 %	
8B	72,3	37,0 %	16,5 %	46,4 %	13,2 %	82,2 %	

Для оценки корреляции между биологическими репликами был использован инструмент HiCRep (см. раздел 2.6), в результате была выявлена высокая корреляция как между биологическими репликами (~98 %, Рисунок 12), так и между часами развития *D. discoideum*. На основании этой высокой корреляции и визуального сходства карт контактов, анализы структуры хроматина были проведены на карте 0 часа развития, если не сказано иное.

Поскольку биологические повторности продемонстрировали высокую корреляцию, для облегчения детекции петель биологические реплики были слиты на уровне пар контактов, а затем полученные карты были приведены к одинаковому количеству картированных прочтений (46 млн. п.н.) для более аккуратных сравнений между разными часами развития. Кроме того, карты Hi-C были нормированы с помощью алгоритма итеративной коррекции [207]. Несмотря на то, что нам удалось получить карты с разрешением 500 п.н. (Рисунок 13а), они содержат множество белых полос, появляющиеся, по-видимому, не из-за недостаточной для такого разрешения глубины секвенирования, а вследствие высокого содержания повторов в геноме. По этой причине основные анализы проводились на картах Hi-C с разрешением 2000 п.н. (пример региона на Рисунке 13б), если не было указано иное. Даже

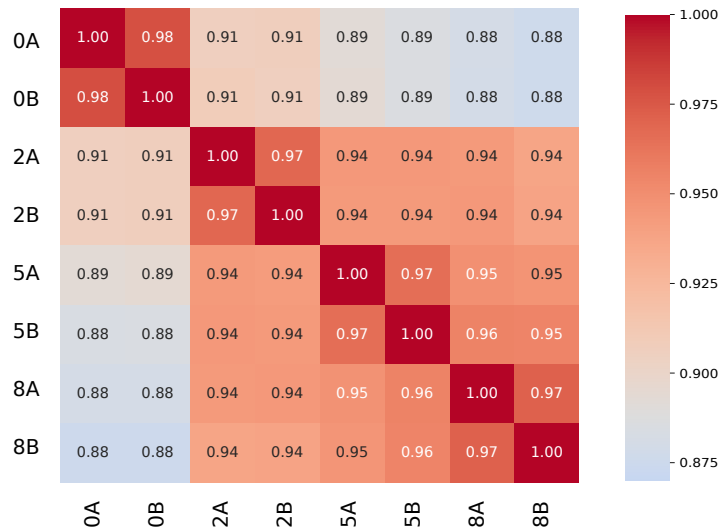
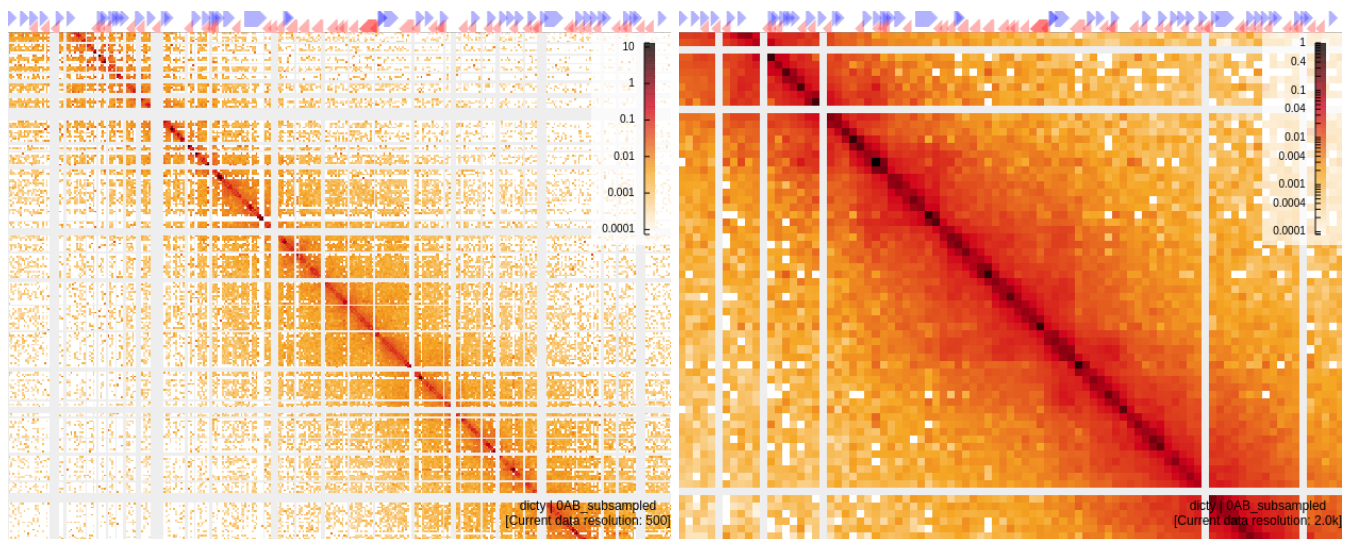


Рисунок 12 — Карта корреляций между биологическими репликами и часами развития. Значения коэффициента корреляции рассчитаны с помощью инструмента HiCRep.

карты этого разрешения имеют значительное количество белых полос, что, вероятно, подтверждает гипотезу связи с большим количеством повторов в геноме *D. discoideum*.



а) Пример карты контактов с размером бина 500 п.н.

б) Пример карты контактов с размером бина 2000 п.н.

Рисунок 13 — Пример карты контактов при размере бина 500 и 2000 п.н. Регион хромосомы 3 длиной 140 тыс. п.н.

3.1.2 Анализ межхромосомных и внутрихромосомных взаимодействий *D. discoideum*

Анализ соотношения цис- и транс- контактов является одним из показателей качества, однако в случае *D. discoideum* у нас отсутствовали референтные значения, поэтому показатели цис-транс соотношений были использованы для сравнительного анализа распределения частот взаимодействий на разных часах развития. Из Таблицы 6 видно, что показатели соотношений цис-транс находятся в достаточно узком диапазоне, от ~3.4 до ~5.0, а для стадий, которые, согласно нашим представлениям о биологии *D. discoideum*, должны быть наиболее похожи — 0 час развития и 6 час после синхронизации по клеточному циклу — значения цис-транс соотношений похожи. Однако важно понимать, что такое соотношение отражает лишь глобальное распределение частот контактов и ничего не говорит о локальной упаковке хромосом.

Таблица 6 — Показатели цис-транс соотношений на разных стадиях развития и клеточного цикла.

образец	отношение цис-транс-контактов
0AB	3,720
2AB	4,337
5AB	4,946
8AB	3,776
0AB (cell cycle)	4,614
1AB (cell cycle)	3,935
3AB (cell cycle)	4,106
6AB (cell cycle)	3,390

На уровне межхромосомных взаимодействий заметны точечные обогащения частот контактов, что, вероятно, соответствует взаимодействиям центромерных регионов (Рисунок 14), расположенных в области 5'-конца хромосом. Однако отметим, что хромосома 5 имеет перевернутый вид, поскольку ее предполагаемое обогащение центромерными контактами в зоне транс-взаимодействий расположено ближе к 3'-концу; во всех дальнейших

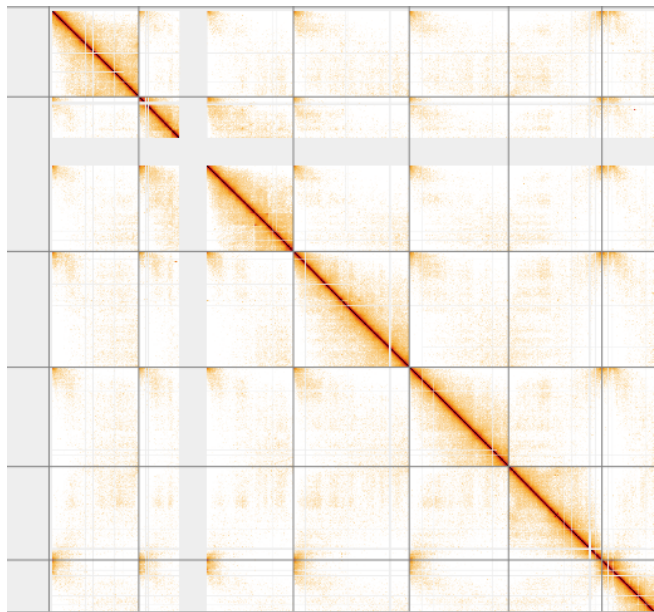


Рисунок 14 — Полногеномная карта контактов *D. discoideum*, иллюстрирующая центромерные взаимодействия между разными хромосомами.

анализах эта хромосома была развернута для унификации. Для анализа силы транс-взаимодействий (предполагаемо центромерных) и подтверждения структуры Рабля был проведен анализ силы таких взаимодействий в хроматине *D. discoideum*.

Вначале была построена карта средних межхромосомных взаимодействий для *D. discoideum* (Рисунок 3). Для этого из полногеномной карты Hi-C были выбраны только межхромосомные взаимодействия, полученные матрицы приведены к одному размеру, и построена средняя карта межхромосомных взаимодействий. При этом важно учитывать особенность генома *D. discoideum*: наличие крупной дупликации на хромосоме 2. Наш анализ центромерных взаимодействий показал, что регион, следующий за дупликацией, по-видимому, активно контактирует со всеми другими 5'-концами хромосом (включая регион на той же хромосоме). Были изучены межхромосомные взаимодействия отдельно для каждой из хромосом, полученные путем усреднения взаимодействий каждой из хромосом со всеми остальными хромосомами (методология описана в разделе 2.5). Результаты анализа показывают, что все шесть хромосом *D. discoideum* демонстрируют повышение частот взаимодействия в верхнем левом углу средней карты межхромосомных взаимодействий, что соответствует межцентромерным контактам. Интересно отметить, что матрица частот межхромосомных контактов хромосомы 6 помимо обогащения в левом верхнем углу имеет еще одно пятно

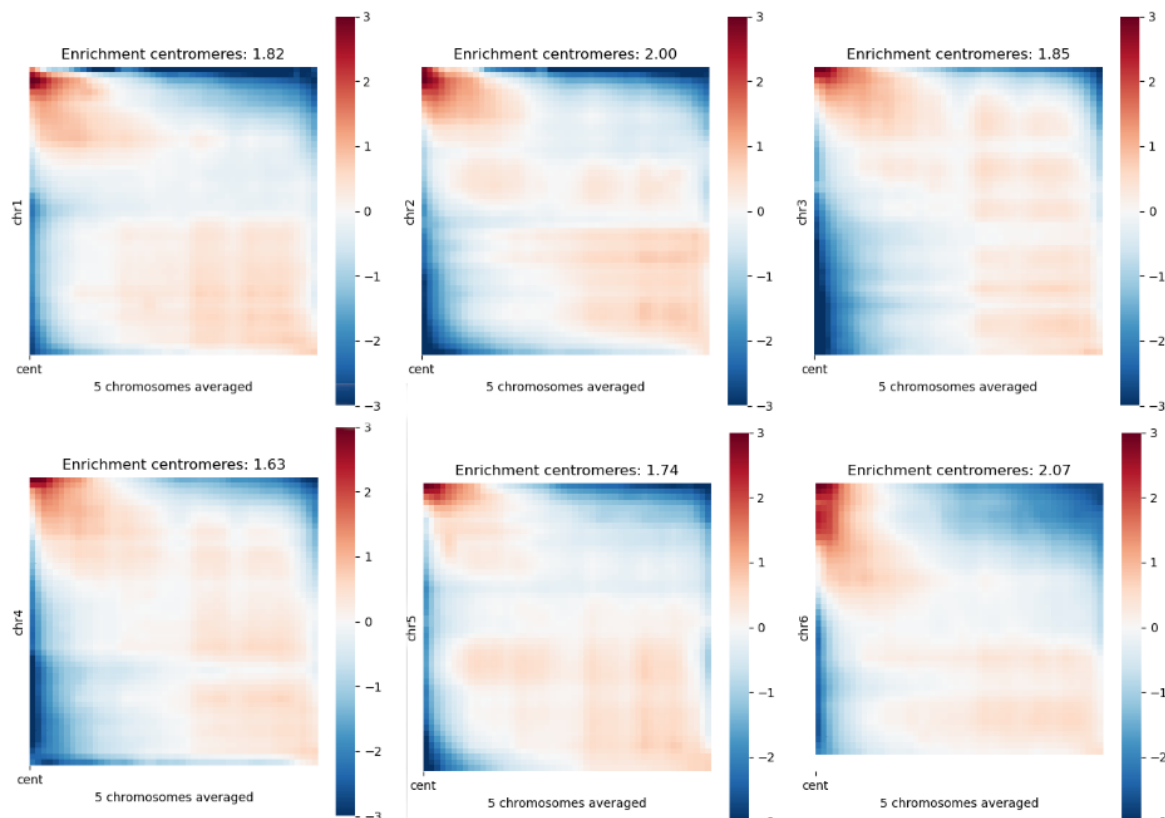


Рисунок 15 — Структура межхромосомных контактов каждой отдельной хромосомы *D. discoideum*. Средняя карта нормализованных частот межхромосомных взаимодействий для хромосом 1, 2, 3 (верхняя панель), хромосом 4, 5, 6 (нижняя панель).

повышенный частот контактов, что может свидетельствовать о наличии контактирующего со всеми прочими центромерами региона.

Далее был проведен анализ межхромосомных взаимодействий на разных стадиях развития (Рисунок 16a) и клеточного цикла (Рисунок 16б). Из анализа видно сильное обогащение частот контактов 5'-концов хромосом друг с другом на всех часах развития, что свидетельствует о сильном центромерном взаимодействии и подтверждает структуру Рабля для хромосом *D. discoideum*, как на часах развития, так и на стадиях клеточного цикла. При этом в процессе развития сила центромерных взаимодействий выше, чем на изученных стадиях клеточного цикла, что может быть, например, результатом большей глубины прочтений карт стадий клеточного цикла, что повышает среднюю частоту межхромосомных контактов (ожидаемое). Однако анализ фрагментов скейлинг-графиков, соответствующих взаимодействиям на расстоянии более 100 тыс. п.н. (Рисунок 17), показывает, что глубина прочтений не является определяющей в ожидаемой частоте

контактов, поскольку ожидаемые частоты контактов для 0 часов развития превышают значения всех прочих часов развития, но не для 3 и 6 часа после синхронизации по клеточному циклу, несмотря на различия в глубине секвенирования. Этот результат свидетельствует о том, что наблюдаемые различия частот межхромосомных взаимодействий между стадиями развития и клеточного цикла, по-видимому, имеют биологическое обоснование, например, изменение экспрессии структурных белков хроматина, однако точный механизм в рамках данной работы не установлен.

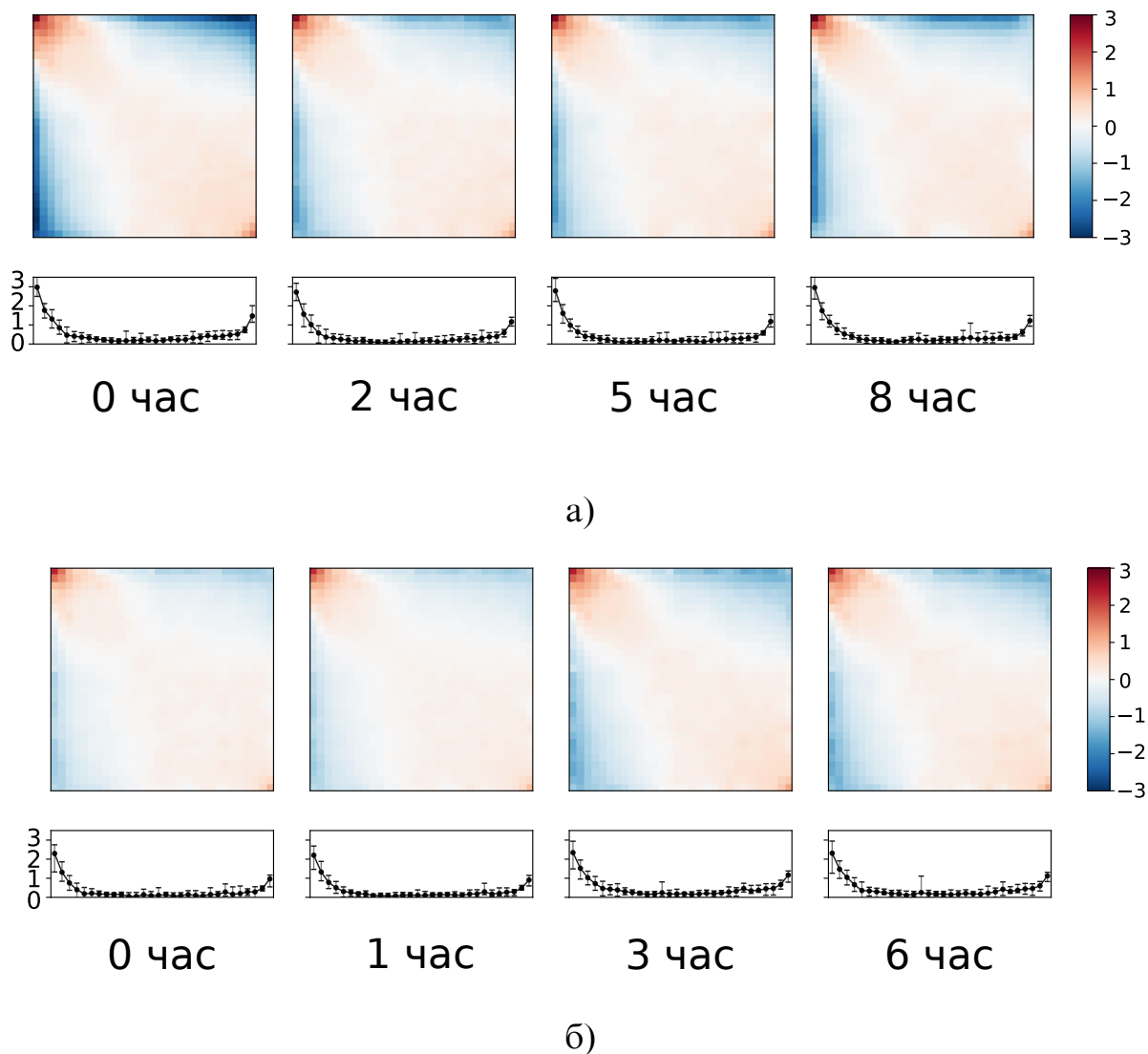


Рисунок 16 — Карта межхромосомных взаимодействий для (а) стадий цикла развития; (б) стадий клеточного цикла.

В то же время взаимодействия вдоль хромосом находятся на уровне среднего межхромосомного взаимодействия (ожидаемого) даже на стадии входа в митоз, что, по-видимому, следует интерпретировать как отсутствие

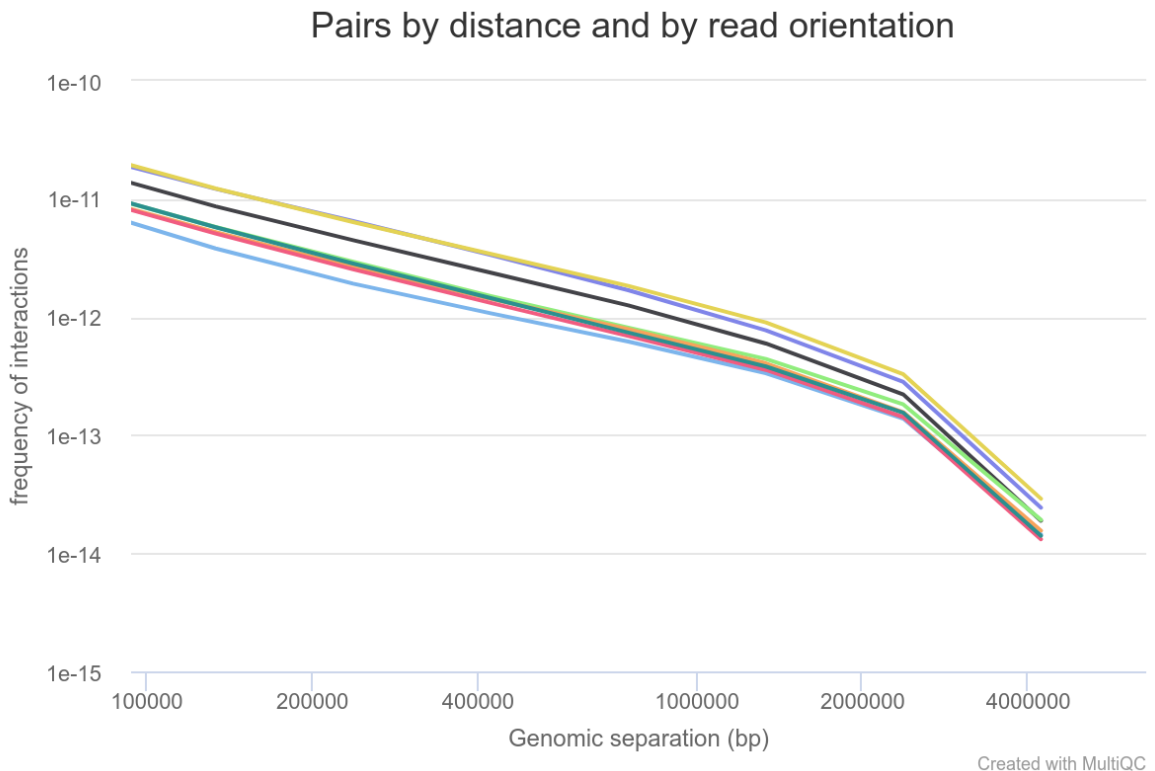


Рисунок 17 — Фрагмент scaling-графика для клеточного цикла и часов развития. 0, 2, 5 и 8 часы развития — черный, оранжевый, розовый и бирюзовый соответственно; 0, 1, 3 и 6 часов после синхронизации по клеточному циклу — голубой, салатовый, фиолетовый и желтый соответственно.

активных взаимодействий по телу хромосомы, что также свидетельствует в пользу структуры Рабля. Важно также отметить обогащение частот контактов 3'-концов хромосом на уровне в 2-3 раза чаще ожидаемого, что является свидетельством теломерных взаимодействий. При сравнении медианных частот контактов на разных часах развития и стадиях клеточного цикла теломерные взаимодействия оказываются более выраженными для всех часов развития по сравнению со стадиями клеточного цикла. Кроме того, необходимо понимать, что повышения частот взаимодействий 5'-концов хромосом нельзя интерпретировать только как центромерные взаимодействия, поскольку хромосомы *D. discoideum* акроцентрические, т.е. хромосомы имеют одно выраженное плечо, а центромеры располагаются близко к теломерам. В соответствии с этим, говоря о повышении частот контактов 5'-концов хромосом, мы имеем в виду не только центромерные взаимодействия, но и неизбежные теломерные взаимодействия.

Одновременно с этим мы наблюдаем понижение по сравнению с ожидаемым частот взаимодействия 5'- и 3'-концов разных хромосом друг с другом,

особенно заметные для графиков отдельных хромосом (Рисунок 15). Кроме того, при сравнительном анализе межхромосомных взаимодействий на 0 часе развития наблюдается понижение частот взаимодействия 5'-конца с телами других хромосом, что может свидетельствовать о значительном отличии хромосомной организации на 0 часе развития; этому посвящен анализ в разделе 3.1.3 текущей главы. Глобально, пониженные частоты контактов двух концов разных хромосом свидетельствует о полярности расположения центромерных и теломерных концов в ядрах *D. discoideum*.

Для проверки этой гипотезы были построены средние карты внутрихромосомных взаимодействий (Рисунок 18a), полученные путем приведения к одному размеру и вычислению средних контактов внутри всех хромосомы. Если бы 5'- и 3'- концы хромосом располагались в ядре близко, то мы бы наблюдали частоты контактов, превышающие среднюю частоту контактов для данных линейных расстояний («ожидаемое»). Однако из графиков (Рисунок 18a) видно значительное обеднение по сравнению с ожидаемым частот контактов 5'-концов хромосом со всей остальной хромосомой и 3'-концами на всех исследованных часах развития (верхняя полоса каждого графика соответственно). На нижней панели для каждого графика усредненных частот внутрихромосомных контактов отображены графики частот контактов вдоль верхней строки матрицы, из которых видно, что отношение средней частоты контактов двух претеломерных областей хромосом к ожидаемому не превышает 0,03, т.е. более чем в тридцать раз ниже. При этом на 0 часу развития это отношение снижается еще сильнее, в то время как на малых расстояниях наблюдается повышение частот взаимодействия по сравнению с остальными часами развития (подробный анализ особенностей организации хроматина 0 часа развития описан в разделе 3.1.3 данной главы). Средняя карта внутрихромосомных взаимодействий, построенная для стадий клеточного цикла (Рисунок 18б), свидетельствует о том, что хромосомы *D. discoideum* на разных стадиях клеточного цикла в среднем характеризуются более высокой частотой контактов 5'-концов с телом хромосом по сравнению с картами стадий развития.

Отдельно важно отметить, что на средних картах внутрихромосомных взаимодействий в хроматине *D. discoideum* (Рисунок 18a) не усматривается существование выраженных хромосомных территорий, которые присутствуют, например, у млекопитающих [249], что также может свидетельствовать

в пользу иной укладки хромосом в ядре, вероятно, структуры Рабля, обсуждаемой в этом разделе.

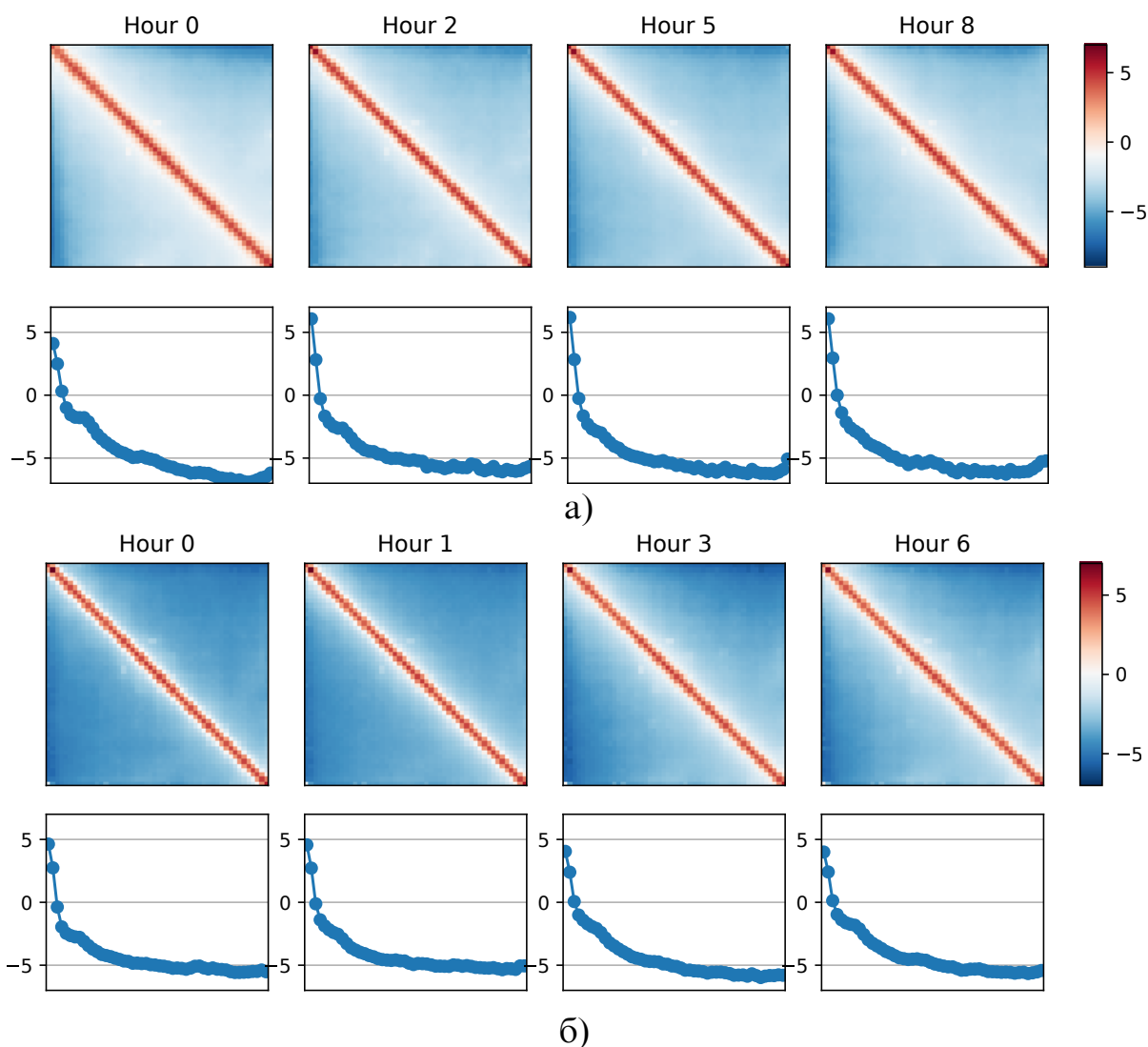


Рисунок 18 — Сравнительный анализ карт средних внутривнутрихромосомных взаимодействий для (а) стадий развития; (б) стадий клеточного цикла. Верхняя панель: средняя карта межхромосомных взаимодействий. Нижняя панель: графики распределения частот контактов вдоль верхней строки карты межхромосомных взаимодействий с верхней панели.

Наконец, был проведен сравнительный анализ карт средних межхромосомных взаимодействий для дрожжей *S. cerevisiae* и *D. discoideum* (Рисунок 19). Выбор дрожжей для сравнения обусловлен тем, что это один из немногих представителей низших эукариот, для которых имеются карты хорошего разрешения и для которых описана структура Рабля. Для сравнительного анализа межхромосомные взаимодействия двух организмов были приведены к одному масштабу, и была построена карта средних межхромосомных

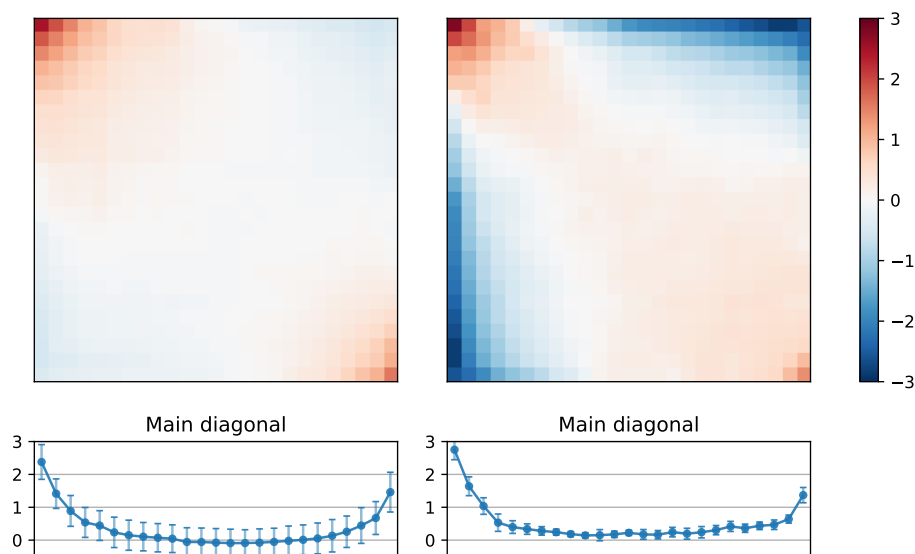


Рисунок 19 — Сравнительный анализ средних карт межхромосомных взаимодействий *D. discoideum* и *S. cerevisiae*. Слева: *D. discoideum*, справа: *S. cerevisiae*. Верхняя панель: средняя карта межхромосомных взаимодействий. Нижняя панель: графики распределения частот контактов вдоль верхней строки карты межхромосомных взаимодействий с верхней панели.

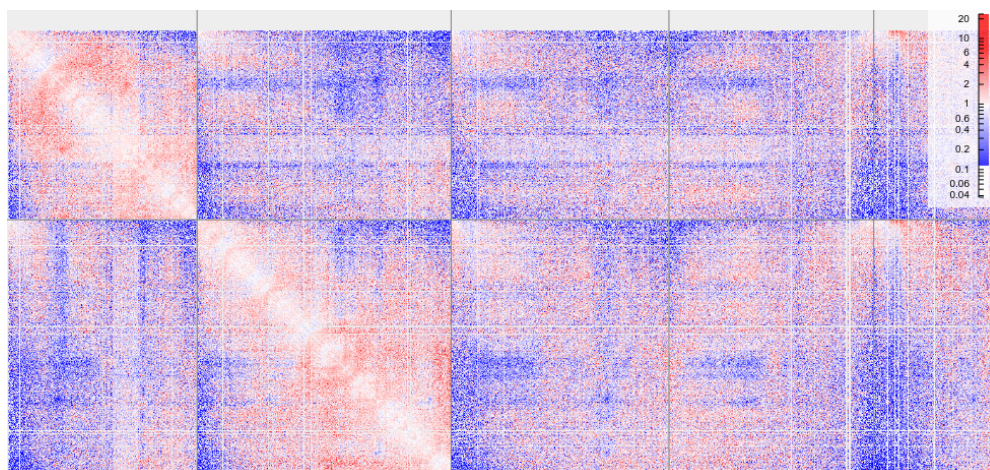
взаимодействий. В результаты были получены схожие частоты взаимодействий, нормализованные на ожидаемое, как для 5'-концов хромосом, так и для 3'-концов хромосом между разными хромосомами, но не между 3'- и 5'-концам разных хромосом. При этом обеднение этих контактов у *D. discoideum* выражено куда сильнее, чем у *S. cerevisiae*. Проведенный сравнительный анализ в сумме с результатами, приведенными выше, свидетельствует об укладке хромосом в ядрах почкующихся дрожжей и *D. discoideum* в структуры Рабля (Рисунок 19), которые, однако, отличаются по силе межхромосомных взаимодействий 5'- (теломера и центромера) и 3'-концов (только теломера) хромосом.

3.1.3 Изучение особенностей упаковки хроматина на 0 часе развития *D. discoideum*

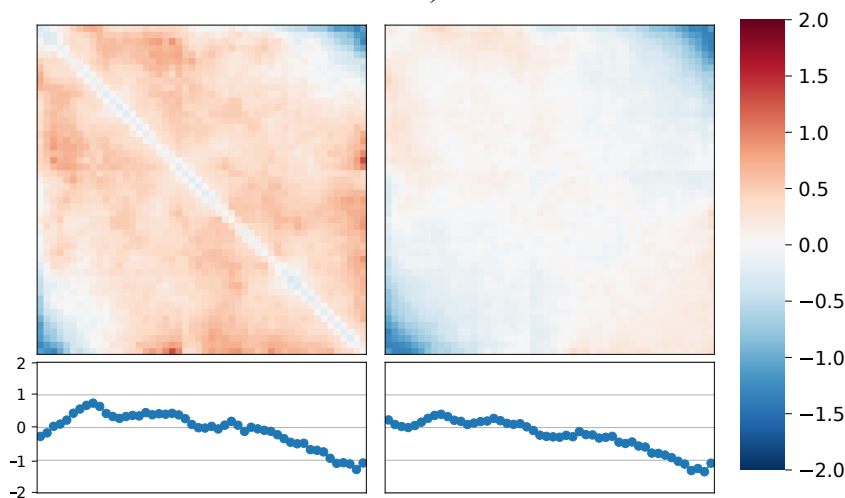
Результаты анализа межхромосомных взаимодействий у *D. discoideum* показали, что взаимодействие 5'-конца (центромерного) хромосом с телом хромосомы и особенно с 3'-концом (теломерным) слабее на 0 часе развития по сравнению с другими часами развития (Рисунок 16). Более того, аналогичная тенденция наблюдается и для карт средних внутривхромосомных взаимодействий (Рисунок 18а). Для изучения изменений в устройстве хроматина на уровне всего генома на 0 часе развития был проанализирован график скейлинга (Рисунок 17). Из него следует, что укладка хроматина на 0 часе развития отличается от остальных часов развития небольшим увеличением средних частот контактов на расстоянии около 80 тыс. п.н. График корреляции (Рисунок 12) также подтверждает, что карты Hi-C 0 часа развития сильно отличаются от других часов развития.

Для изучения этого эффекта по всему геному все значения матриц контактов на 0 часе развития были поделены на соответствующие значения на матрице 8 часа развития (Рисунок 20а). Результаты анализа показывают, что на 0 часе развития наблюдается более активное внутривхромосомное взаимодействие и понижение частот контактов концов хромосом по сравнению с 8 часом развития.

Для количественного анализа этого феномена был проведен сравнительный анализ частот внутри- и межхромосомных взаимодействий (см. раздел 2.7). На Рисунке 20б представлен график различий внутри- и межхромосомных взаимодействий между 0 и 8 часами развития. Мы наблюдаем меньшую частоту контактов внутри хромосом, а также большую частоту контактов концов хромосом на 8 часе развития по сравнению с 0 часом развития (левая панель Рисунка 20б). Среди межхромосомных взаимодействий выделяется большая в ~2-3 раза частота контактов центромерных и теломерных концов разных хромосом друг с другом на 8 часе развития по сравнению с 0 часом развития (правая панель Рисунка 20б). Полученные результаты свидетельствуют в пользу пониженных частот взаимодействия концов хромосом на 0 часа развития при, возможно, большей компактизации хромосом.



а)



б)

Рисунок 20 — Различия хромосомных взаимодействий между часами развития. (а) Пример изменения частот контактов между 0 и 8 часами. (б) Усредненные изменения внутри- и межхромосомных контактов между 0 и 8 часами. Положительные значения означают понижение частот контактов на 8 часе развития в сравнении с 0 часом, а отрицательные значения — повышение частот контактов на 8 часе развития по сравнению с 0 часом развития. Матрицы частот контактов 0 часа развития поделены на соответствующие значения матриц частот контактов 8 часа развития. Средние карты для внутрихромосомных и межхромосомных взаимодействий получены как описано в разделе 2.5. Размер бина матрицы — 16 тыс. п.н.

Причиной организации хроматина на 0 часе развития, отличной от других часов развития, может быть неоднородность клеточной культуры в начале эксперимента, так как *D. discoideum* был инициирован из культуры, состоящей из обычных и спорных клеток. Споры проросли и превращались в амёб, которые, как и обычные клетки, вступали в цикл развития; клеточная

культура ко 2 часу развития становилась однородной. Чтобы проверить это наблюдение с помощью вычислительных методов, мы провели анализ GO-категорий для генов, повышенно экспрессирующихся на 0 часе развития в сравнении с 8 часом развития. Результаты анализа, приведенные в Таблице 7, демонстрируют статистически значимую связь генов с повышенной экспрессией на 0 часе развития и спорообразования и образования клеточных стенок спор.

Таблица 7 — GO термины генов с повышенной экспрессией на 0 часе развития

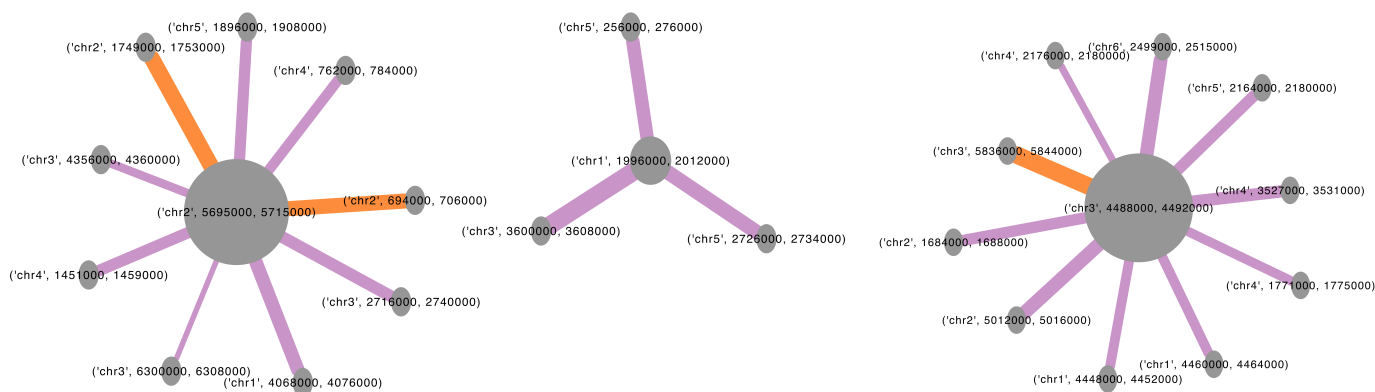
GO-термин	Кол-во генов в группе	Функция	<i>p</i> -значение
GO:0042244	39/94	образование спор- рочной стенки	0,001
GO:0031153	39/101	развитие слизи	0,003
GO:0090665	39/90	гликопротеиновый комплекс	< 0,001
GO:0031160	39/101	споровая стенка	0,001

3.1.4 Дальние контакты *D. discoideum*

При визуальном анализе карт Hi-C в HiGlass было выявлено обогащение точечных дальних взаимодействий по сравнению с окружением на карте Hi-C (как межхромосомных, так и внутрихромосомных). Вследствие их небольшого количества и отсутствия специальных инструментов для их аннотации, все дальние взаимодействия были аннотированы вручную на разрешении карты 4 тыс. п.н. с использованием HiGlass [230]. На 0 и 8 часах развития были обнаружены соответственно 12 и 11 точечных взаимодействий; оба набора дальних контактов специфичны только для одного часа развития и визуально не детектируются на других часах развития.

Поскольку визуальный анализ свидетельствует, что многие локусы вовлечены сразу в несколько дальних контактов, мы построили графы взаимо-

действий между всеми локусами, аннотированными как основания дальних контактов для 0 и 8 часов соответственно (Рисунки 21а и 21б). Полученные графы свидетельствует о том, что существуют локусы, которые являются одним из оснований подавляющего большинства дальних контактов одновременно.



а) Граф дальних взаимодействий на 0 часе развития.

б) Граф дальних взаимодействий на 8 часе развития.

Рисунок 21 — Графы дальних взаимодействий на 0 и 8 часах развития. Узел — основание контакта. Частота контактов отображена на графе как ширина соответствующего ребра; оранжевый цвет ребра обозначает цис-взаимодействия, фиолетовый цвет ребра — транс-взаимодействия. Размер узла пропорционален количеству дальних взаимодействий, образуемых данным локусом.

В связи с существованием локусов, которые являются одним из оснований подавляющего большинства дальних контактов одновременно, была выдвинута гипотеза о принадлежности дальних контактов к центромерным регионам. Поскольку взаимодействия центромер в карте контактов имеет наибольшую частоту контактов среди межхромосомных взаимодействий, для проверки центромерной природы дальних контактов необходимо провести сравнительный анализ частот контактов для центромерных взаимодействий и дальних контактов. Сила контактов была рассчитана как сумма частот контактов в окне 3×3 на разрешении 20 тыс. п.н., преобразованных с помощью Формулы 2.1. Проведенный анализ показывает, что, хотя частота центромерных взаимодействий (левая панель Рисунка 22) меняется в ходе развития *D. discoideum*, она остается значительно выше частот дальних контактов на любом из часов развития (центральная и правая панели Рисунка

22). Полученные результаты позволяют отвергнуть центрмерную гипотезу дальних контактов.

Кроме того, проведенный анализ позволяет подтвердить, что дальние контакты являются специфичными для определенного часа развития. Анализ распределения силы дальних контактов на всех часах развития (центральная и правая панели Рисунка 22) показывает, что для дальних контактов, специфичных для 0 и 8 часов развития, наблюдаются максимальные частоты на этих часах развития. При этом, на других стадиях медианные значения меньше минимум на порядок. Полученные результаты подтверждают визуальные наблюдения дальних контактов, специфичных для одной стадии.

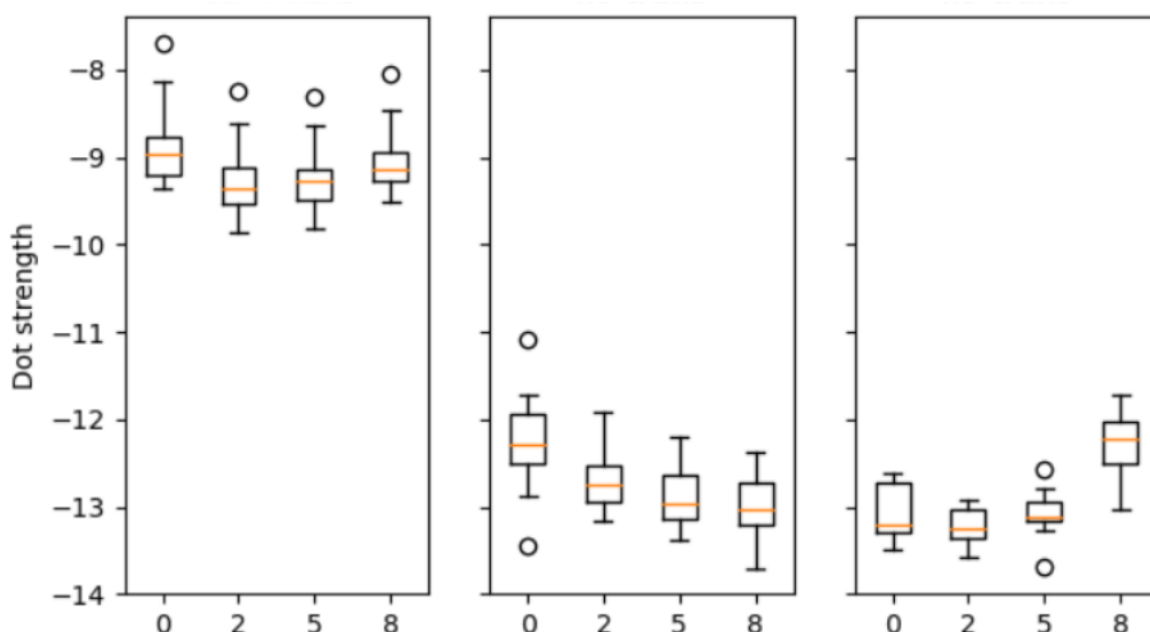


Рисунок 22 — Характеристики дальних контактов. Сравнение частот контактов для транс-центромерных взаимодействий (cent-cent) на часах развития, а также для аннотированных дальних контактов на 0 и 8 часах развития (h0-trans, h8-trans). Ось x — часы развития, ось y — сила дальних контактов, рассчитанная как сумма частот контактов в окне 3×3 на разрешении 4 тыс. п.н., преобразованных с помощью Формулы 2.1.

Основываясь на данных литературы о пространственных кластерах некодирующих РНК [250], мы предположили, что наблюдаемые дальние контакты нужны для пространственного сближения локусов генов некодирующих РНК. Для проверки этой гипотезы были взяты разметки генов длинных некодирующих РНК и антисмысловых РНК из работы [236] и подсчитано,

сколько из них пересекаются с основаниями дальних контактов, а в качестве контроля 1000 раз были выбраны случайные регионы и подсчитано количество пересечений с каждым из 1000 наборов случайных регионов. Проведенный анализ показал, что в основаниях дальних контактов длинные некодирующие РНК представлены чаще, чем в случайных областях, как на 0, так и на 8 часах развития (Рисунки 23а и 23б соответственно), что статистически достоверно (p -значение $< 0,001$, пермутационный тест). Всего было найдено 22 и 17 генов длинных некодирующих РНК, пересекающих основания дальних контактов на 0 и 8 часах развития соответственно.

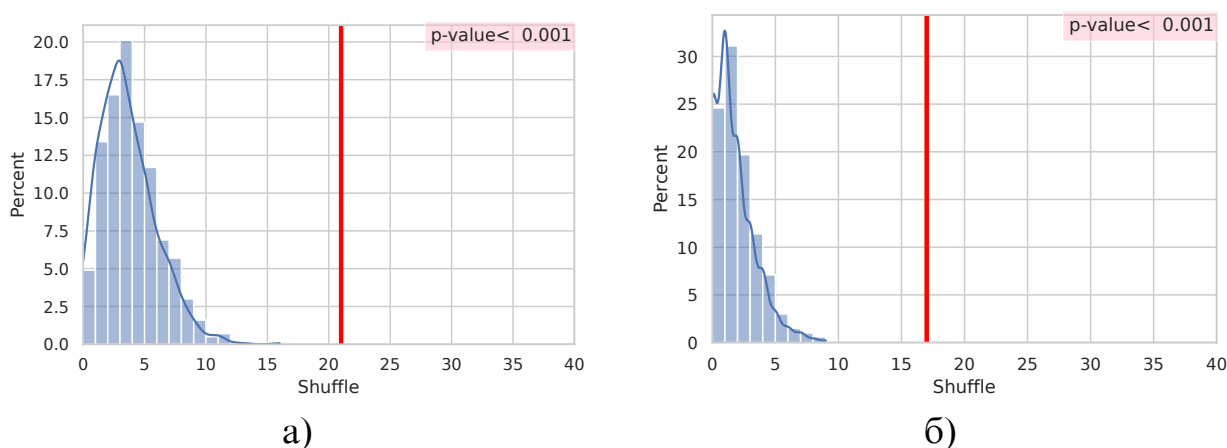


Рисунок 23 — Количество генов длинных некодирующих РНК, пересекающих основания дальних контактов на 0 (а) и 8 (б) часах развития; p -значение рассчитано пермутационным тестом. Синее распределение — распределение количества генов длинных некодирующих РНК, пересекающих основания дальних контактов в пермутациях, красная линия — количество генов длинных некодирующих РНК, пересекающих основания дальних контактов в реальных данных. Ось x — количество генов длинных некодирующих РНК, пересекающих основания дальних контактов, ось y — процент пермутаций.

Кроме того, был проведен сравнительный анализ экспрессии генов длинных некодирующих РНК, пересекающих основания дальних контактов и случайного контроля. Уровни экспрессии генов оказались статистически неотличимы от контроля ни на 0, ни на 8 часах развития (результаты не представлены). Однако затем вместо абсолютных значений экспрессии мы использовали разность экспрессии между 8 и 0 часами развития. Результаты этого анализа для оснований дальних контактов, специфичных как для 0 часа развития, так и для 8 часа развития (Рисунки 24а и 24б соответственно) демонстрируют статистически значимые отличия от контроля (p -значения

$< 0,0001$ и $< 0,01$ соответственно для контактов 0 и 8 часов развития, тест Манна-Уитни).

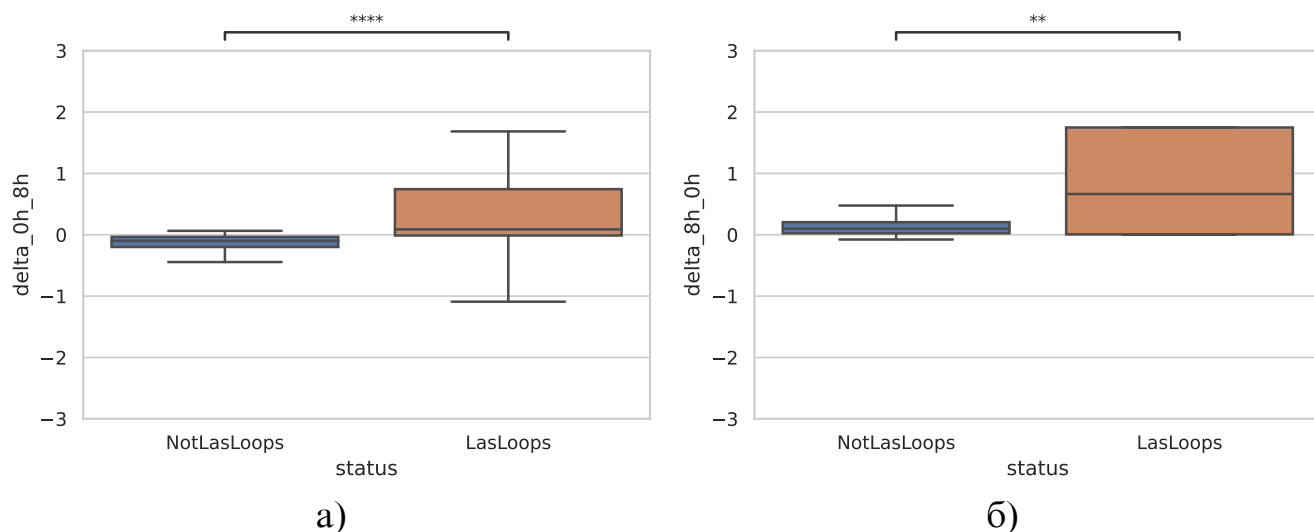


Рисунок 24 — Различия экспрессии генов длинных некодирующих РНК, лежащих в основаниях дальних контактов в сравнении со случайно выбранными генами длинных некодирующих РНК (а) для дальних контактов, специфичных для 0 часа развития, (б) для дальних контактов, специфичных для 8 часа развития. Ось y — разность экспрессии между часами развития, на которой наблюдается дальний контакт и часа развития, на котором дальний контакт отсутствует. ** p -значение $< 0,01$; **** p -значение $< 0,0001$, тест Манна-Уитни.

Эти результаты свидетельствуют о том, что специфические для стадий дальние точечные взаимодействия у *D. discoideum*, по-видимому, обусловлены контактами генов длинных некодирующих РНК. В литературе было показано наличие у длинных некодирующих РНК стадий-специфичных уровней экспрессии [15; 251]; мы показали, что именно длинные некодирующие РНК с измененными между стадиями развития уровнями экспрессии располагаются в основаниях дальних контактов.

3.1.5 Численная оценка компарментов у *D. discoideum*

На карте контактов хроматина *D. discoideum* была обнаружена очень слабая структура вида «шахматная доска». Для численной оценки ком-

партментализации была проведена классическая вычислительная процедура, описанная в работе [133]. Были подсчитаны значения корреляции Пирсона между профилями контактов каждого бина генома со всеми остальными (строками матрицы Hi-C) на разрешении 20 тыс. п.н., а затем рассчитаны собственные вектора полученных матриц корреляции. Проведенный анализ демонстрирует слабое разделение генома на два компартмента с помощью первого собственного вектора, которые к тому же плохо соотносятся с полученной матрицей контактов (Рисунок 25). Несколько лучше предполагаемое разделение отражают второй и третий собственные вектора. Однако для разных хромосом лучше отражающим визуальные наблюдения является то один, то другой из этих двух векторов. Это наблюдение свидетельствует о том, что собственные вектора улавливают скорее некоторые особенности укладки каждой из хромосом.

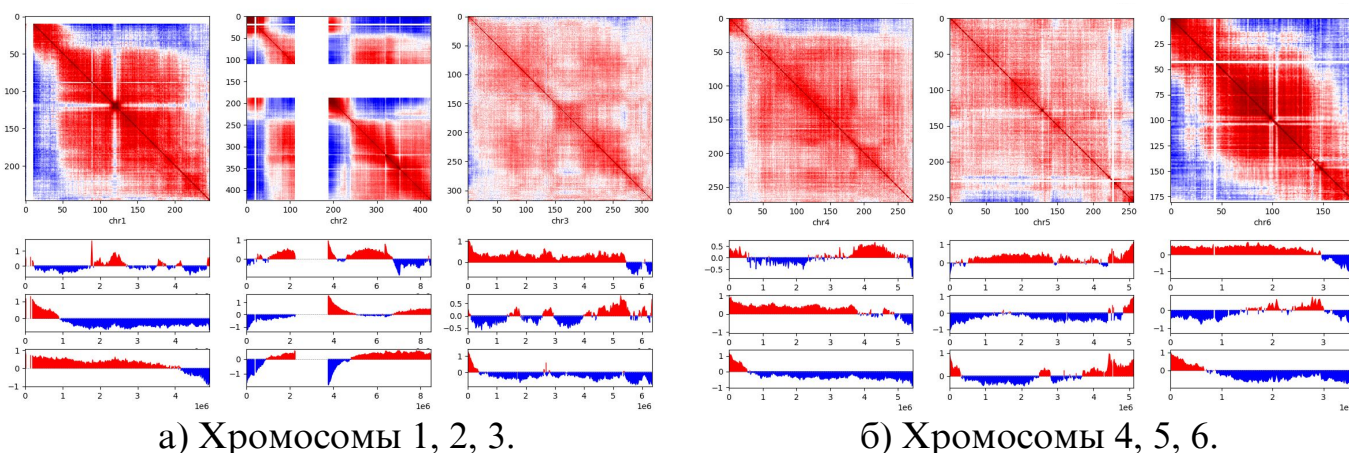


Рисунок 25 — Матрица корреляций Пирсона и первые три вектора собственных вектора для всех хромосом.

Для изучения взаимодействия предполагаемых компартментов была также построена седловая диаграмма (Рисунок 26a). Из рисунка видно, что частоты контактов в левом верхнем и нижнем правом углах графика повышены относительно средней частоты контактов на данном линейном расстоянии («ожидаемое»). Это свидетельствует о том, что существуют отдельные активно взаимодействующие регионы как внутри активного, так и неактивного компартментов, но большинство частот контактов внутри компартментов в 1,1–1,3 раза выше, чем ожидаемое, в то время как для млекопитающих эти соотношения составляют около 3,5–3,8 [237]. Частота контактов между локусами, принадлежащими разным компартментам, снижена относительно ожидаемого в 1,2–1,4 раза, что также значительно

меньше, чем у млекопитающих. Был также построен график силы компартментов (см. раздел 2.12), представленный на Рисунке 266). Полученный график показывает, что сила компартментов в *D. discoideum* составляет 4, в сравнении с 12-15 для человека (оценка была проведена автором на данных из работы [187]).

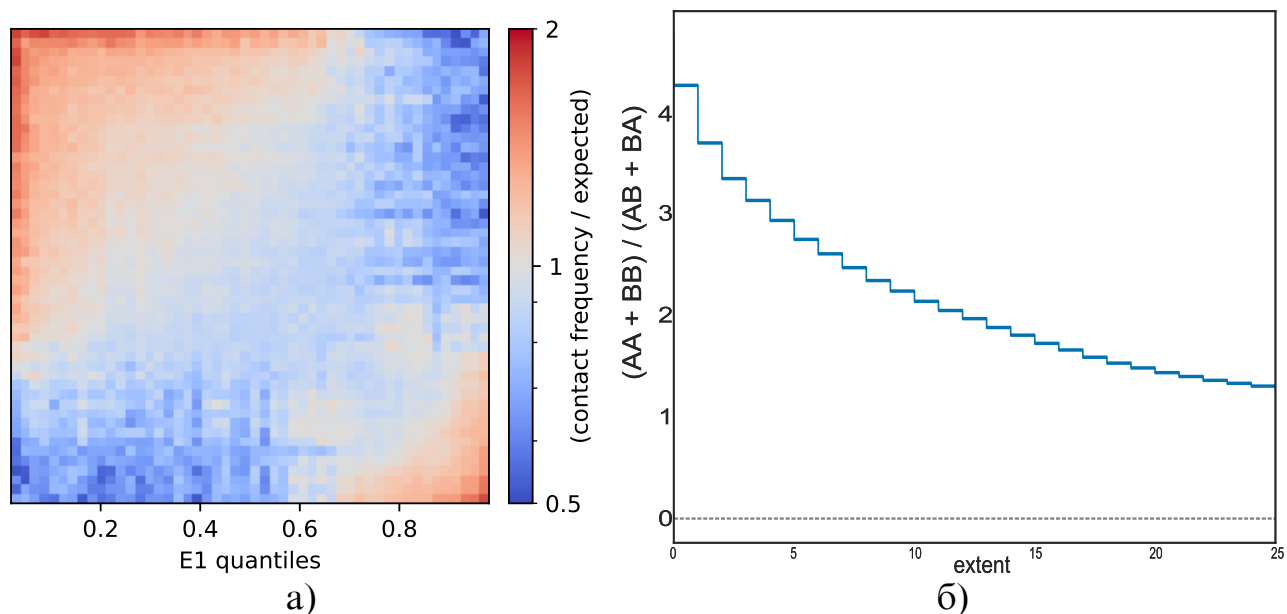


Рисунок 26 — Сила компартментов хроматина *D. discoideum* на 0 часе развития. (а) Седловая диаграмма. Левый верхний и правый нижний углы соответствуют взаимодействиям локусов, принадлежащих к одному компартменту. Левый нижний и правый верхний углы графика соответствуют взаимодействиям локусов, принадлежащих к разным компартментам. (б) График силы компартментов. Ось x — размер окна, для которого рассчитано значение силы компартментов, ось y — значение силы компартментов, рассчитанное как отношение сумм взаимодействия локусов внутри одного компартмента и взаимодействия локусов между компартментами.

На основании проведенных анализов можно заключить, что компартменты в хроматине *D. discoideum* очень слабые, т.е. плохо разделены между собой и слабо компактизованы. По-видимому, столь слабый эффект связан с активной экструзией, которая, по данным литературы, может ослаблять силу компартментов [128; 129].

3.1.6 Анализ точечных цис-взаимодействий *D. discoideum*

Визуальная оценка карт контактов показывает, что в хроматине *D. discoideum* присутствуют множественные точечные взаимодействия. Для них характерно наличие общего основания с ближайшим соседом, что проявляется как последовательное расположение точек на карте (левая панель Рисунка 27). Эти точечные взаимодействия устойчивы к сэмплированию карты контактов (правая панель Рисунка 27), что свидетельствует о том, что наблюдаемый сигнал не является шумом.

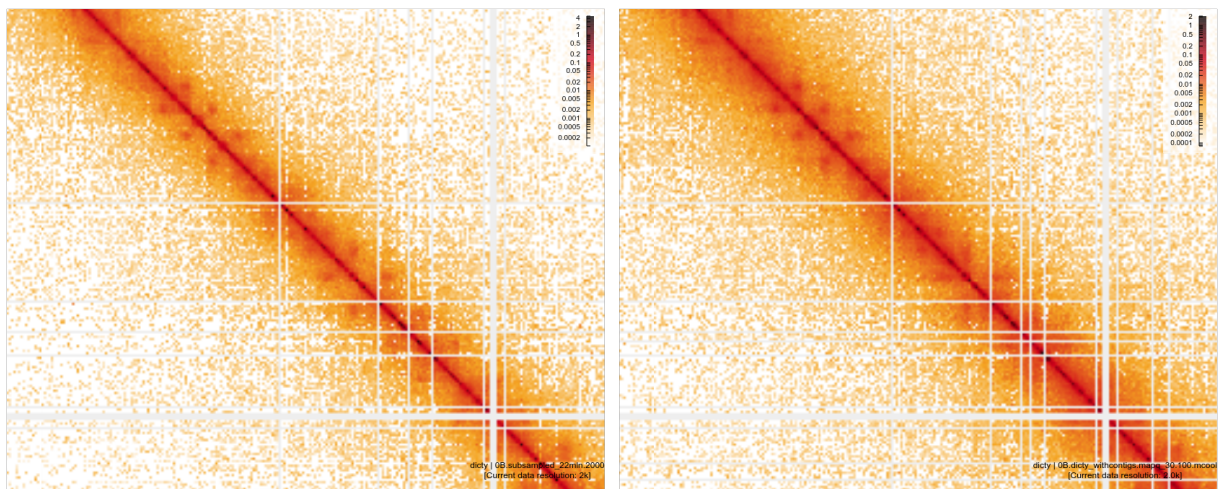


Рисунок 27 — Пример карты контактов *D. discoideum* в браузере HiGlass. Левая панель: исходная карта, 55 млн контактов, правая панель: сэмплированная карта, 46 млн пар. Размер бина матрицы — 2 тыс. п.н.

Для анализа этих контактов во всем геноме *D. discoideum* был построен график скейлинга (Рисунок 28). Полученные графики для всех часов развития демонстрируют отклонение от модели фрактальной глобулы (1.11) на линейном расстоянии между локусами около 20 тыс. п.н., что свидетельствует о повышении частот контактов между локусами на этом расстоянии и, тем самым, подтверждает результаты визуального анализа о наличии точечных взаимодействий по всему геному.

Для изучения особенностей этих точечных взаимодействий (в дальнейшем называемыми петлевыми) были использованы инструменты для аннотации петель LASCA [238] и chromosight [239]. Финальной разметкой был выбран набор петель, размеченный с помощью обоих инструментов одновременно; такой подход позволил снизить долю ложных находок, что

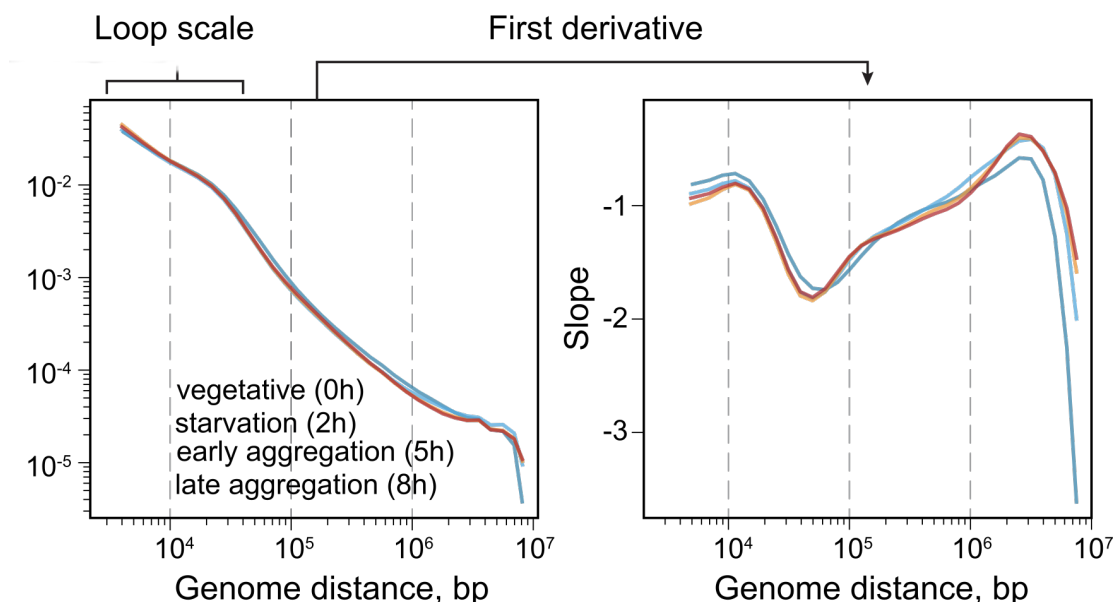


Рисунок 28 — График скейлинга и первой производной для 4 часов развития *D. discoideum*. Ось x — линейное расстояние между локусами, ось y — средняя частота контактов и их производная.

оценивалось в сравнении с визуальным анализом. Из аннотации петель следует, что большинство из них (75,5 %) действительно является последовательными (Рисунок 29а), а распределения их размеров не отличаются друг от друга на всех часах развития (Рисунок 29б).

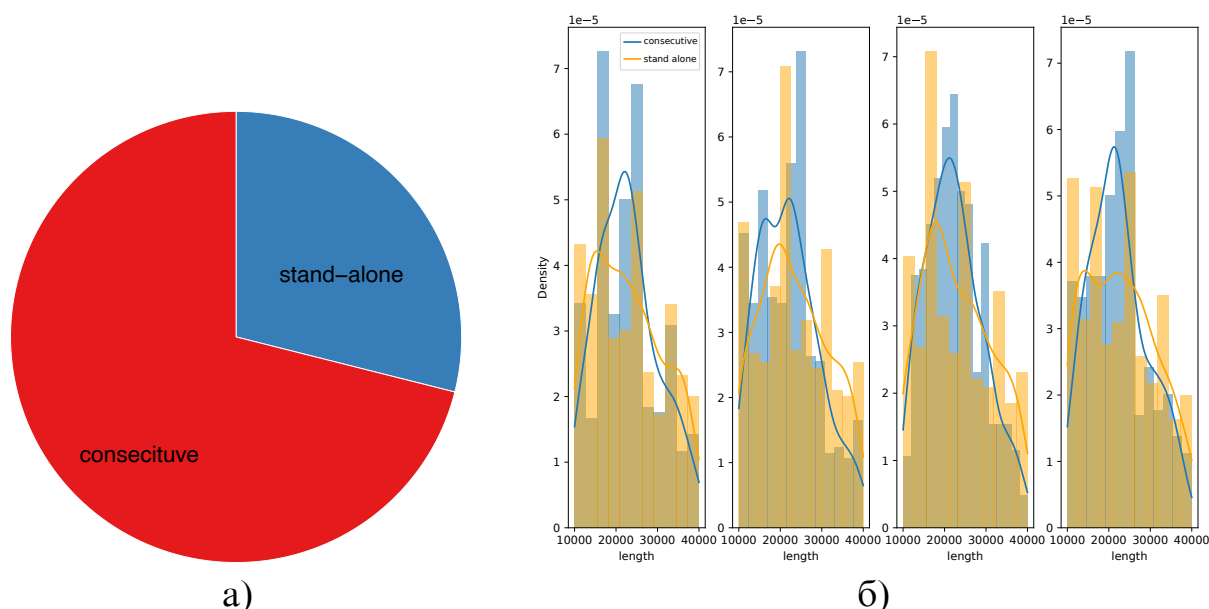


Рисунок 29 — Сравнительная характеристика количества и размеров последовательных и одиночных петель. (а) Круговая диаграмма количества последовательных (consecutive) и одиночных (stand-alone) петель. (б) Гистограммы размеров для четырех часов развития. Синее распределение: последовательные петли; оранжевое распределение: одиночные петли.

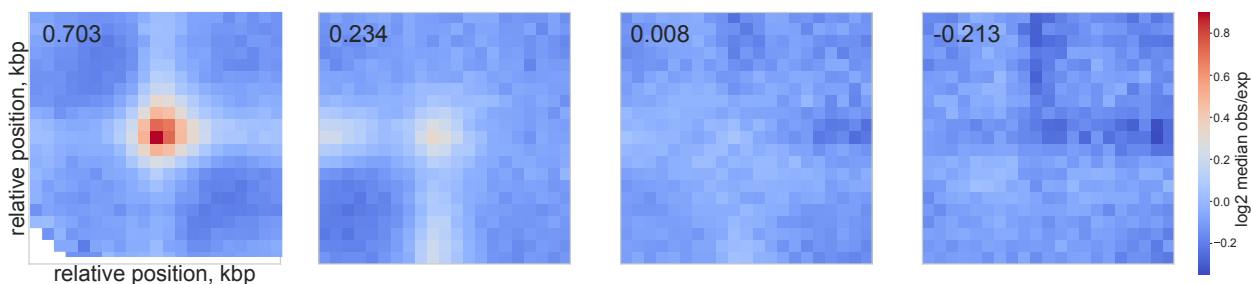


Рисунок 30 — Изучение иерархии в петлях хроматина *D. discoideum*. Слева направо представлены график средней петли первого, второго, третьего и четвертого порядков.

У млекопитающих петли часто являются иерархическими, когда две петли меньшего размера являются частью другой, большей, при этом последняя имеет по одному общему основанию с меньшими петлями. Эта иерархичность является результатом большой процессивности когезина и частичной проницаемости барьеров экструзии. Визуальный анализ карт хроматина *D. discoideum* показывает отсутствие иерархии у петель. Для подтверждения визуальных наблюдений был проведен анализ средней хроматиновой петли разных уровней иерархии. Для этого петлями первого порядка названы петли, полученные из разметки; петлями второго порядка — те, для которых 5'-основание взято из одной петли, а 3'-основание из соседней. Для петель третьего и четвертого порядков 3'-основание взято из петли, находящейся через одну или через две от петли, чье 5'-основание взято в анализ. Была построена средняя карта контактов, центрированная вокруг петель. Результат такого анализа показывает (Рисунок 30), что сила петли, рассчитанная как усреднение в окне 3×3 по позициям первого порядка, в 1,63 раза выше, чем ожидаемое (0,705 в логарифмической шкале), в то время как для петель второго порядка сила петли составляет 1,17, что может свидетельствовать об отдельных проскоках экструдера. В то же время интенсивность петли третьего порядка неотличима от ожидаемого, а интенсивность петли четвертого порядка даже ниже ожидаемого. Совместно, полученные результаты свидетельствуют об отсутствии иерархии петель *D. discoideum*, что, возможно, связано с низкой процессивностью когезина или с очень низкой проницаемостью барьеров экструзии.

Визуальный сравнительный анализ карт контактов показывает высокую консервативность петель между часами развития. Для численной оценки этого наблюдения был использован коэффициент Жаккара (Рисунок 31a).

Значения коэффициента получились не менее 0,76 для двух последовательных часов развития, поскольку недопредсказанность отдельных петель снижает значения коэффициента Жаккара. Мы заметили также, что при аннотации петель, визуальнo расположенных в неизменных позициях на картах разных часов развития, алгоритмы могут ошибаться на 1 бин при определении границ. Для преодоления этих смещений был разработан показатель Loop Anchor Intersection (LAI), рассчитанный как доля оснований петель для одного часа развития в виде отрезков 2 тыс. п.н. (1 бин), которые пересекаются с основаниями петель на другом часе развития в виде отрезков длиной 6 тыс. п.н. (3 бина). Такое расширение границ петель необходимо, чтобы учесть вариабельность предсказания границ одной и той же петли на картах разных часов развития. Кроме того, поскольку здесь мы работаем с основаниями петель, отдельные крупные недопредсказанные петли будут вносить такой же вклад, что и небольшие по размеру. Итоговая матрица асимметрична, но значения в соответствующих ячейках похожи, что демонстрирует стабильность разработанного показателя. Результаты такого анализа показывают куда более высокие значения в сравнении с коэффициентом Жаккара — не менее 0,91 при попарном сравнении между часами развития (Рисунок 316), что свидетельствует о высокой консервативности петель между часами развития. Таким образом, мы разработали метрику консервативности границ петель между часами развития, которая устойчива к шуму и позволила нам получить значения, лучше соответствующие результатам визуального анализа.

В связи с высокой консервативностью петель на разных часах развития мы решили изучить возможную связь хроматиновых петель *D. discoideum* с расположением генов домашнего хозяйства. Гены домашнего хозяйства у *D. discoideum* были определены с использованием данных из работы [241] (поскольку они содержат данные экспрессии для всех точек развития) в соответствии с процедурой, описанной в работе [242] (подробнее — см. раздел 2.15). Анализ показал, что внутри петель гены домашнего хозяйства встречаются чаще в сравнении со случайными регионами (Рисунок 32а), в то время как покрытие всеми генами или дифференциально экспрессируемыми генами неотлично от покрытия случайными регионами (Рисунки 32б и 32в). Полученные результаты позволяют говорить о возможной роли хроматиновых

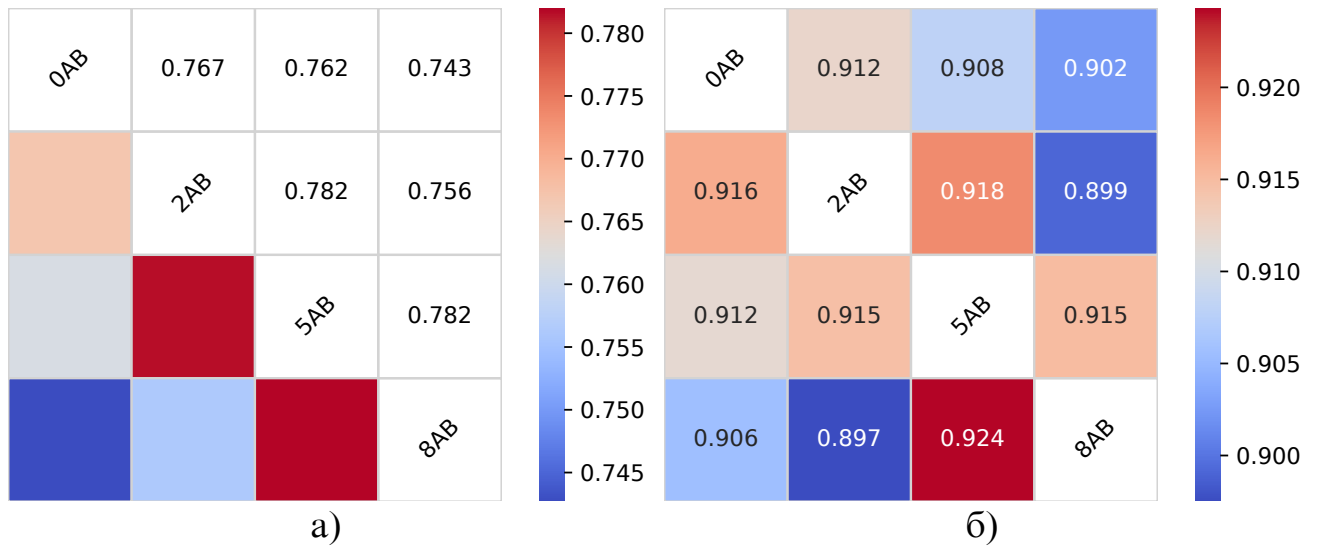


Рисунок 31 — Консервативность петель между часами развития *D. discoideum*. (а) Коэффициенты Жаккара. (б) Показатели Loop Anchor Intersection.

петель в поддержании экспрессии генов домашнего хозяйства, но не в поддержании дифференциальной экспрессии между часами развития.

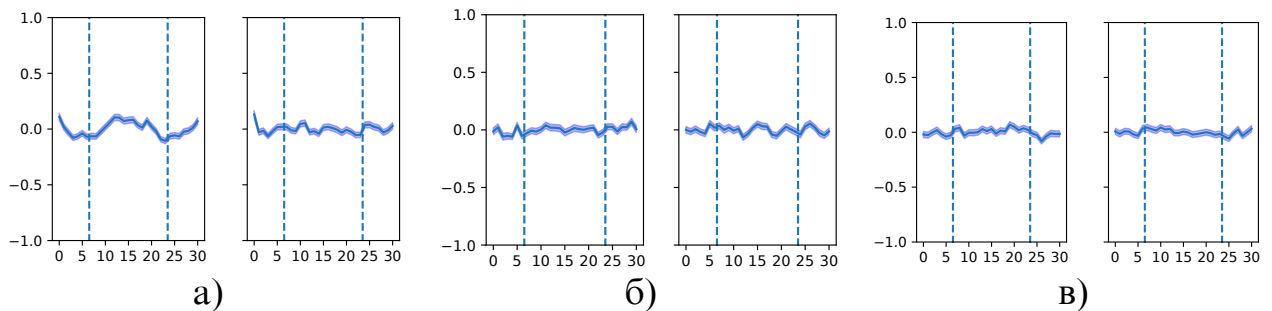


Рисунок 32 — Изучение покрытия генами разных категорий вокруг петель. (а) Покрытие генами домашнего хозяйства петель в сравнении с контролем. (б) Покрытие всеми генами петель в сравнении с контролем. (в) Покрытие дифференциально экспрессируемыми между 0 и 8 часами развития генами петель в сравнении с контролем. Левая панель: гены разных категорий. Правая панель: контроль. Границы петель обозначены пунктирными линиями. Медианные покрытия генами каждой категории были z-трансформированы и отрисованы вокруг петель.

3.1.7 Анализ синтенных блоков между *D. discoideum* и *D. purpureum*

Для косвенной оценки функциональной важности хроматиновых петель у *D. discoideum* мы решили использовать синтенные блоки, представляющие собой регионы, в которых у родственных видов сохранился порядок генов; такой же порядок генов был у общего предка. Поскольку *D. discoideum* и *D. purpureum* разделяют около 400 миллионов лет эволюции, колокализация синтенных блоков и петель может свидетельствовать о важности самих этих петель для транскрипционной программы клеток. Для изучения консервативности петель *D. discoideum* были использованы данные из работы [252], в которых определены синтенные блоки между двумя видами рода *Dictyostelium*: *D. discoideum* и *D. purpureum*.

Визуальный анализ показал, что границы синтенных блоков часто соответствуют границам петель (Рисунок 33а), а отдельные синтенные блоки покрывают сразу несколько петель. Пермутационный тест свидетельствует о более частой колокализации границ синтенных блоков и границ петель, чем ожидается для структур такого количества и размеров (Рисунок 33б). Таким образом, проведенный анализ свидетельствует о том, что петли часто являются частью синтенных блоков, а границы петель и границы синтенных блоков статистически значимо колокализуются.

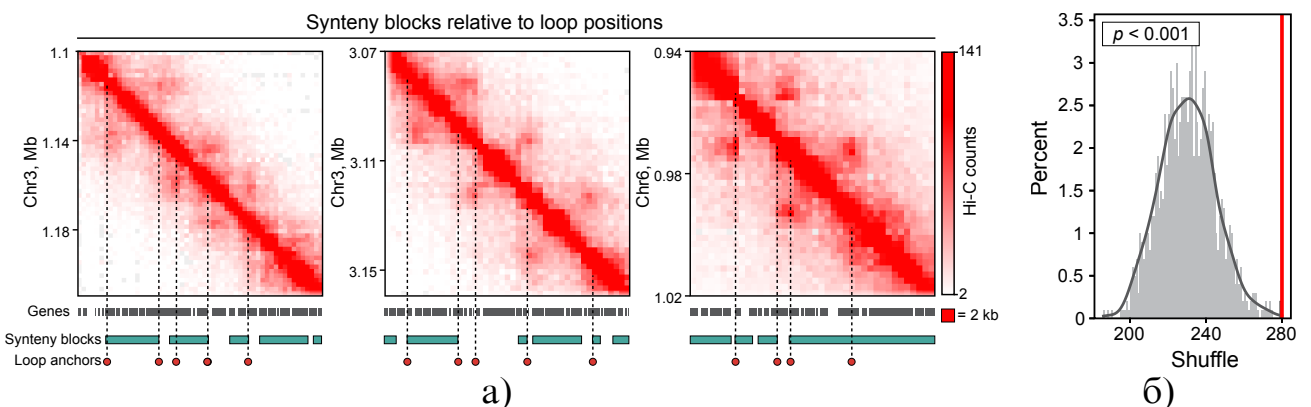


Рисунок 33 — Колокализация синтенных блоков и петель *D. discoideum*. (а) Примеры колокализации синтенных блоков и границ петель *D. discoideum*. Серыми отрезками под картой контактов показаны гены, их границы визуально сливаются из-за близкого расположения. (б) Численная оценка колокализации границ синтенных блоков и границ петель *D. discoideum*, выполненная с помощью пермутационного теста.

3.1.8 Сравнительный анализ оснований последовательных и одиночных петель

Для того, чтобы выяснить, обладают ли последовательные и одиночные петли разными свойствами, мы провели сравнительный анализ оснований последовательных и одиночных петель. Контролем были выбраны случайные участки генома в том же количестве и той же медианной длины. В частности, нас интересовало сравнение профилей покрытия генами, лежащими на разных цепях ДНК, бинов вокруг оснований последовательных и одиночных петель. Оказалось, что медианные значения покрытия генами разных цепей в соседних с основаниями бинах составляют 0,7 и 0,6 для оснований последовательных и одиночных петель соответственно, что в 2 раза выше, чем в контроле (Рисунки 34а, 34б).

Кроме того, мы заметили, что профили покрытия генами, лежащими на плюс-цепи и минус-цепи ДНК, зеркально симметричны относительно основания петель. Это наблюдение свидетельствует о неслучайной ориентации генов на цепях вокруг оснований петель, а именно, что гены ориентированы конвергентно. Для оценки этого наблюдения мы использовали показатель конвергентности (подробнее о процедуре в разделе 2.20). Показатель конвергентности позволяет оценивать направления генов на цепях ДНК в заданном окне (20 тыс. п.н.) вокруг каждого бина генома, причем чем выше показатель, тем больше генов на плюс-цепи расположено левее от бина, для которого рассчитывается показатель, и больше генов на минус-цепи ДНК правее бина, для которого рассчитывается показатель. В результате оказалось (Рисунок 34в), что основания последовательных петель имеют более выраженный пик показателя конвергентности (медианные значения достигают 1,5), что приблизительно в 3 раза выше, чем соответствующий показатель для одиночных петель. Важно также отметить, что медианный профиль показателя конвергентности вокруг контроля представляет собой ровную линию около нуля, что свидетельствует о том, что пик показателя конвергентности для одиночных петель также отличен от случайных значений. Таким образом, проведенный анализ свидетельствует о роли конвергентно ориентированных генов в образовании границ петель, причем пик показателя

конвергентности значительно более выражен для оснований последовательных петель, чем для оснований одиночных петель.

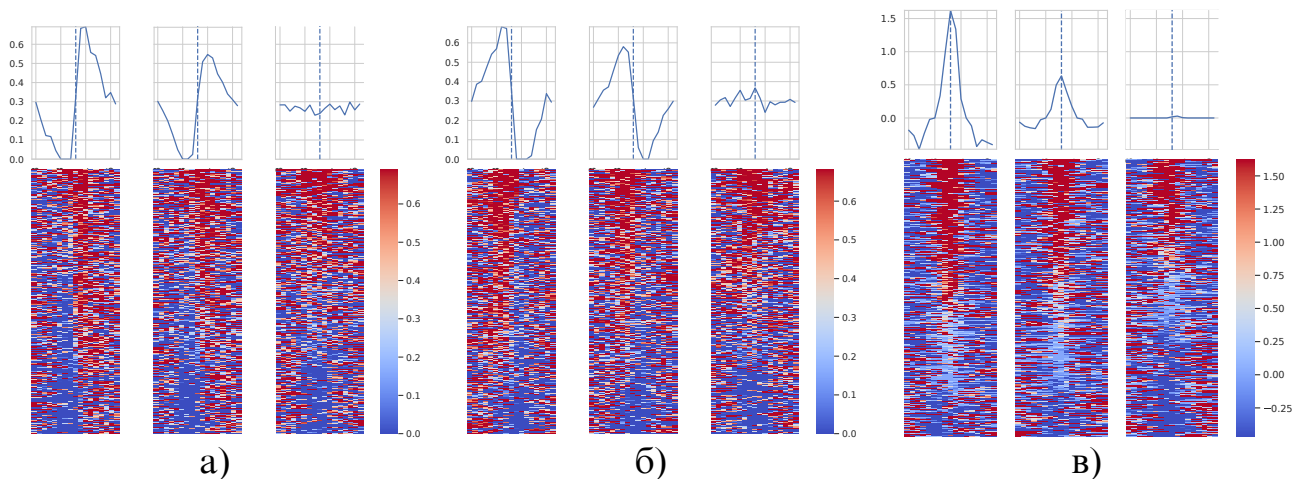


Рисунок 34 — Профиль доли покрытия бинов генами, лежащими на разных цепях ДНК, и значений показателя конвергентности в бинах вокруг оснований последовательных и одиночных петель в сравнении с контролем. (а) Профиль покрытия генами минус-цепи бинов вокруг оснований последовательных и одиночных петель в сравнении с контролем. (б) Профиль покрытия генами плюс-цепи бинов вокруг оснований последовательных и одиночных петель в сравнении с контролем. (в) Профиль показателей конвергентности вокруг оснований последовательных и одиночных петель в сравнении с контролем. Верхняя панель: медианные значения показателей вокруг оснований петель; нижняя панель: тепловые карты профиля сигнала вокруг каждого отдельного основания петель. Пунктиром обозначены основания петель.

Возможным объяснением этих различий может быть принципиально иной механизм формирования петель двух категорий, ошибки алгоритма при детектировании одиночных петель или же роль некоторого дополнительного фактора, в проведенном анализе не учтенного. Для ответа на этот вопрос мы обратились к принципиально иному феномену.

3.1.9 Анализ асимметричных вытянутых петель

При визуальном анализе хроматина *D. discoideum* с помощью Hiclass [230] мы заметили, что, помимо симметричных петель, LASCA [228] также ан-

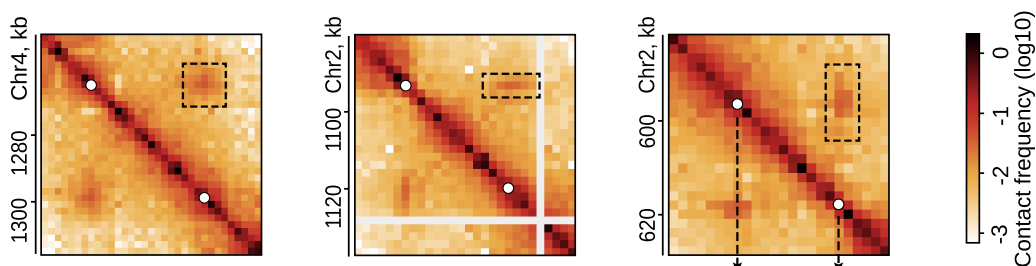


Рисунок 35 — Примеры симметричных и вытянутых петель в матрице хроматиновых контактов. Левая панель: симметричная петля, центральная панель: петля, вытянутая вдоль 5'-основания, правая панель: петля, вытянутая вдоль 3'-основания. Петли выделены черным пунктиром.

нотирует асимметричные петли, вытянутые как вдоль 5'-основания, так и 3'-основания (примеры см. на Рисунке 35).

Для того, чтобы выделить эти удлиненные петли (названные «el-петли») из существующей аннотации хроматиновых петель *D. discoideum*, мы разработали простой алгоритм, основанный на отношении частот контактов 5'-основания с регионами, фланкирующими 3'-основание, по сравнению с частотами контактов 3'-основания с регионами, фланкирующими 5'-основание (см. раздел 2.17). Пороговые значения были выбраны как 20- и 80-перцентили распределения отношения частот контактов для вытянутых как вдоль 5'-основания, так и вдоль 3'-основания.

Чтобы убедиться, что алгоритм работает правильно, мы построили среднюю карту, центрированную на вытянутых петлях, с помощью `coolpuppy` [253]. Полученные графики свидетельствуют, что нам удалось успешно отделить 3'- и 5'-вытянутые петли от симметричных (Рисунок 36). Итоговые списки содержат в среднем 550 петель для каждого из четырех часов развития.

Далее мы провели сравнительный анализ симметричных петель и el-петель, чтобы лучше понять их свойства и, возможно, механизм образования. Для этого мы проверили, как распределяется покрытие генами, лежащими на разных цепях, вокруг петель разных групп. Для этого мы получили значения покрытия генами, лежащими на плюс- или минус-цепях ДНК, геномных бинов размером 500 п.н. Используя `deeptools` [254], мы построили профиль покрытия генами плюс и минус цепей вокруг петель (Рисунок 37). Мы обнаружили, что для всех типов петель график медианного покрытия оказывается

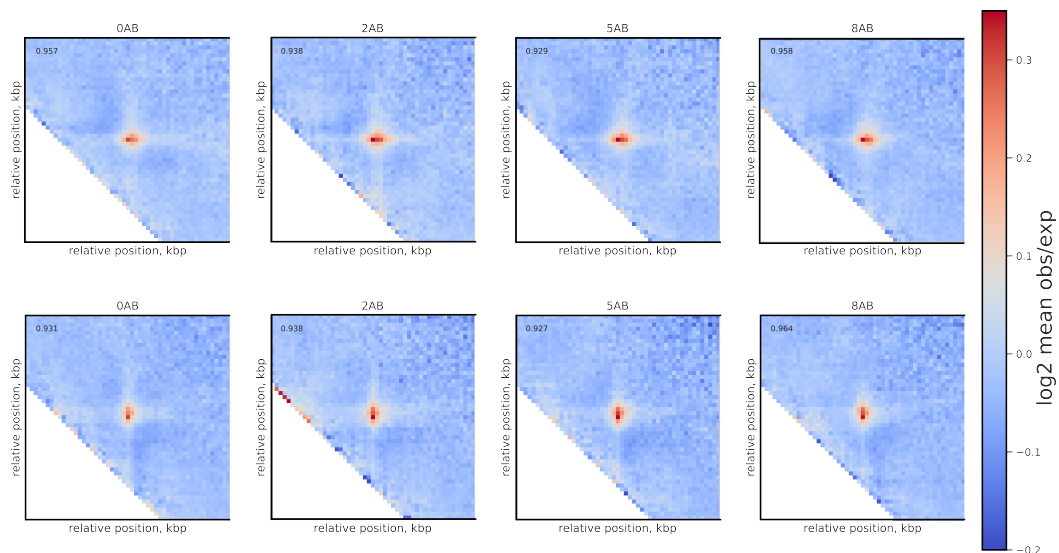


Рисунок 36 — Усредненная карта, центрированная на вытянутых петлях, для разных часов развития. Столбцы: 0, 2, 5 и 8 часы развития. Верхняя панель: петли, вытянутые вдоль 5'-оснований. Нижняя панель: петли, вытянутые вдоль 3'-оснований. Все значения нормированы на ожидаемое, усреднены между отдельными петлями и преобразованы в логарифмическую шкалу.

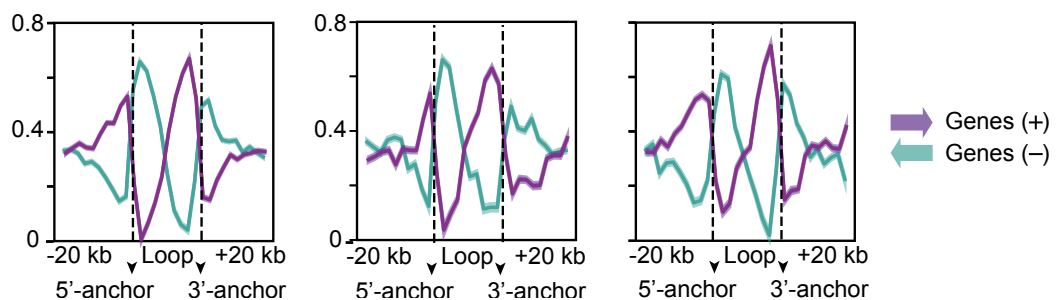


Рисунок 37 — График медианного покрытия генами плюс и минус цепи. Левая панель: симметричные петли, центральная панель: петли, вытянутые вдоль 5'-основания, правая панель: петли, вытянутые вдоль 3'-основания. Пунктиром обозначены основания петель.

схожим, свидетельствуя о том, что все типы петель, по-видимому, образованы конвергентно ориентированными генами.

Кроме того, мы провели анализ распределения значений экспрессии генов и показателя инсуляции вокруг петель. Для этого e1-петли были ориентированы так, чтобы петля была вытянута вдоль 5'-основания. Полученный график для показателя инсуляции (Рисунок 38a) свидетельствует о сильной инсуляции 5'-основания и слабой инсуляции 3'-основания. Эти результаты соотносятся также с профилем экспрессии (Рисунок 38б) вокруг петель, который демонстрирует более низкий уровень экспрессии в 3'-основаниях

e1-петель по сравнению с 5'-основанием e1-петель, а также — симметричным сигналом в обоих основаниях симметричных петель. Это свидетельствует о том, что сигнал петли вытягивается вдоль одного основания, если уровень транскрипции в другом ниже. Таким образом, предполагаемым барьером экстрезии в хроматине *D. discoideum* является активная транскрипция, что было показано для хроматина млекопитающих [187].

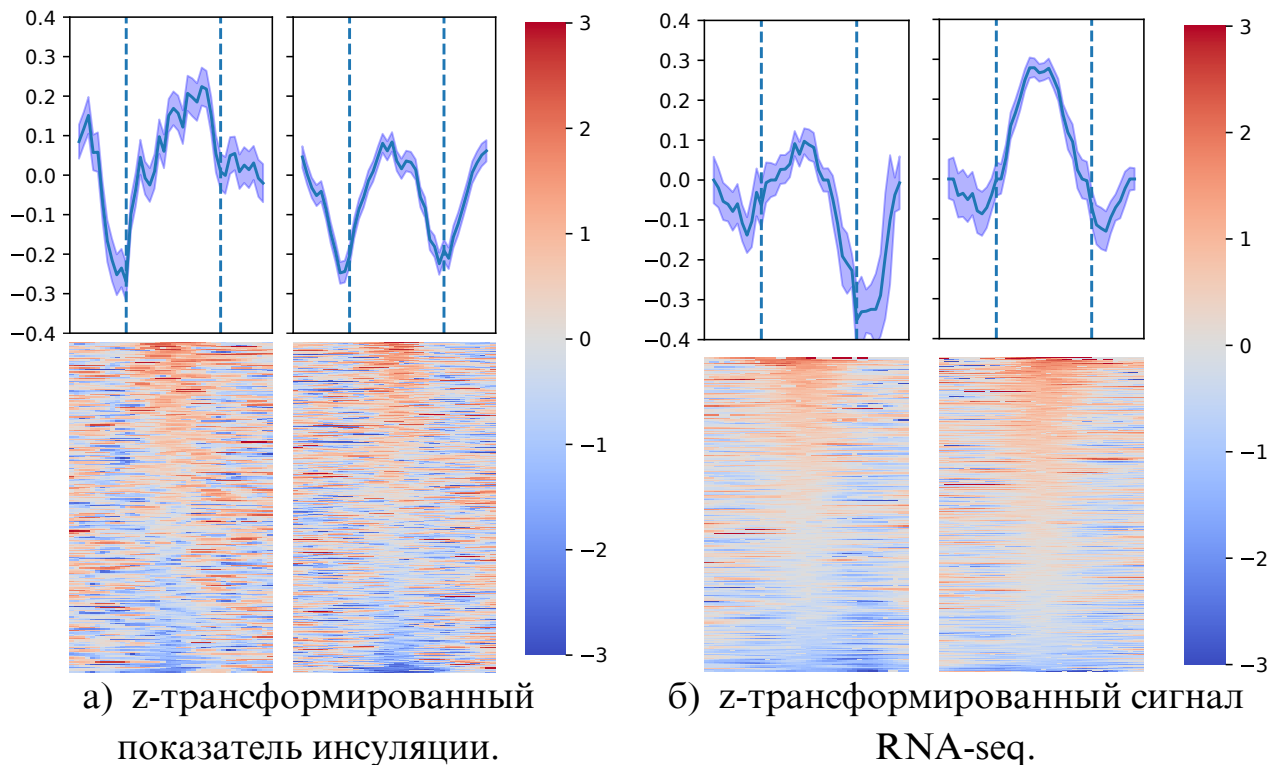


Рисунок 38 — Графики значений сигнала RNA-seq (а) и показателя инсуляции (б) вокруг вытянутых (левые панели) и симметричных (правые панели) петель. Тепловая карта показывает z-трансформированные значения вокруг отдельных петель, верхние панели — медианные значения этих параметров.

Пунктирами обозначены основания петель.

3.1.10 Взаимосвязь уровня транскрипции и силы петель

Полученные для вытянутых петель результаты свидетельствуют о влиянии уровня транскрипции на структуру петель. В то же время, мы показали повышение показателей конвергентности в основаниях петель по сравнению с контролем. Для более подробного понимания общих закономерностей

связи формы и силы петли и транскрипции, мы провели анализ уровней транскрипции конвергентно ориентированных генов вокруг оснований петель. Для этого был рассчитан показатель конвергентной транскрипции вокруг оснований как сумма уровня транскрипции генов, лежащих на плюс цепи, расположенных на 2-4 тыс. п.н. влево от основания петли, и уровня транскрипции генов, лежащих на минус цепи, расположенных на 2-4 тыс. п.н. вправо от основания петли. В результате каждой петле соответствуют два значения показателей конвергентной транскрипции вокруг 5'- и 3'-оснований петель, которые были отложены по y - и x -осям соответственно (Рисунок 39а). Каждая петля представляет собой точку, расположенную на графике в соответствии со значениями экспрессии вокруг оснований. Сила петли рассчитана как сумма нормализованной на ожидаемое частоты контактов в окне 3×3 бина размером 2 тыс. п.н. каждый и показана на графике цветом соответствующей точки (Рисунок 39а).

Для того, чтобы разделить все петли на группы по уровню транскрипции вокруг оснований, мы построили графики плотности распределения значений экспрессии для каждого основания отдельно и производных плотности (верхние и боковые панели у графика точек на Рисунке 39а). Значения экспрессии вокруг основания, соответствующие изменению знака первой производной, были использованы как пороги для деления на группы. Таким образом, петли были разделены на четыре группы в соответствии с уровнями конвергентной транскрипции вокруг оснований. Для каждой группы был отрисован график распределения силы петли (правая панель Рисунка 39а). Рассчитанные с помощью одностороннего теста Манна-Уитни p -значения для силы петли в разных группах демонстрируют высокую значимость отличий между группами (Рисунок 39а).

Аналогичный анализ был проведен для дивергентно ориентированных генов, т.е. был рассчитан показатель дивергентной транскрипции вокруг оснований как сумма уровня транскрипции генов, лежащих на минус-цепи ДНК, расположенных на 2-4 тыс. п.н. влево от основания петли, и уровня транскрипции генов, лежащих на плюс-цепи ДНК, расположенных на 2-4 тыс. п.н. вправо от основания петли. Затем петли были разделены на четыре группы на основании значений дивергентной транскрипции вокруг оснований, в которых первая производная плотности распределения меняет знак (Рисунок 39б), и для сил петель между группами были рассчитаны

p-значения с помощью одностороннего теста Манна-Уитни; статистически значимых отличий обнаружено не было. Проведённый анализ свидетельствует в пользу наличия связи между уровнем конвергентной транскрипции вокруг оснований петель и силой петель, а также отсутствием такой связи между дивергентной транскрипцией вокруг оснований петель и силой петель.

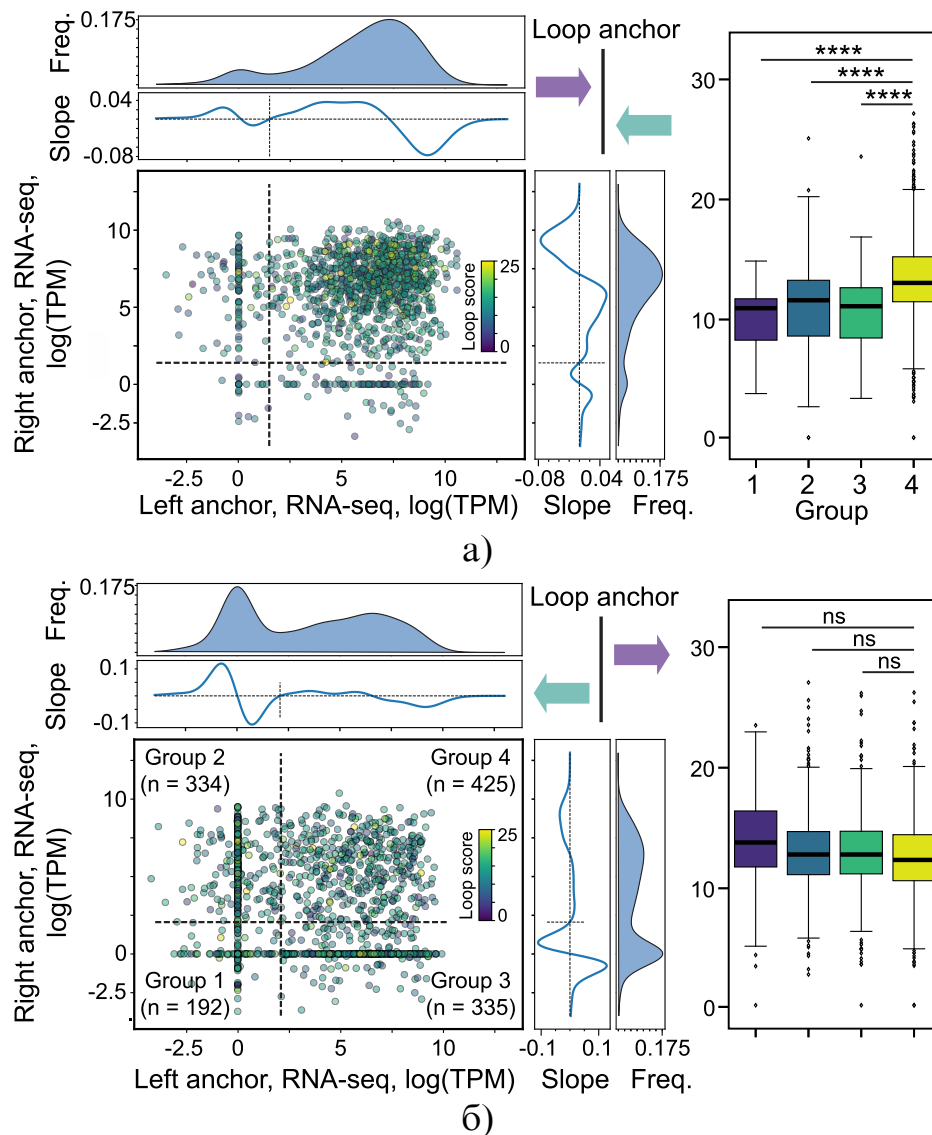


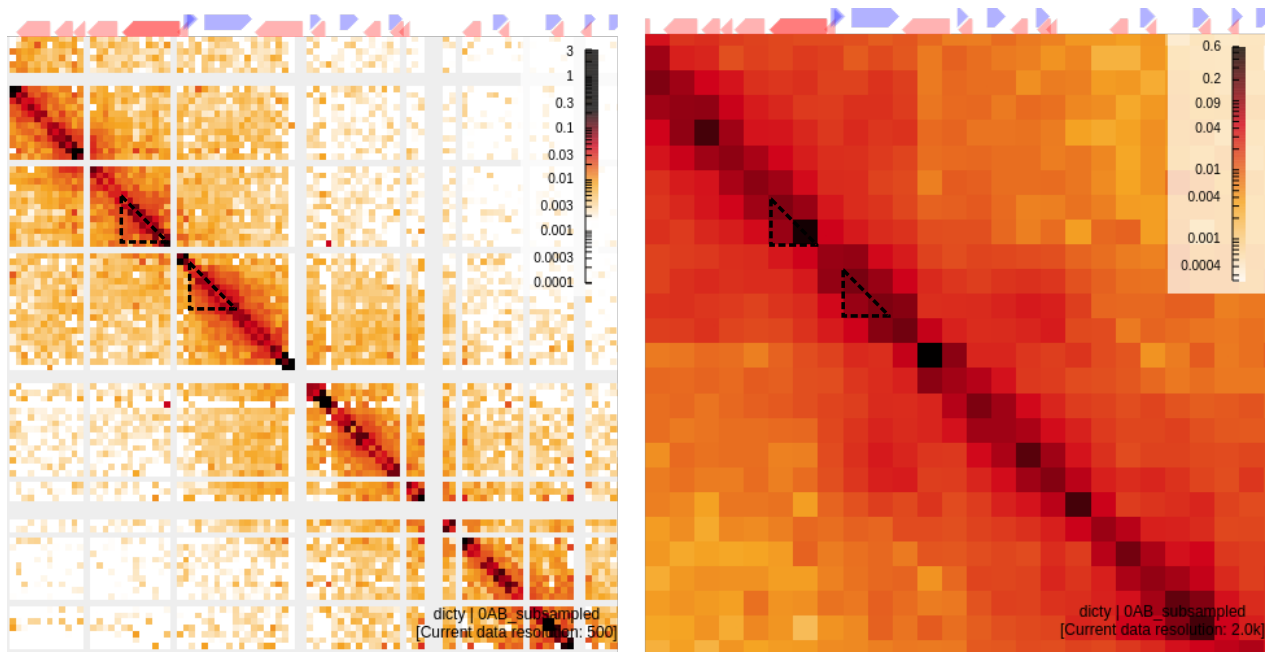
Рисунок 39 — Зависимость силы петли от уровня экспрессии в основаниях. (а) Уровень конвергентной транскрипции вокруг оснований петель и его связь с силой петель. (б) Уровень дивергентной транскрипции вокруг оснований петель и его связь с силой петель. Ось x — сумма конвергентной транскрипции вокруг 3'-оснований петель, ось y — сумма конвергентной транскрипции вокруг 5'-оснований петель.

Ранее нами было отмечено, что анализ петель хроматина необходимо производить на разрешении 2 тыс. п.н. Однако визуальный анализ карты

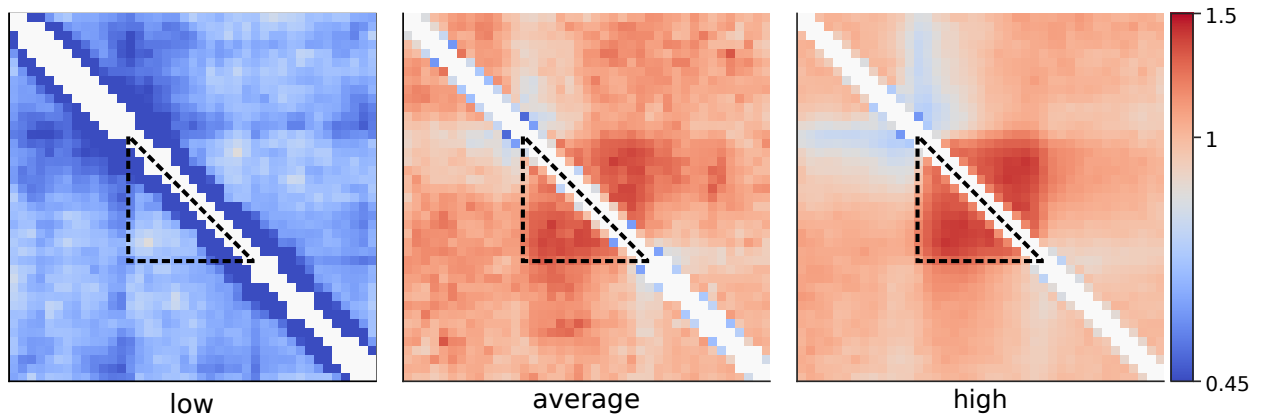
контактов на разрешении 500 п.н. свидетельствует о повышении частоты контактов в отдельных участках хроматина около диагонали (левая панель на Рисунке 40a). Такие участки хроматина с повышенными частотами взаимодействия между друг другом в литературе называют домены хромосомных взаимодействий (chromosomal interaction domains — CIDs [255—257]). При этом сравнение одних и тех же доменов на разрешении 500 п.н. и 2 тыс. п.н. (правая панель на Рисунке 40a), показывает, что на разрешении 2 тыс. п.н. домены сливаются с диагональю.

Мы также заметили, что часто границы контактных доменов совпадают с границами отдельных генов. Контактный домен представляет собой участок хроматина с повышенной частотой контактов внутри него в сравнении с окружением, визуально выглядит как треугольник около диагонали на матрице Hi-C. Принимая во внимание наблюдаемую связь между уровнем транскрипции и силой петель, мы изучили связь между частотами контактов внутри домена (косвенное свидетельство его плотности) и уровнем транскрипции генов. Для этого мы разделили все гены на три группы по уровню транскрипции и с помощью `coolpupru` отрисовали среднюю карту хроматиновых взаимодействий вокруг генов (Рисунок 40б). Результаты анализа показали, что гены с низким уровнем транскрипции в среднем не образуют контактного домена; гены группы средней и высокой транскрипции образуют контактный домен, причем для генов с высоким уровнем транскрипции частота контактов по телу гена наибольшая. Таким образом, мы показали, что в хроматине *D. discoideum* образуются контактные домены, частоты контактов в которых связаны с уровнем транскрипции гена, формирующего домен. Одной из предполагаемых функций такой структуры может быть облегчение перемещений РНК-полимеразы II из 3'-области в промотор экспрессируемого гена для повторного использования полимеразы [187].

Полученные в разделе 3.1.8 результаты свидетельствуют, что именно конвергентная ориентация генов важна для формирования основания петель. Поэтому далее мы сосредоточили внимание на генах, образующих конвергентные пары. Конвергентными парами (далее — КПГ) были названы пары генов, которые расположены на цепях в конвергентной ориентации, расстояние между 3'-концами этих генов не более 8 тысяч п.н., и такие гены являются ближайшими соседями друг друга. В результате мы получили 4672 конвергентные пары генов. Пермутационный тест показал, что КПГ статистически



а)



б)

Рисунок 40 — Контактные домены, образованные конвергентными парами генов. (а) Пример региона карты Hi-C на разрешении 500 п.н. (левые панели) и 2 тыс. п.н. (правые панели). Хромосома 3, 50 тыс. п.н. Домены выделены черными пунктирами. (б) Графики средней частоты контактов для генов, разделенных на три группы в зависимости от уровня транскрипции в них.

значимо перепредставлены в основаниях петель ($p < 0,001$, Рисунок 41, левая панель) и недопредставлены во внутренних частях петли ($p < 0,001$, Рисунок 41, правая панель). В дальнейшем КППГ, расположенные в геномном бине размером 2 тыс. п.н., идентифицированном как основание петли, и/или в его окрестности размером 4 тыс. п.н., называются КППГ, ассоциированными с основанием петли (оп-КППГ). Кроме того, мы аннотировали дивергентные пары генов с помощью аналогичной процедуры за исключением выбора ге-

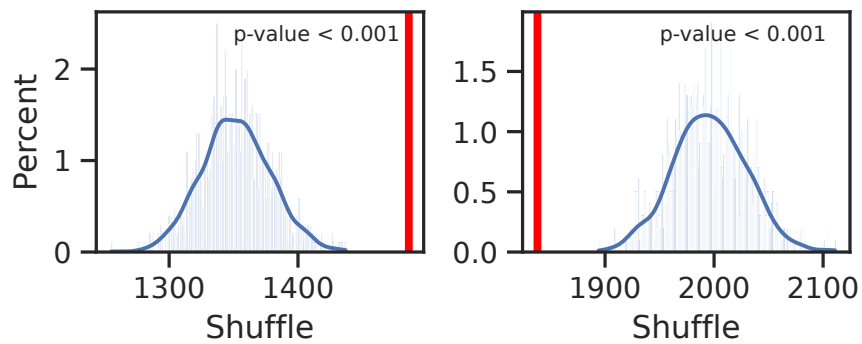


Рисунок 41 — Количество конвергентных пар генов, пересекающих основания петель и внутренние части петель. Левая панель: анализ КПГ, пересекающих основания петель. Правая панель: анализ КПГ, пересекающих внутренности петель. Синее распределение — распределение количества КПГ, пересекающих основания или внутренности петель в пермутациях, красная линия — количество КПГ, пересекающих основания или внутренности петель в реальных данных. Ось x — количество КПГ, пересекающих основания или внутренности петель, ось y — процент пермутаций с данным числом пересечений.

нов противоположной ориентации, и использовали этот набор пар в качестве контроля.

Далее был проведен анализ влияния транскрипции генов, входящих в оп-КПГ, на контакты хроматина. Для этого мы разделили гены, входящие в оп-КПГ, на высокоэкспрессируемые и низкоэкспрессируемые, используя пороговые значения 8 и 2 Transcripts Per Million (TPM) соответственно, рассчитанные по Формуле 2.2. Мы получили три группы пар: оба гена с низким уровнем транскрипции (нн-КПГ), оба гена с высоким уровнем транскрипции (вв-КПГ) и один ген с высоким уровнем транскрипции, а другой с низким (вн-КПГ). Для полученных трех категорий были построены карты средних хроматиновых взаимодействий, причем последний тип пар был ориентирован так, чтобы высокоэкспрессируемый ген находился с 5'-стороны пары (Рисунок 42). Анализ показал, что вв-КПГ организованы в контактные домены, границы которых идеально совпадают с промоторами генов (центральная верхняя панель Рисунка 42). Кроме того, была выявлена умеренная частичная и, что самое главное, односторонняя деградация контактного домена для вн-КПГ с потерей контактов именно на стороне слабоэкспрессируемого гена (верхняя правая панель Рисунка 42). Ни пары слабоэкспрессируемых или молчащих конвергентных генов, ни ди-

вергентные пары генов, независимо от уровня их экспрессии, не образуют контактных доменов и полос (верхняя левая и три нижние панели Рисунка 42).

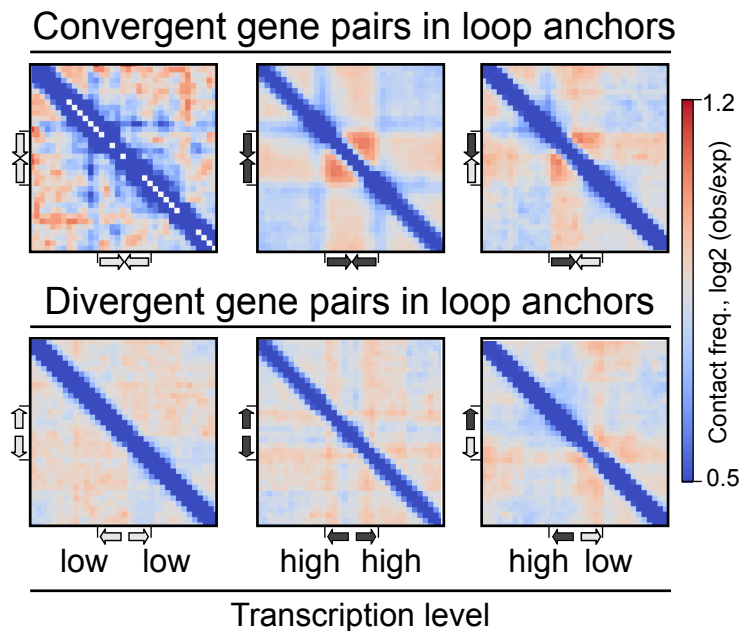


Рисунок 42 — Домен, образованный на телах конвергентных генов с разными уровнями транскрипции. Верхняя панель: средняя карта Hi-C, центрированная на КПГ с низкой, высокой и с низкой и высокой экспрессией; нижняя панель: средняя карта Hi-C, центрированная на дивергентных парах генов с низкой, высокой и с низкой и высокой экспрессией.

После этого мы изучили эффект уровней транскрипции КПГ на больших геномных расстояниях и повторили анализ с длиной фланкирующего региона равной 50 тыс. п.н. (Рисунок 43а). Результаты анализа показали, что вв-КПГ образуют контактные полосы («stripe» на Рисунке 43а) [131; 187; 258] как с 5'-конца (upstream), так и с 3'-конца (downstream). Кроме того, слабоэкспрессируемый ген из вн-КПГ теряет контакты с участком выше пары (со стороны высокоэкспрессируемого гена; третья панель Рисунка 43а, черная стрелка). Мы также наблюдали асимметричный прирост контактов между участками, расположенными с 5'-конца и 3'-конца вн-КПГ, поскольку участки, расположенные с 5'-конца вн-КПГ, увеличивают частоту взаимодействия с регионами, непосредственно примыкающими с 3'-конца к вн-КПГ (со стороны слабоэкспрессируемого гена; третья панель Рисунка 43а, пунктирная область). Наконец, контактная полоса с 3'-конца пары значительно ослаблена в вн-КПГ по сравнению с вв-КПГ, тогда как контактная полоса с 5'-конца

пары, как и профиль инсуляции в вн-КПГ сохраняются. Для оценки наблюдаемых изменений было произведено вычитание логарифмированных карт вн-КПГ и вв-КПГ (правая панель Рисунка 43а). На основании полученных результатов была построена схема (Рисунок 43б), иллюстрирующая изменение средних частот контактов для КПГ с разными уровнями экспрессии.

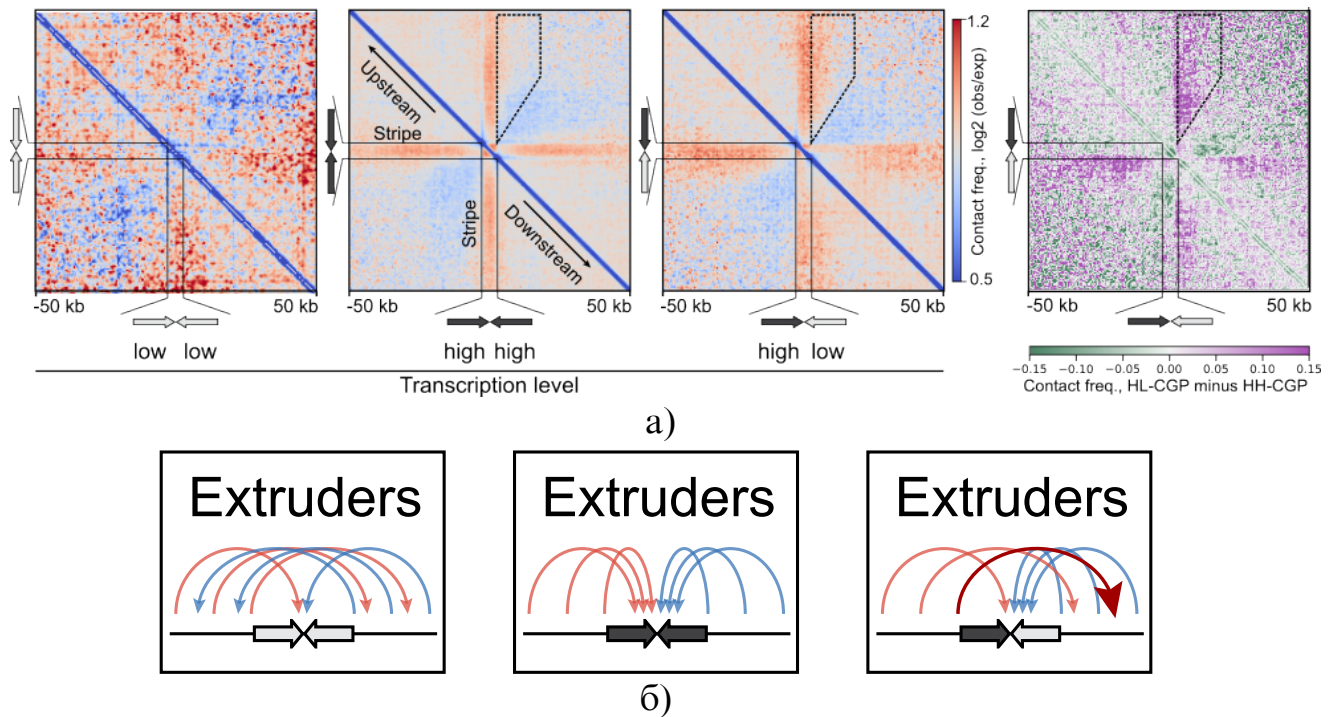


Рисунок 43 — Влияние уровня транскрипции пары конвергентных генов на частоты контактов в окне 100 тыс. п.н. (а) Средняя карта хроматинового окружения конвергентных пар генов. Панели слева направо: средняя карта хроматинового окружения нн-КПГ, вв-КПГ и вн-КПГ; результат вычитания средних карт контактов вокруг вн-КПГ и вв-КПГ. Три левые панели содержат нормализованные частоты контактов между регионами, правая — результат вычитания средних карт контактов вокруг вн-КПГ и вв-КПГ. (б) Схема экструзии для трех групп уровней транскрипции. Красными и синими стрелками обозначены контакты между областью со стороны 5'- и 3'-концов КПГ, соответственно, с областями, обозначенным концом стрелки. Яркая стрелка крупнее остальных обозначает контакты регионов в выделенном четырехугольнике. Ген с высокой экспрессией обозначен стрелкой с черной заливкой, ген с низкой экспрессией обозначен стрелкой с белой заливкой.

Проведенный анализ свидетельствует, что пара генов с низкой экспрессией слабо тормозит когезин (отсутствует контактная полоса), пара

высокой экспрессией является высокоэффективным барьером для экстрезии (симметричная контактная полоса), а пара генов с высокой и с низкой экспрессией является, по-видимому, сильным барьером со стороны гена с высокой экспрессией и слабым барьером со стороны гена с низкой экспрессией, о чем свидетельствует асимметричность контактной полосы. Описанная модель изображена на Рисунке 436. Полученные результаты свидетельствуют о важности не только конвергентной ориентации генов, но и их уровня транскрипции для образования основания петли.

3.1.11 Анализ регуляторных последовательностей *D. discoideum*

Для изучения регуляторных областей генома была проведена аннотация энхансер-подобных элементов (далее — энхансеры) путем пересечения пиков ATAC-seq и ChIP-seq для гистоновых модификаций H3K4me1 и H3K27ac (схема поиска на Рисунке 11). Подавляющее большинство полученных энхансеров располагаются внутри тел генов, что, как минимум частично, является результатом высокой плотности генов у *D. discoideum* (несколько примеров приведены на Рисунке 44).

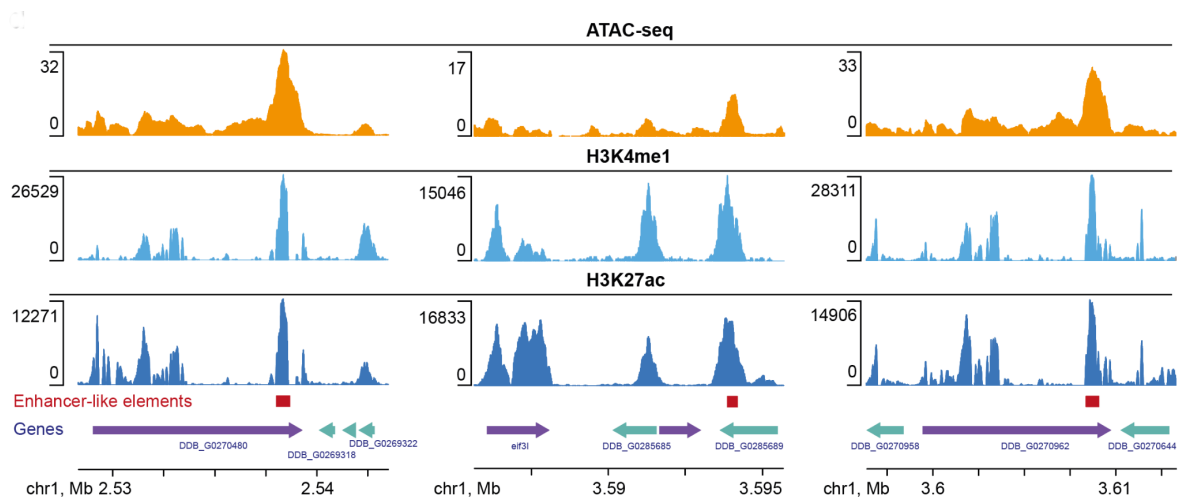


Рисунок 44 — Примеры аннотации энхансеров.

Каждый полученный энхансер был присвоен ближайшему гену с помощью функции `bedtools closest`, и для таких генов был построен график распределения уровней транскрипции; статистическая значимость разницы

экспрессии между генами с энхансером и без него была подсчитана с помощью двустороннего теста Манна-Уитни (Рисунок 45а). Группа генов, не содержащая энхансеров, ожидаемо имеет меньшую экспрессию, чем гены, содержащие энхансер. Более того, подсчитана доля основания петель, содержащих энхансеры, в сравнении с пермутационным тестом — около 30% оснований содержат энхансеры (p -значение $< 0,001$, пермутационный тест) (Рисунок 45б). Интересно также отметить, что гены, содержащие энхансеры, имеют значительно более высокий в среднем уровень экспрессии на всех часах развития, а не только на тех часах развития, для которых они были аннотированы (Рисунок 45в).

Для выяснения процессов, которые регулируются выявленными энхансерами, был проведен анализ обогащения GO-терминов для генов, содержащих энхансеры. Результат анализа показал большое разнообразие обогащенных терминов (Рисунок 46): регуляция транскрипции, метаболические процессы, реакция на стресс и др.

По совокупности полученных результатов можно предположить, что энхансеры в *D. discoideum* представляют собой не специфичные для стадий регуляторные элементы, как это известно для млекопитающих [259] и дрозофилы [260], а скорее элементы, поддерживающие стабильную экспрессию.

3.1.12 Ключевые особенности хроматина *D. discoideum*

Наши данные свидетельствуют о том, что

- Хроматин *D. discoideum* организован в преимущественно последовательные, неиерархические петли, внутренняя часть которых лишь слабо изолирована от соседних локусов (Рисунки 29, 30).
- Основания петель представляют собой сайты конвергентной ориентации генов (Рисунок 34в) и часто колокализуются с границами синтенных блоков (Рисунок 33).
- Сила оснований петель, отражающая эффективность барьерной функции для экстрезии, зависит от силы окружающей его конвергентной транскрипции.

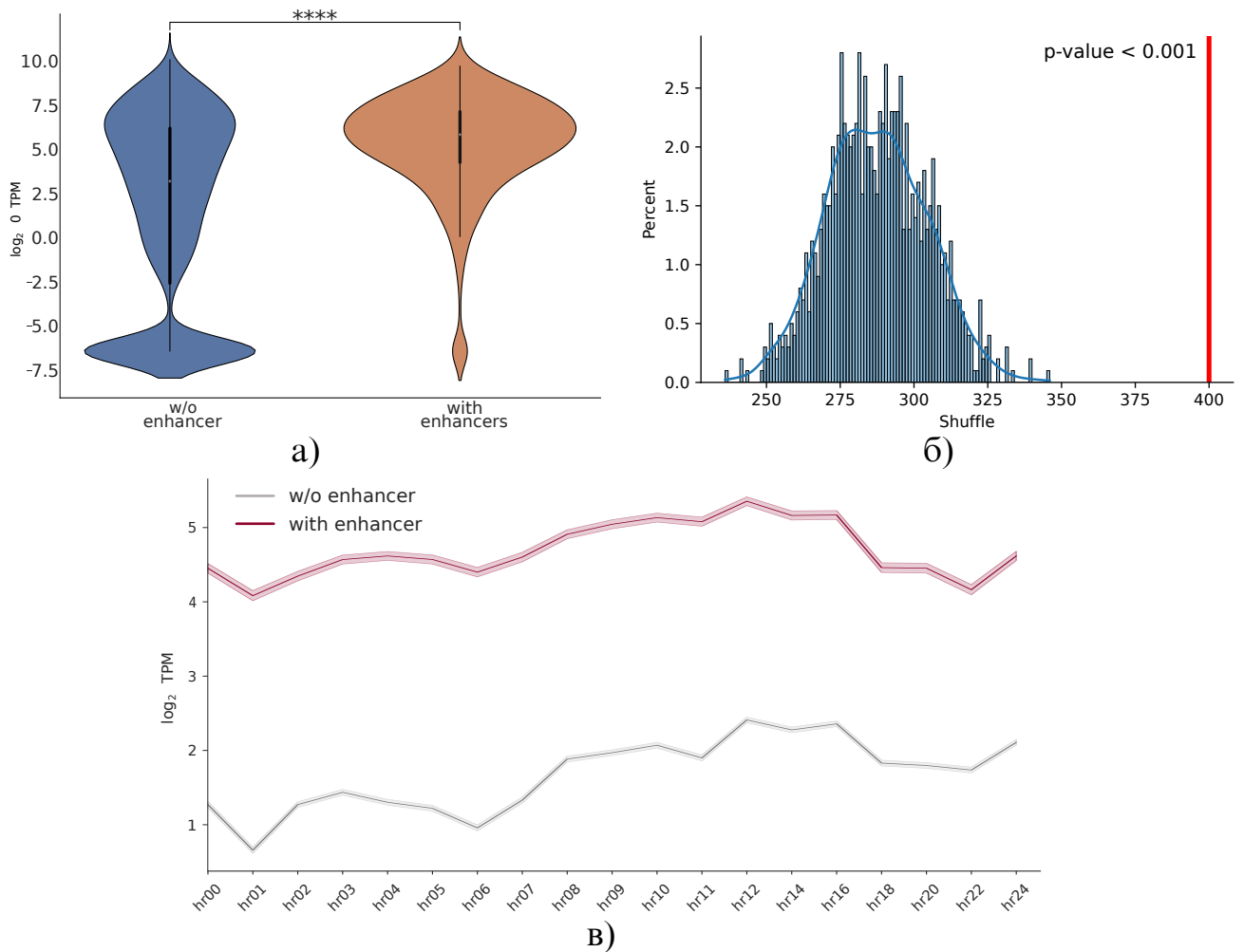


Рисунок 45 — Анализ энхансеров: (а) Уровни транскрипции генов, содержащих энхансер, в сравнении с уровнями транскрипции генов, энхансеров не содержащих. Ось y — TPM-нормализованные значения транскрипции генов, преобразованные в логарифмические значения. (б) Количество энхансеров, пересекающих основания петель, в сравнении со случайными последовательностями. значимость рассчитана пермутационным тестом. (в) Уровни экспрессии генов, содержащих энхансеры, на всех часах развития в сравнении с генами без энхансеров.

- Конвергентные пары генов, расположенные в основания петель, образуют контактные домены, структура которых зависит от уровня транскрипции обоих генов в паре.

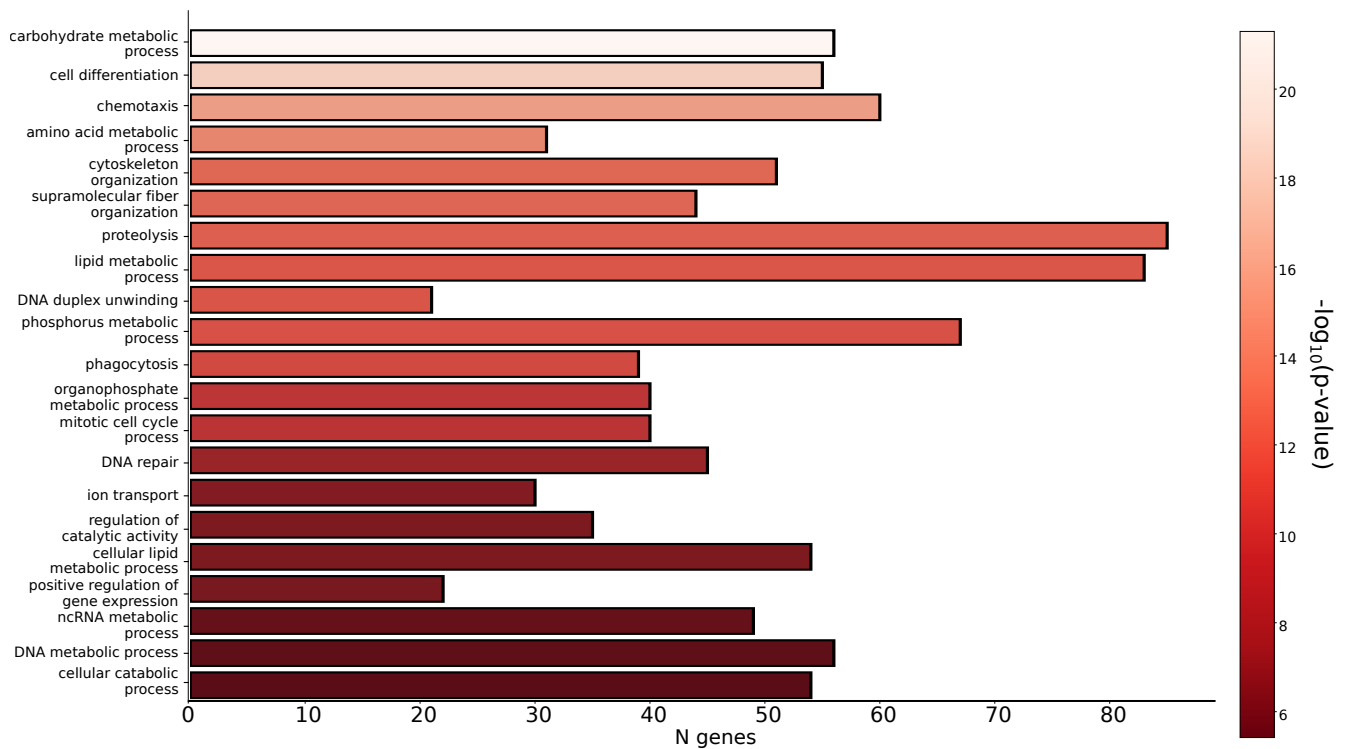


Рисунок 46 — Анализ функций энхансеров *D. discoideum*.

3.2 Организация хроматина *Halisarca dujardini*

3.2.1 Контроль качества карт Hi-C для стадий реагрегации *H. dujardini*

Референсная сборка генома губки *H. dujardini* была реализована А. Черкасовым и В. Зубаревым с использованием Hi-C прочтений, парного полногеномного секвенирования Illumina, а также прочтений, полученных из Nanopore, и переупорядочена в скаффолды приблизительно хромосомного размера с использованием конвейера 3D-DNA с последующим ручным курированием.

Прочтения Hi-C, используемые в данной работе, были получены А. Черкасовым в коллаборации с Институтом биологии развития РАН им. Кольцова и Институтом биологии гена РАН. Карты Hi-C были построены с использованием конвейера `distiller` с опцией `--walks-policy all`, что позволяет получить большее число закартированных прочтений по сравнению со стандартным режимом. Все исследованные образцы содержат значительную долю прочтений, для которых было закартировано только одно прочтение из пары (Рисунок 47), а общая доля закартированных пар со-

ставляет порядка 50 %, что, по-видимому, свидетельствует о несовершенстве текущей версии сборки. Доля дубликатов составляет 5-10 %, что говорит о хорошем качестве секвенирования.

Кроме того, специфичным для карт Hi-C является показатель цис-транс-отношения, рассчитываемый как отношение количества внутрихромосомных и межхромосомных контактов и усредненное для всех хромосом. Результаты для карт Hi-C губок представлены в Таблице 8. Доля внутрихромосомных контактов сильно отличается между образцами и достигает 45 % для образцов Cells-1 и Cells-2, что не является типичным случаем, поэтому требует дополнительного внимания при анализе. Однако следует отметить, что цис-транс-отношение, посчитанное без учета первых двух диагоналей, становится схожим для всех образцов, что, по-видимому, свидетельствует о особенностях процедуры подготовки библиотек Hi-C в первом наборе данных. По этой причине эти образцы были исключены из дальнейшего анализа.

Таблица 8 — Показатели цис-транс-отношения для карт Hi-C губки

Образец	Цис-транс-отношение	Цис-транс-отношение без 2-х диагоналей
Aggregates-1	3,174	0,723
Aggregates-2	3,241	0,581
Cells-1	1,466	0,620
Cells-2	1,145	0,569
Cells-3	3,200	0,568
Tissue-1	12,514	0,601
Tissue-2	19,495	0,554

3.2.2 Анализ межхромосомных и внутрихромосомных взаимодействий *H. dujardini*

Для анализа взаимодействий между центромерами и теломерами хромосом *H. dujardini* была использована та же методология, что и для хрома-

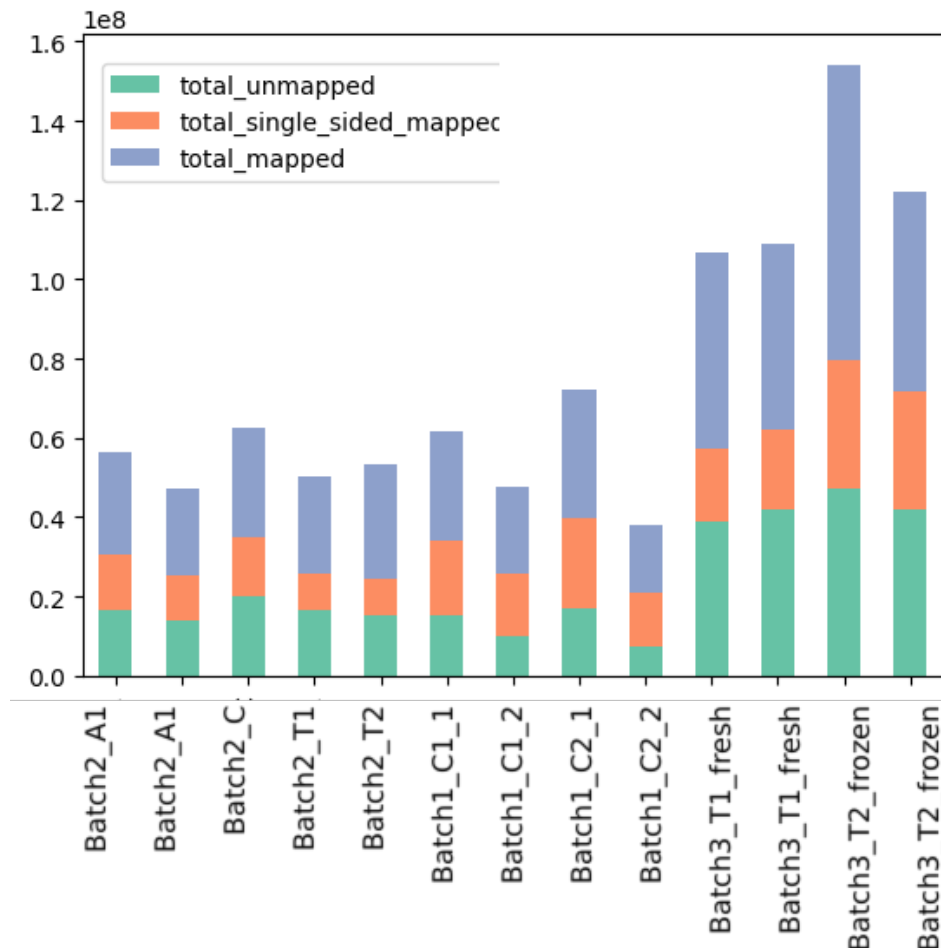


Рисунок 47 — Число прочтений для карт Hi-C губок. Изумрудный — некартированные прочтения, оранжевый — пары прочтений, в которых только одно прочтение из пары было закартировано на геном, сине-фиолетовое — картированные пары прочтений.

тина *D. discoideum* (см. раздел 2.5). Однако, в отличие от акроцентрических хромосом *D. discoideum*, у губок центромера разделяет два примерно одинаковых плеча хромосомы, поэтому для губки возможно разделить теломерные и центромерные взаимодействия (Рисунок 48). Полученный график свидетельствует о наличии центромер-центромерных и теломер-теломерных взаимодействий между хромосомами (центромер-центромерные взаимодействия представлены центральной точкой на Рисунке 48, теломер-теломерные — взаимодействия в углах графика). По результатам анализа, значения частот центромер-центромерных и теломер-теломерных примерно в полтора раза выше, чем средняя частота межхромосомных взаимодействий. Однако важно ещё отметить, что для хромосом *H. dujardini* также наблюдаются центромер-теломерные взаимодействия между разными хромосомами (по-

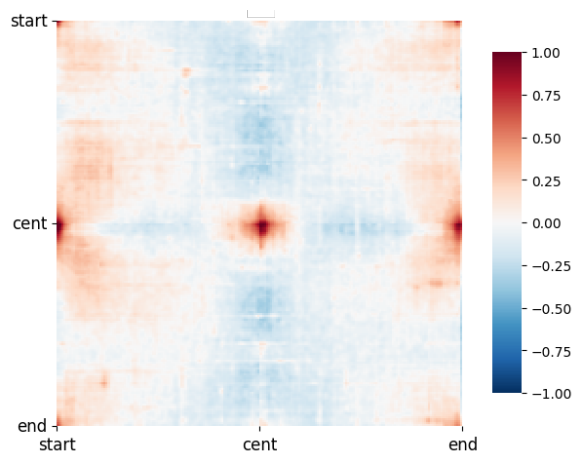


Рисунок 48 — Карта среднего хромосомного взаимодействия для карт Hi-C губок.

вышенные частоты контактов, расположенные в середине около границ графика, Рисунок 48).

Описанные характеристики свидетельствуют о том, что хромосомы *H. dujardini* образуют структуру Рабля с повышенными частотами контактов между теломерами и между центромерами разных хромосом, но не центромер-теломерными взаимодействиями.

3.2.3 Анализ компартментов у *H. dujardini*

Для изучения компартментной организации хроматина *H. dujardini* были посчитаны собственные вектора карты Hi-C для каждого бина генома на разрешении 25 тысяч п.н. Для фазирования собственного вектора был использован динуклеотидный состав последовательности для определения того, соответствуют положительные или отрицательные проекции на первый собственный вектор активному или неактивному компартменту [191]. Фазирование основано на различии динуклеотидного состава регионов, принадлежащих активному и неактивному хроматину. На Рисунке 49а изображен пример смены знака для первого собственного вектора. Визуальный анализ показывает, что в хроматине *H. dujardini* присутствуют выраженная «шахматная доска», а первый собственный вектор хорошо отражает изменения в паттернах взаимодействия хромосом. Была также построена седловая диаграмма для изучения силы предполагаемых компартментов (Рисунок

496). Она свидетельствует о наличии выраженных взаимодействий как между локусами активного компартмента, так и внутри неактивного компартмента. При этом частоты взаимодействий локусов, принадлежащих разным компартментам, оказываются ниже ожидаемых. Проведенный анализ свидетельствует о выраженной компартментализации хроматина *H. dujardini*.

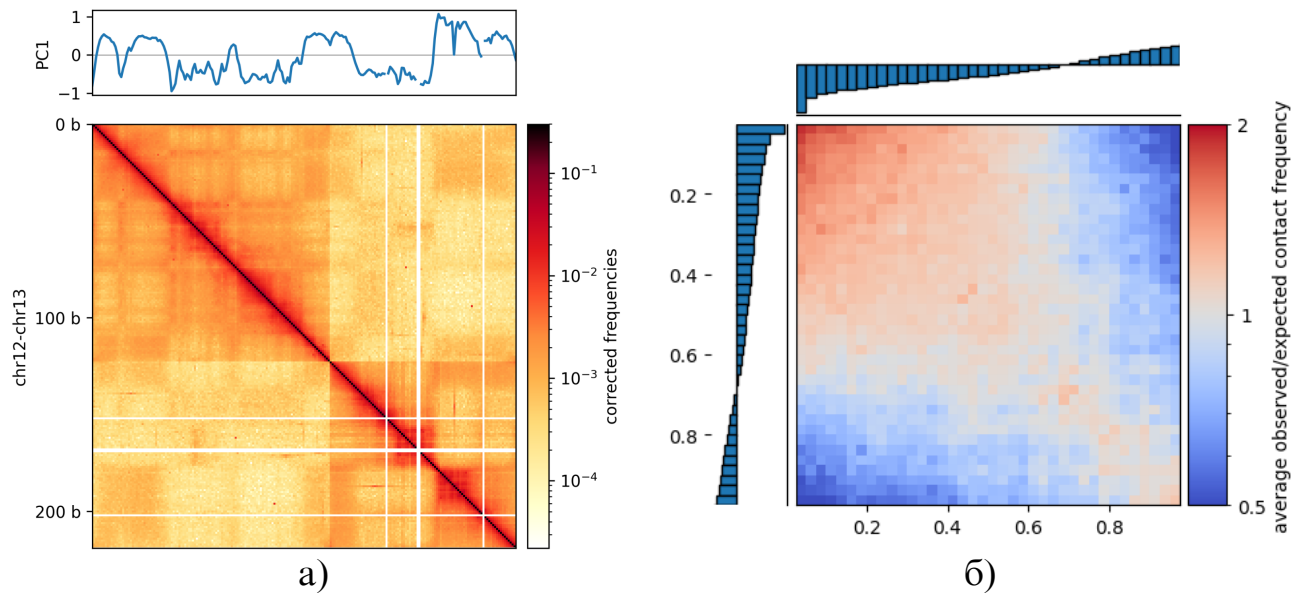


Рисунок 49 — Компартменты в хроматине *H. dujardini*. (а) Пример смены знака проекции на первый собственный вектор для нескольких хромосом *H. dujardini*. (б) Седловая диаграмма для хроматина *H. dujardini*.

3.2.4 Анализ хроматиновых шпилек/фонтанов *H. dujardini*

При визуальной оценке карты контактов губки *H. dujardini* с использованием хроматинового браузера HiGlass [230] на разрешении 3-5 тысяч п.н. мы обнаружили яркие вытянутые структуры, похожие на хроматиновые шпильки, или фонтаны. Для аннотации данных объектов был использован инструмент fontanka [159], специально разработанный для поиска подобных структур. В ходе поиска фонтанов у губки *H. dujardini* были подобраны параметры фильтрации (см. раздел 2.14) на основании предлагаемых в работе [159], позволяющее оставить фонтаны, лучше соответствующие визуальной оценке (Рисунок 506). Такой подход позволил нам избавиться от шума (левая панель Рисунка 51а) и получить 986 фонтанов со средней силой 0,17 (правая

панель Рисунка 51a). Для губки *Petrosia ficiformis* данные Hi-C были взяты из проекта Tree of Life (SRA: ERR10297828). Аналогичная процедура, проведенная с картой контактов губки *P. ficiformis*, позволила получить 1309 фонтанов со средней силой 0,13 (Рисунок 51б). Проведенный анализ позволяет говорить о большой распространенности фонтанов в хроматине представителей Porifera.

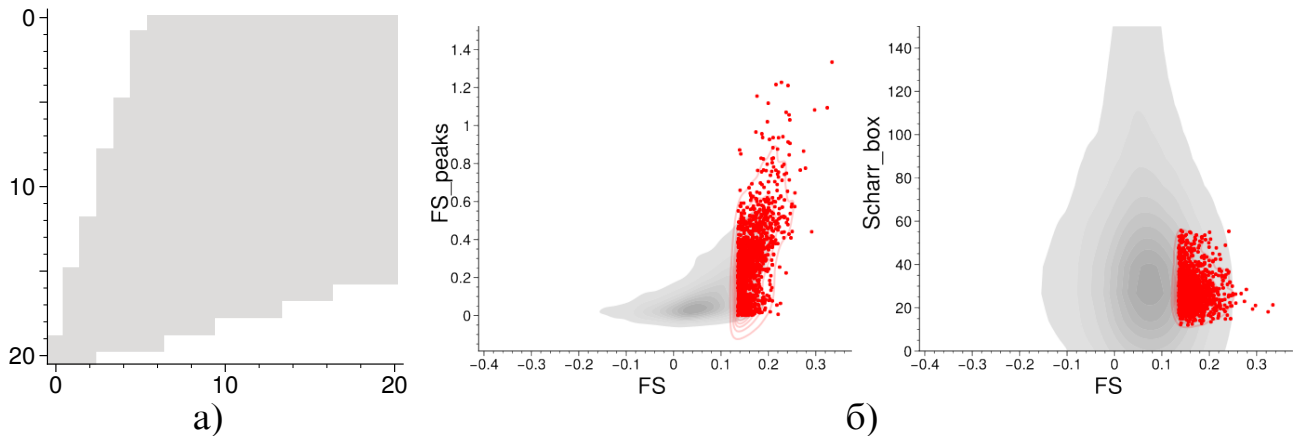


Рисунок 50 — Поиск фонтанов в хроматине губки. (а) Маска, использованная при поиске фонтанов. (б) Фильтрация найденных фонтанов по индексу Шарра и силе фонтана. Детали методов описаны в разделе 2.14).

3.2.5 Сравнительный анализ ферритинов у представителей типа Porifera

Для морских губок наличие ионов железа является обязательным условием для роста [101], при этом важно отметить, что у губок есть богатое разнообразие симбионтов, которые обладают способностью усваивать железо, полученное ими от губок [102; 103]. Метаболизм железа требует наличия системы белков-переносчиков, систем детоксикации и запасания железа в клетках, закодированных в геноме, поэтому отдельно интересно изучить, какие именно гены участвуют в данном процессе и какую роль в этом играет организация хроматина.

В геноме губки *H. dujardini* обнаружено три последовательности со статистически значимым сходством с аминокислотной последовательностью ферритина человека (HuHF, см. раздел 2.22), причем две последовательности, обладающие наибольшим сходством (далее *HdF1a* и *HdF1b*), располо-

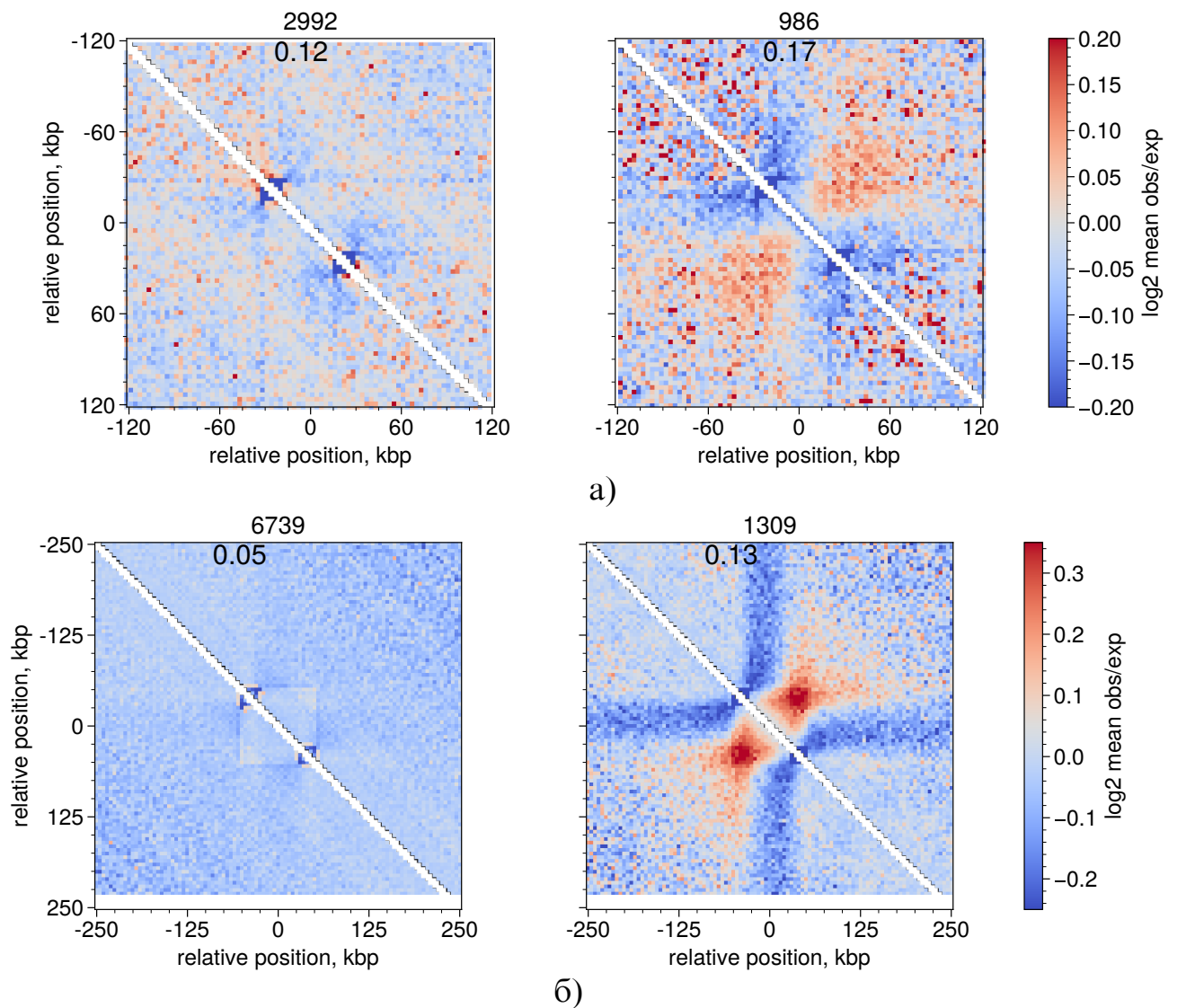


Рисунок 51 — Средняя карта контактов вокруг хроматиновых фонтанов. Левая панель: все найденные фонтаны, правая панель: применение дополнительной фильтрации. (а) Средняя карта контактов вокруг хроматиновых фонтанов *N. dujardini*. (б) Средняя карта контактов вокруг хроматиновых фонтанов *P. ficiformis*.

жены на расстоянии чуть больше 3 тысяч п.н. друг от друга на скаффолде 11, в то время как *HdF2* располагается на скаффолде 3. Дальнейший анализ показывает, что кодирующую часть генов *HdF1a* и *HdF1b* между собой отличают только две нуклеотидные замены, одна из которых синонимичная. В то же время, кодирующая часть гена *HdF2* значительно отличается от генов *HdF1a/b*: уровень аминокислотного сходства между ними порядка 70 %.

Далее мы попарно подсчитали процент сходства аминокислотных последовательностей гомологов ферритинов из различных представителей типа Porifera и NuHF, а также количество аминокислотных замен в консер-

вативных позициях активных центров. Результаты анализа, представленные в Таблице 9, показывают, что у всех исследованных представителей типа Porifera за исключением *H. dujardini* обнаружено по два гена ферритинов. Интересно также отметить, что для большинства последовательностей число аминокислотных замен в сравнении с HuHF не превышает трех, за исключением упомянутого выше белка *H. dujardini* HdF2 (шесть аминокислотных замен) и ферритина губки *Ephydatia muelleri* EmF2 (шесть аминокислотных замен). По совокупности признаков было решено называть ферритины HdF2 и EmF2, имеющие более пяти аминокислотных замен в функциональных доменах по сравнению в HuHF, атипичными ферритинами. Из литературы [245; 246] известно, что мутации в консервативных позициях функциональных доменов значительно снижают эффективность накопления, транспортировки и детоксикации железа. Кроме того, для атипичных ферритинов HdF2 и EmF2 не удалось предсказать наличие железосвязывающей способности с помощью веб-сервера SIRE [247]. В совокупности это указывает на принципиальные отличия двух групп ферритинов, однако какие функции атипичные ферритины выполняют в клетке остается неизвестными.

Таблица 9 — Процент аминокислотного сходства с HuHF и количество аминокислотных замен в функциональных доменах у ферритинов представителей Porifera, класс Demospongiae.

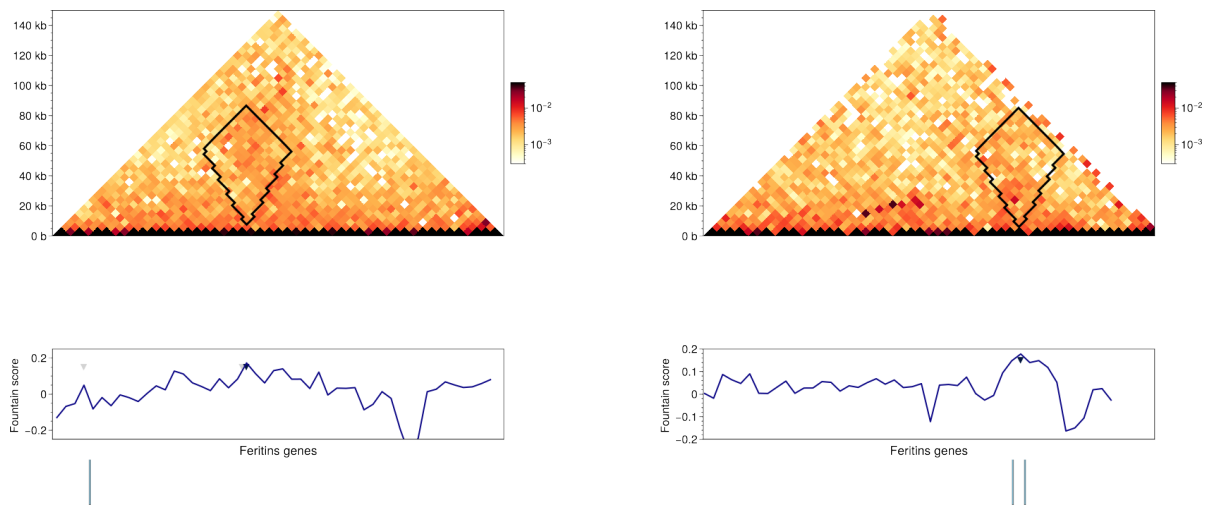
Организм	Белок	Процент сходства	Кол-во замен
<i>Halisarca dujardini</i>	HdF1a	69.9	2
	HdF1b	69.4	2
	HdF2	55.2	6
<i>Petrosia ficiformis</i>	PfF1a	71.8	4
	PfF1b	69.9	4
<i>Amphimedon queenslandica</i>	AqF1a	68.6	1
	AqF1b	67.5	1
<i>Lubomirskia baikalensis</i>	LbF1a	71.0	1
	LbF1b	71.6	1
<i>Ephydatia muelleri</i>	EmF1	70.5	1

Организм	Название гена	Процент сходства	Кол-во замен
	EmF2	61.2	6

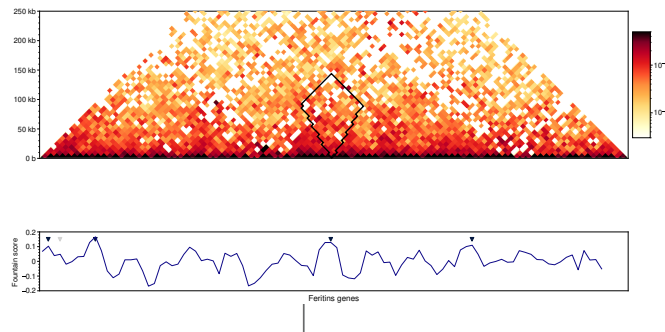
3.2.6 Сравнительный анализ трехмерного окружения ферритинов у губок и других представителей Metazoa

Как было сказано выше, в геноме губки *H. dujardini* найдено три гена ферритинов. Используя ранее сделанную (см. раздел 3.2.4) аннотацию фонтанов, мы проверили наличие этих структур в окрестностях локусов генов ферритинов. Для генов консервативных ферритинов основание фонтана расположено между гена HdF1a и HdF1b (Рисунок 52a), а для гена атипичного ферритина основание ближайшего фонтана находится на расстоянии десятка тысяч пар оснований. Для губки *P. ficiformis* также была использована аннотация хроматиновых фонтанов и аннотация генов ферритинов. Стоит отметить, что предсказанные гены ферритинов *PfF1a* и *PfF1b* расположены на расстоянии 700 нуклеотидов друг от друга, а потому, вероятно, являются частями одного гена из-за ошибки программы аннотации. Результат анализа хроматинового окружения генов ферритина (Рисунок 52б) показывает, что основание одного из хроматиновых фонтанов расположено в непосредственной близости от локуса ферритина. Полученные результаты свидетельствуют о колокализации генов консервативных, но не атипичных ферритинов, и оснований хроматиновых фонтанов. Данные литературы [158; 159] показывают, что основания фонтанов преимущественно образуются на активных энхансерах, а сам фонтан является некоторой промежуточной стадией между загрузкой когезина и формированием энхансер-промоторной петли. На основании этих данных и наших выводов можно предполагать, что и в хроматине двух представителей типа Porifera фонтаны помогают поддерживать уровень транскрипции генов консервативных ферритинов путем реализации энхансер-промоторных взаимодействий.

Для изучения того, как меняется укладка хроматина вокруг ферритинового локуса у организмов с эмбриональным развитием, были взяты



а)



б)

Рисунок 52 — Карты хроматиновых контактов ферритинов у представителей типа Porifera. (а) Карта хроматиновых контактов генов ферритинов у губки *H. dujardini*. Левая панель: ген атипичного ферритина. Правая панель: гены консервативных ферритинов. (б) Карта хроматиновых контактов генов ферритинов у губки *P. ficiformis*. Хроматиновый фонтан выделен по форме маски, использованной для его поиска, а также отмечен как пик значений fountain score. Размер региона: 500 тыс. п.н.

данные для четырех стадий раннего эмбрионального развития рыбы *D. rerio* из работы [160]. Согласно данным работы [159], фонтаны у этого организма появляются на стадии 4 часа после оплодотворения и превращаются в топологически ассоциированные домены (ТАДы) — плотно упакованную в пространстве глобулу ДНК — к 20 часам после оплодотворения. В отличие от представителей типа Porifera с весьма компактным геномом, в хроматине *D. rerio* образуются фонтаны между регионами, расположенными на линейном расстоянии друг от друга нескольких сотен тысяч пар нуклеотидов (Рису-

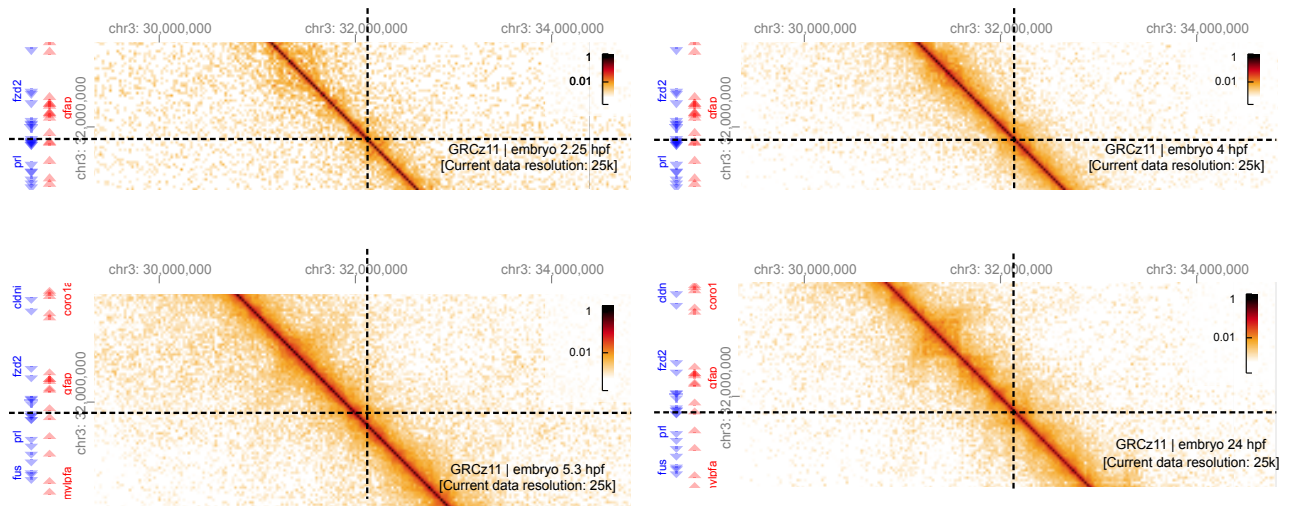
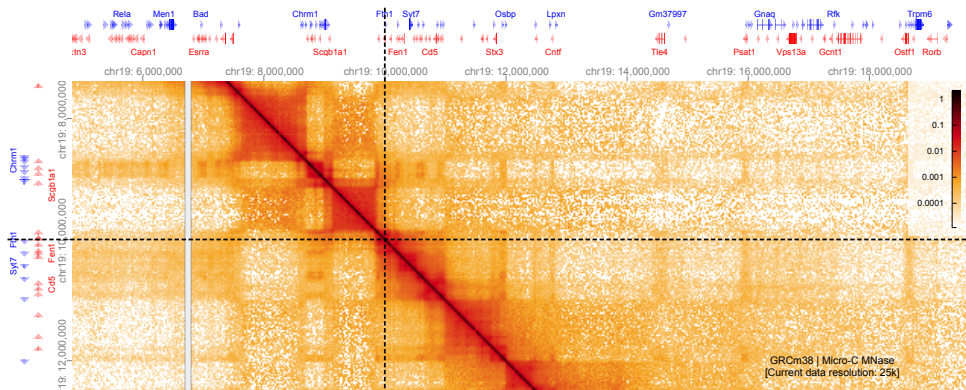


Рисунок 53 — Карта хроматиновых контактов ферритинов у *D. rerio*. Расположение гена *fth1*, указано пунктирными линиями. Размер бина карты — 25 тыс п.н.

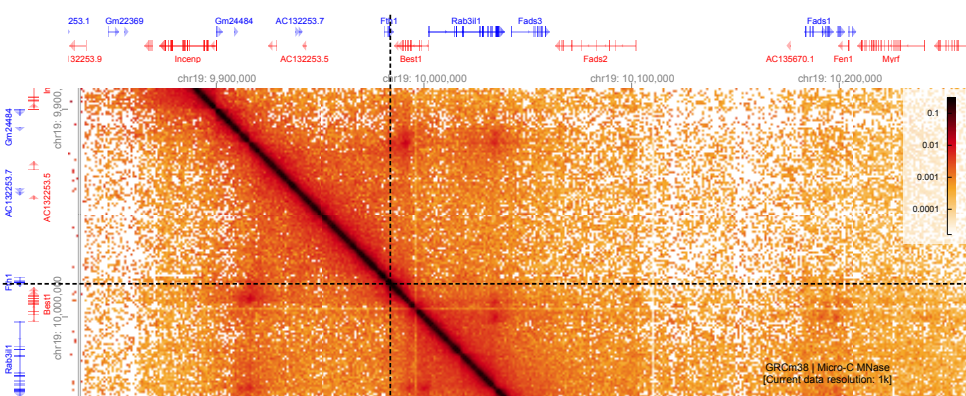
нок 53). В то же время визуальный анализ демонстрирует, что сам локус гена ферритина *fth1* не является основанием фонтана, а к 20-му часу после оплодотворения локус входит в состав ТАДа. Полученные результаты свидетельствуют о том, что хроматиновое окружение локуса ген ферритина *fth1* у рыбы *D. rerio* меняется в ходе эмбрионального развития: ген *fth1* удален от оснований фонтанов, возникающих в раннем эмбриональном развитии, но входит в состав ТАДа на стадии 20 часов после оплодотворения. Наблюдаемые изменения, вероятно, являются отражением глобальной реорганизации хроматина в ходе эмбрионального развития, например, вследствие зиготической активации генома [261—263].

Данные Hi-C из работы [264] для раннего эмбрионального развития мыши *Mus musculus* на ранней и поздней стадии двух клеток, а также стадии восьми клеток, не позволяют говорить о существенных изменениях в локусе ферритина, что, по-видимому, является отражением недостаточной глубины секвенирования для проведения подобного анализа (результаты не показаны). Однако применение метода Micro-C позволяет получить более детальное представление о трехмерной организации локуса ферритина для взрослой мыши. Анализ карты контактов из работы [265] на разрешении 25 тысяч п.н. позволяет сделать вывод о расположении гена внутри довольно крупного ТАДа (Рисунок 54а). Анализ карты контактов на разрешении 1

тысяча п.н. позволяет детектировать множество иерархических петель по геному, в том числе и вблизи локуса гена ферритина (Рисунок 54б).



а)



б)

Рисунок 54 — Карта хроматиновых контактов ферритинов у мыши *M. musculus*, полученная методом Micro-C (а) на разрешении 25 тыс. п.н.; (б) на разрешении 1 тыс. п.н. Расположение гена *fth1* указано пунктирными линиями.

Согласно литературным данным [187], активная транскрипция является барьером для экструзии, а потому полное её выключение ведёт к потере большинства петель и ТАДов. Однако открытым вопросом остается изменение структуры хроматина при изменении уровня экспрессии в несколько раз. Для изучения этого на примере хроматинового окружения локуса *fth1* в тканях человека данные Hi-C были взяты из работы [266]. Согласно данным Human Protein Atlas [267], в ткани легкого человека уровень экспрессии гена *fth1* в 1,5-2 раза выше, чем в ткани яичника. Визуальный анализ показывает (Рисунок 55), что в обеих тканях ген ферритина расположен внутри рыхлого ТАДа, структура которого слабо отличается для двух рассмотренных образцов несмотря на разницу в уровне транскрипции этого гена между тканями.

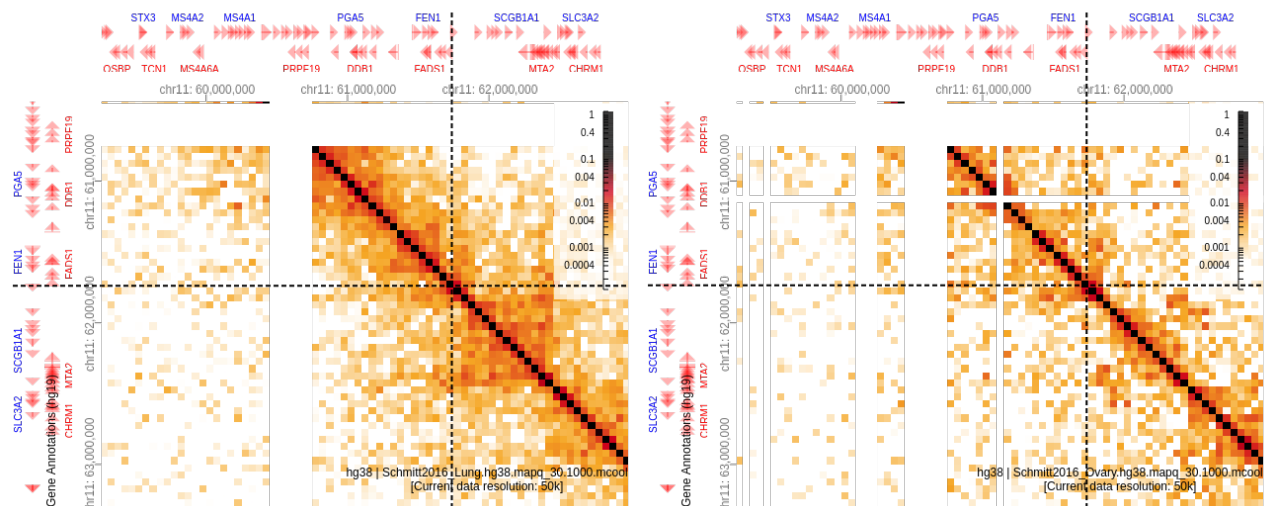


Рисунок 55 — Карта хроматиновых контактов ферритинов для ткани легкого и яичника человека *H. sapiens*. Расположение гена *fth1* указано пунктирными линиями.

Таким образом, мы провели сравнительный анализ хроматинового окружения локуса ферритина у двух представителей типа Porifera, *P. ficiformis* и *H. dujardini*, рыбы *D. rerio*, мыши *M. musculus* и человека *H. sapiens* и показали, что в хроматине губок, но не *D. rerio*, основание хроматинового фонтана находится в непосредственной близости от генов консервативных ферритинов, а в хроматине *D. rerio* на стадии 20 часов после оплодотворения, а также у млекопитающих *M. musculus* и *H. sapiens* ген ферритина расположен внутри ТАДа, чья структура только незначительно меняется при изменении уровня экспрессии.

3.3 Организация хроматина при анеуплоидиях

3.3.1 Изучение межхромосомных взаимодействий на примере анеуплоидий

Для изучения возможной взаимосвязи между экспрессией генов и организацией хроматина была выбрана модель анеуплоидий, в частности, трисомии по хромосомам 13, 16 и 18. Коллегами из Института биологии гена РАН были выполнены эксперименты Hi-C на клетках хориона, несущих до-

полнительную хромосому 13 (синдром Патау, Tr13) или хромосому 16 (Tr16, наиболее частая трисомия при анализе материала самопроизвольных абортов), а также на культивируемых первичных фибробластах кожи с третьей копией хромосомы 18 (синдром Эдвардса, Tr18).

Анализ парных взаимодействий между хромосомами в нормальных клетках показал, что хромосомы небольшого размера образуют контактный кластер (т.е. имеют повышенные частоты контактов взаимодействия друг с другом) и имеют низкую частоту взаимодействия с крупными хромосомами, которые образуют менее выраженный, но заметный кластер как в клетках хориона, так и в фибробластах (Рисунок 56). Поэтому мы определили хромосомы 16-22 как «малые», а хромосомы 1-15, X — как «большие». Эти наблюдения согласуются с предыдущими сообщениями о большей частоте взаимодействий между малыми хромосомами по сравнению со средней частотой взаимодействия между хромосомами [191]. Это можно объяснить различной локализацией малых и больших хромосом внутри ядра: в то время как крупные хромосомы имеют тенденцию располагаться на периферии ядра и контактировать с ламиной, малые хромосомы преимущественно локализуются внутри ядра [268]. Следует отметить, что в клетках хориона обеденная генами хромосома 18 взаимодействует с относительно высокой частотой только с хромосомами 20 и 21, которые содержат меньше генов по сравнению с другими малыми хромосомами (Рисунок 57). В фибробластах с Tr18 хромосома 18 имеет аномально сильный контакт с хромосомой 9 (Рисунок 56, нижняя левая панель). Как выяснилось при визуальном анализе карт Hi-C, это, вероятно, вызвано локальной перестановкой между 18q и околоцентроммерной областью 9q. Во избежание систематических ошибок хромосома 9 была исключена из всех последующих сравнений нормальных фибробластов и фибробластов с Tr18.

Паттерн межхромосомных контактов в целом схож для образцов с различными дополнительными хромосомами: Tr13, Tr16 и Tr18 (сравните левую и среднюю панели на Рисунке 56). Однако мы обнаружили, что наличие дополнительных хромосом вследствие трисомии вызывает умеренные, но статистически значимые и противоположные изменения в контактах между малыми и большими хромосомами (Рисунок 58). Для Tr13 и Tr18 мы наблюдаем значительное различие между образцами с трисомией и контролем в кластерах больших и малых хромосом, сопровождающееся увеличением ча-

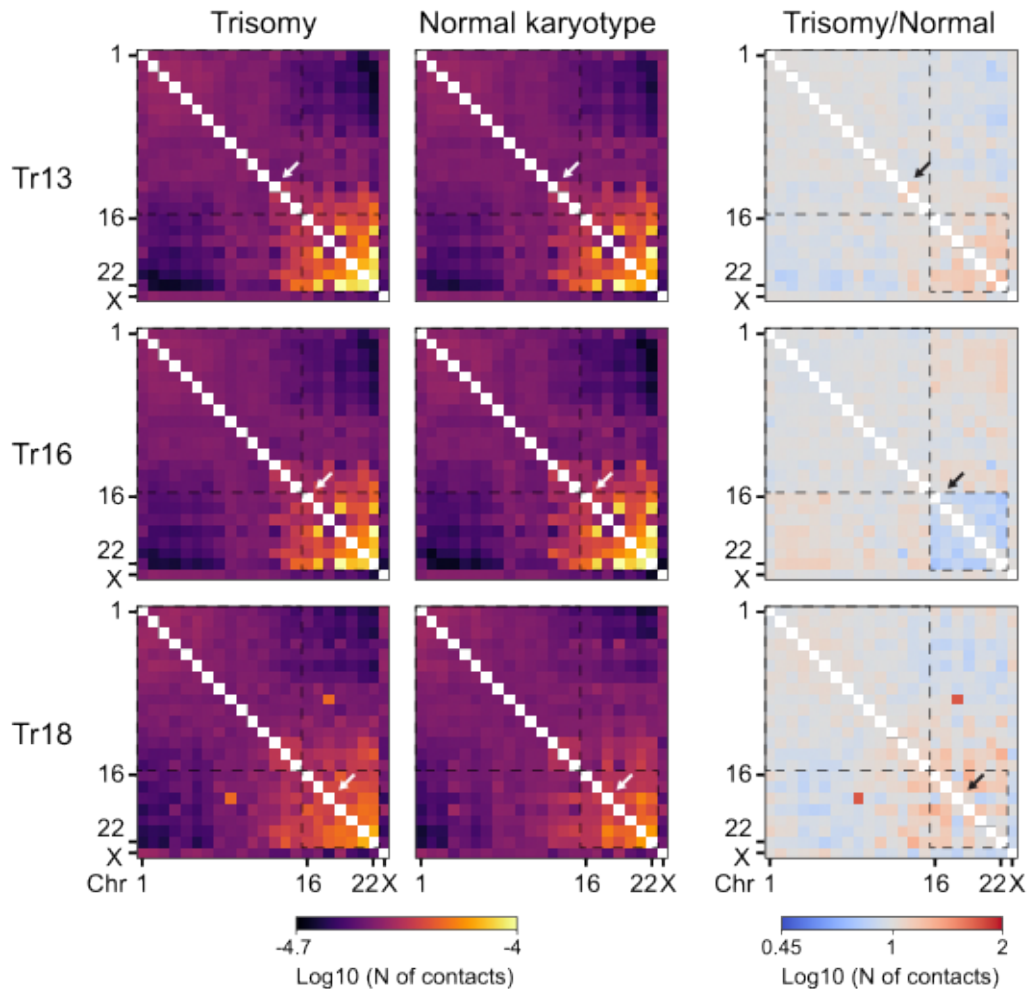


Рисунок 56 — Карта межхромосомных контактов для образцов с трисомией по 13, 16, 18 хромосомам человека *H. sapiens*. Стрелка указывает на трисомийную хромосому.

стоты контактов между малыми хромосомами, тогда как для образцов Tr16 характерно частичное перемешивание кластеров больших и малых хромосом, а также выраженное снижение взаимодействий малых хромосом. Важно отметить, что различия в характере контактов хромосом при попарном сравнении нормальных клеток от разных доноров значительно менее выражены (особенно для малых хромосом).

Таким образом, наличие лишней хромосомы в ядре может систематически влиять на частоту контактов между другими хромосомами и превышает степень изменчивости между отдельными здоровыми донорами.

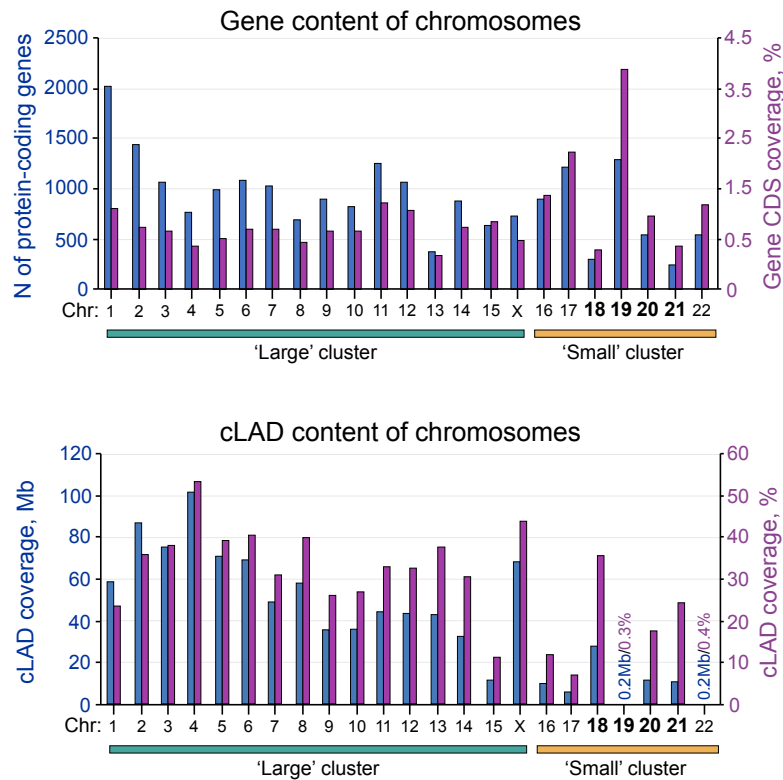


Рисунок 57 — Абсолютные значения и доля длины хромосомы, покрытой белок-кодирующей частью генов и ламина-ассоциированными доменами в геноме человека *H. sapiens*.

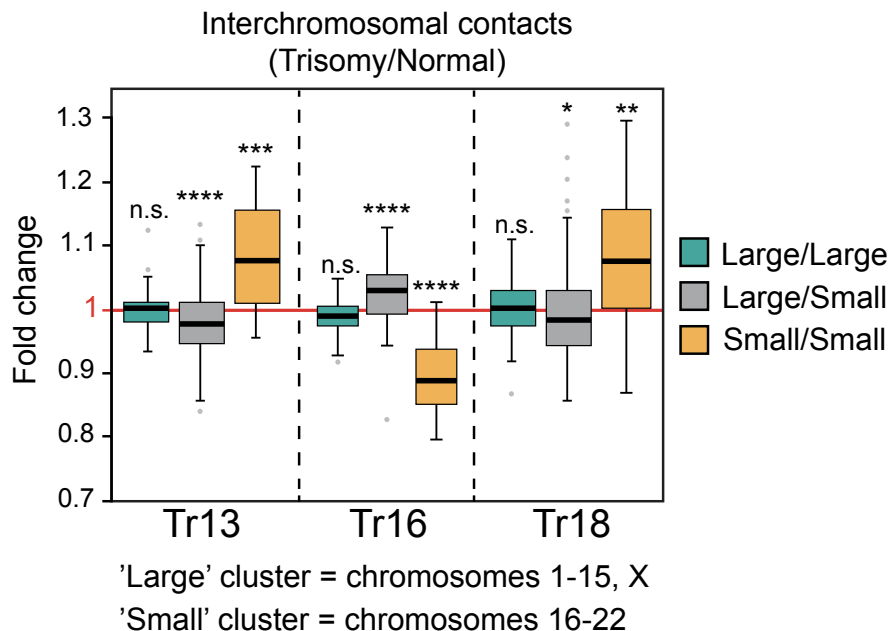


Рисунок 58 — Различия частот межхромосомных контактов для больших и малых хромосом с трисомиями по хромосомам 13, 16, 18. ns p -значение $> 0,05$; * p -значение $< 0,05$; ** p -значение $< 0,01$; *** p -значение $< 0,001$; **** p -значение $< 0,0001$ по тесту Манна-Уитни.

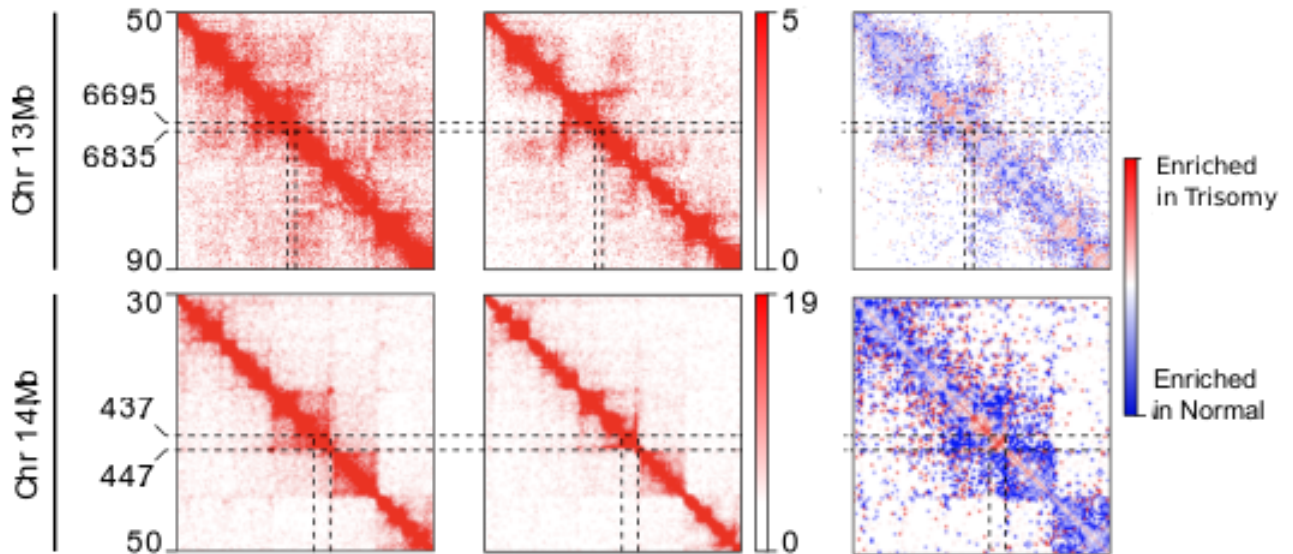


Рисунок 59 — Примеры карт хроматиновых контактов для двух регионов, которые меняют свою плотность в фибробластах с Tr18 сравнении с контрольным образцом. Верхняя панель: хромосома 13, 40 млн. п.н. Нижняя панель: хромосома 14, 20 млн п.н.

3.3.2 Изменения локальной упаковки хроматина образцов с трисомиями на уровне всего генома

Визуальный сравнительный анализ карт хроматиновых контактов образцов с трисомиями показал значительные изменения локальной упаковки хроматина в образцах с трисомиями (Рисунок 59).

Для численной оценки изменения локальной упаковки в образцах с трисомией, было введено и подсчитано R — отношение долей частот контактов на расстоянии не более 1 млн. п.н. к сумме всех внутрихромосомных контактов данного локуса для образцов с трисомиями и контроля. Компактизованными регионами являются те, для которых значение $R > 1$; декомпактизованными — при значениях $R < 1$. Для всех хромосом мы подсчитали долю суммарной длины регионов, изменивших упаковку, к общей длине хромосом, отдельно для больших и малых хромосом. Результаты численной оценки показывают, что для образцов Tr18 и Tr13, но не образца Tr16, доли длины хромосомы с измененной упаковкой значительно больше, чем для малых (Рисунок 60a). В образцах Tr13 и Tr18 мы обна-

ружили статистически значимую корреляцию между покрытием хромосом ламина-ассоциированными доменами и долей компактизованных участков на больших, но не на малых хромосомах (Рисунок 60б), что также свидетельствует о существенных различиях между этими двумя группами.

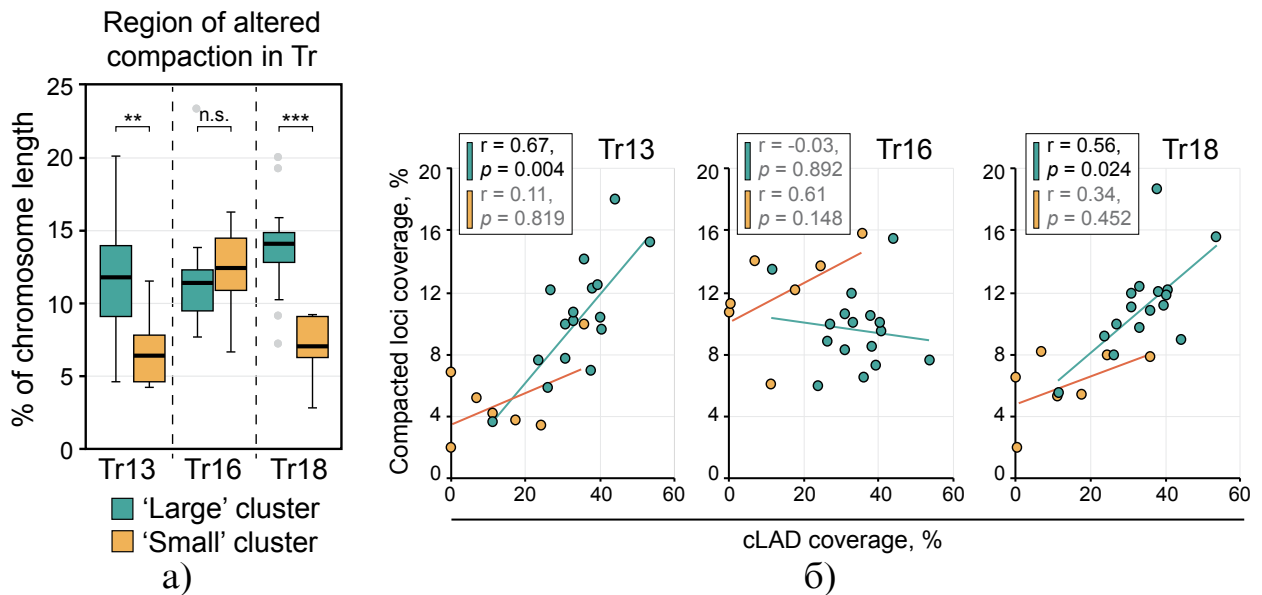


Рисунок 60 — Сравнительная характеристика регионов хроматина с измененной упаковкой. (а) Сравнение доли длины хромосом с измененной упаковкой для образцов с анеуплоидиями для больших и малых хромосом. ns p -значение $> 0,05$; * p -значение $< 0,05$; ** p -значение $< 0,01$; *** p -значение $< 0,001$ по тесту Манна-Уитни. (б) Взаимоотношения доли длины хромосом, для которых изменилась компактизация хроматина, и доли покрытия ламина-ассоциированными доменами. r — коэффициент корреляции Спирмена.

Более того, если оценить принадлежность компактизованных ($R > 1$) и декомпактизованных ($R < 1$) регионов к эу- или гетерохроматиновым регионам, оказывается, что компактизованные регионы с примерно равной вероятностью принадлежат как А-, так и В-компартаментами (Рисунок 61). Однако для декомпактизованных регионов отмечена более частая принадлежность к В-компарменту, которая достигает 90 % для Tr18, что свидетельствует о преимущественной декомпактизации именно регионов гетерохроматина.

Далее мы изучили представленность белок-кодирующих генов и генов домашнего хозяйства в компактизованных и декомпактизованных регионах (Рисунок 62). Результаты анализа указывают на недопредставленность

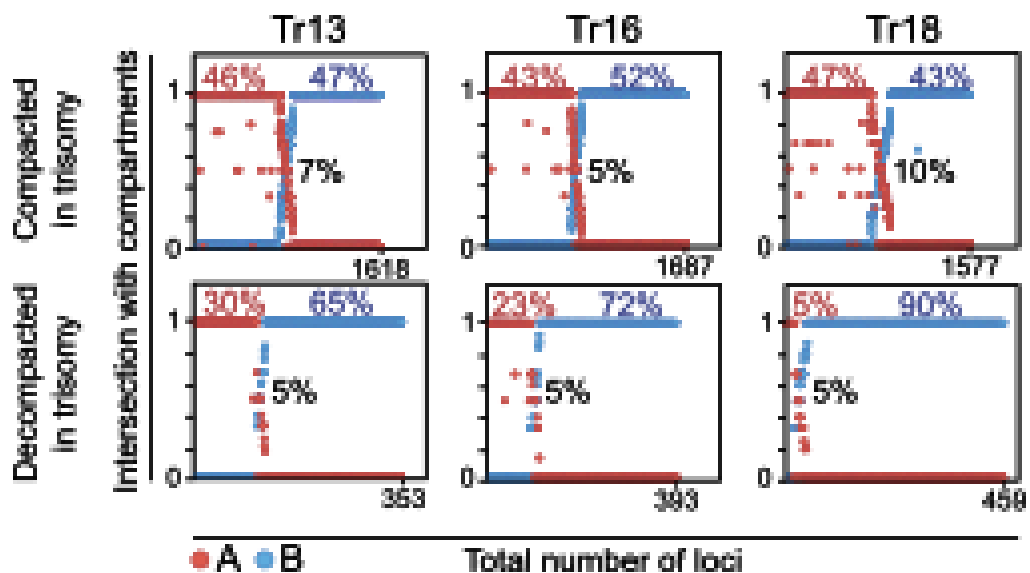


Рисунок 61 — Принадлежность компактизованных и декомпактизованных регионов к эу- и гетерохроматину.

белок-кодирующих генов в декомпактизованных регионах, особенно выраженную для крупных хромосом (Рисунок 62а). Более того, если взять только гены домашнего хозяйства, то есть те, которые стабильно и сильно экспрессируют в различных тканях, оказывается, что они значительно недопредставлены в декомпактизованных регионах как малых, так и больших хромосом и наоборот, немного перепредставлены в регионах с увеличенной в трисомии хромосомы 18 плотности (Рисунок 62б). Такие локальные изменения плотности упаковки хроматина по всему геному, обусловленные появлением дополнительной копии хромосомы, могут обеспечивать изменения уровней экспрессии различных генов, что может влиять на фенотипические изменения.

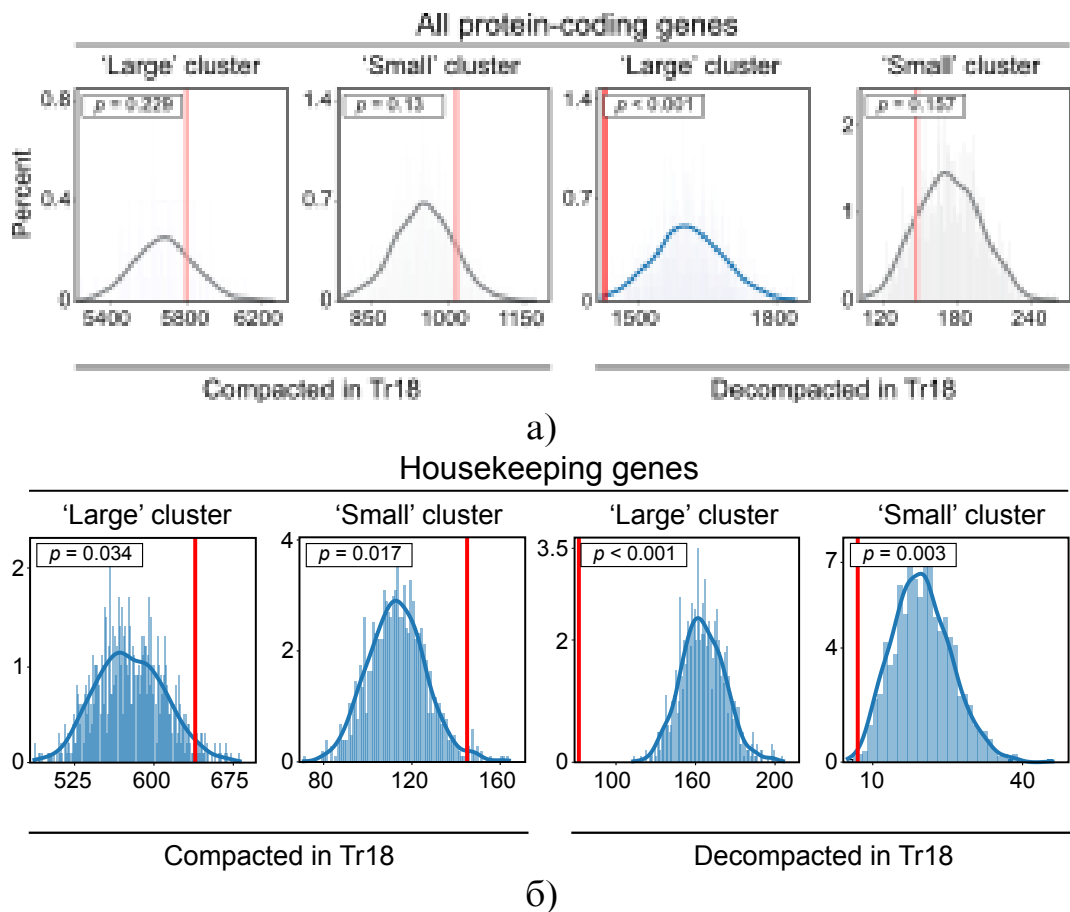


Рисунок 62 — Характеристика регионов хроматина с измененной упаковкой. (а) Сравнение доли длины хромосом, для которых изменилась компактизация хроматина, у образцов с анеуплоидиями в сравнении с контролем. (б) Взаимоотношения доли длины хромосом, для которых изменилась компактизация хроматина, и доли покрытия ламина-ассоциированными доменами.

Глава 4. Обсуждение результатов

Феномен взаимодействия центромер хорошо известен в литературе; структура Рабля представляет собой центромер-центромерные и теломер-теломерные взаимодействия различных хромосом. Подобные структуры в литературе описаны, например, в клетках кончика корня лука *Allium cepa* [269], делящихся дрожжах *Schizosaccharomyces pombe* [270], дрозофиле *Drosophila melanogaster* [271] и почкующихся дрожжах *Saccharomyces cerevisiae* [272]. В основе подобных взаимодействий, по-видимому, лежит дефектный или отсутствующий конденсин II [273]. Однако недавняя работа [274] представляет интересный феномен перехода от структуры Рабля к хромосомным территориям по мере дифференциации нейронов, потенциально указывая на роль других факторов в укладке хромосом. В настоящей работе было изучено, образуют ли акроцентрические хромосомы *Dictyostelium discoideum* и хромосомы *Halisarca dujardini* такие же структуры Рабля. Было обнаружено, что хромосомы *D. discoideum* не склонны к образованию явных хромосомных территорий, а средние карты межхромосомных контактов как для *D. discoideum*, так и для *H. dujardini* показывают обогащение контактов между центромерами и теломерами разными хромосомами. Это свидетельствует о главенствующей структуре Рабля, но не хромосомных территорий, в укладке интерфазных хромосом *D. discoideum* и *H. dujardini*. Среднее цис-транс-отношение для карт *D. discoideum* составляет 4,19, что является близким к цис-транс-отношениям, указанным для типичных карт Hi-C млекопитающих [275]. Мы также отметили, что все хромосомы *D. discoideum*, кроме хромосомы 5, имеют свои центромеры близко к 5'-концу хромосом, а хромосома 5 — «перевернута» в сборке, и ее центромера расположена у 3'-конца хромосомы. В связи с большим количеством повторов, влияющий на качество карты, уточнить положение центромер и теломер в хромосомах *D. discoideum* не удалось.

Из литературы известен феномен дальних взаимодействий локусов *piwi*РНК у *D. melanogaster*; по-видимому, они являются регуляторами активности транспозонов [250]. Также были описаны структуры KNOT, которые представляют собой дальние взаимодействия локусов, которые обогащены генами малых РНК, а также являются предпочтительным местом посадки

транспозонов в геноме *A. thaliana* [276; 277]. Кроме того, было показано, что на картах Hi-C такие структуры выглядят как сеть точечных дальних взаимодействий [277]. В нашей работе были обнаружены дальние цис- и транс-контакты, специфичные для 0 и 8 часов развития *D. discoideum*. Более того, многие локусы образуют несколько дальних контактов одновременно. Нами было показано, что в основаниях таких контактов перепредставлены гены длинных некодирующих РНК; рассчитанная по пермутационному тесту значимость отличия от случайного для пересечений оснований и таких генов составила $< 0,001$. Было также показано, что гены длинных некодирующих РНК в основаниях дальних контактов меняют свой уровень экспрессии между часами развития, однако, вследствие недостаточно точной аннотации генов длинных некодирующих РНК, выяснить их функции в рамках данной работы не удалось.

Компартменты представляют собой один из уровней организации хроматина, в основе которого лежит жидкость-жидкостное разделение фаз [131]. В литературе описаны компартменты у растений *Arabidopsis thaliana*, *Oryza sativa*, *Zea mays* [278–281], *Drosophila melanogaster* [282], *Danio rerio* [283] и др., а также различных у млекопитающих [284]. Настоящая работа позволяет добавить губку *Halisarca dujardini* к этому ряду организмов, поскольку в её хроматине присутствует выраженный паттерн «шахматная доска», а численная оценка подтверждает визуальные наблюдения. Также, по-видимому, по результатам настоящей работы, к таким организмам стоит отнести *D. discoideum*, для которого результаты численного анализа свидетельствуют о наличии слабой компартментализации генома.

Кроме того, описаны эффекты нарушения компартментализации как часть патогенеза различных заболеваний [285]. В нашей работе на примере трисомий человека мы показали, что наличие дополнительной копии хромосомы нарушает глобальную укладку хроматина, что проявляется в виде компактизации одних регионов и декомпактизации других. Мы показали также, что для декомпактизованных регионов наблюдается более частая принадлежность к неактивному компартменту, которая достигает 90 % для Tr18, что свидетельствует о преимущественной декомпактизации именно регионов гетерохроматина.

Хромосомы в ядрах эукариот прикрепляются к ламине и ядерную оболочку, а участки их крепления называются, соответственно, ламина-

ассоциированные домены (ЛАДы) и ядрышко-ассоциированные домены [286]. В работе [287] было показано, что лишь 30 % ЛАДов ассоциированы с ламиной в каждой отдельно клетке. Было также показано, что именно расположение на ядерной периферии вблизи ламины позволяет поддерживать репрессивное состояние генов [288]. Таким образом, хромосомы, содержащие значительную долю ЛАДов, могут конкурировать за локализацию на ядерной периферии, а дополнительная копия хромосомы при трисомии выступает еще одним конкурентом в этом взаимодействии. Это согласуется с нашими наблюдениями по изменению профилей контактов практически всех хромосом в клетках с трисомией. Мы обнаружили, что именно для трисомий по хромосомам 13 и 18, но не для трисомии по хромосоме 16, характерна сильная зависимость между долей локусов с измененной компактизацией и содержанием ЛАДов в хромосоме. Поскольку покрытие ламина-ассоциированными доменами у хромосом 13 и 18 очень похоже (и примерно в три раза выше, чем у хромосомы 16), это указывает на то, что нарушения взаимодействий хроматин-ламина являются потенциальной причиной изменений в компактизации хроматина внутри хромосом.

Ведущей теорией формирования петель и ТАДов в настоящее время является экструзия петель [289]. У высших эукариот когезин-зависимые петли возникают преимущественно благодаря наличию двух сайтов связывания CTCF в конвергентной ориентации [150]. Кроме того, недавние работы показывают роль других факторов, например, натяжения ДНК [182], РНК-ДНК гибридов [290] и длинных некодирующих РНК [291] в регуляции эффективности и проницаемости этого барьера. Среди других барьеров выделяют движущуюся РНК-полимеразу [187; 292] и МСМ, один из белков репликационного комплекса [293]. В настоящей работе большое внимание было уделено точечным цис-взаимодействиям *D. discoideum*. Было показано, что такие взаимодействия (названные хроматиновыми петлями) представляют собой высококонсервативные между часами развития последовательно расположенные структуры. Кроме того, были изучены вытянутые хроматиновые петли и показано, что их асимметрия может быть обусловлена различиями экспрессии генов вокруг оснований, что также исследуется для хроматиновых фонтанов у *C. elegans* [158], аналогично асимметричным фонтанам в хроматине *Caenorhabditis elegans* [158]. Основываясь на полученных данных,

мы предполагаем, что границы петель формируются за счет взаимодействия между транскрипцией генов конвергентных пар и экструзией когезина.

В работе [294] с помощью анализа данных MicroC-XL на почкующихся дрожжах *Saccharomyces cerevisiae* были охарактеризованы иерархические хроматиновые петли, найденные как в S-фазе, так и в митозе. Для аннотированных петель была показана высокая доля последовательно расположенных оснований, а дальнейшие работы показали, что барьерами для экструзии, по-видимому, выступают репликационная вилка [184] на S-фазе, а также когезин, участвующий в когезии сестринских хроматид [185].

Из литературы известны также особенности организации генома представителя динофлагеллят *Breviolum minutum* [295] — хроматин уложен в крупные ТАДы (названные dinoTADs), в границах которых также расположены конвергентно ориентированные гены. Однако dinoTADs обладают и другим примечательным свойством — в Hi-C отсутствуют точечные (петлевые) цис-взаимодействия. Полученные результаты подтверждаются работой на другом представителе динофлагеллят — *Symbiodinium microadriaticum* [296].

О конвергентных генах, участвующих в формировании границ петель и ТАДов, известно и из экспериментов на других организмах, например *Saccharomyces cerevisiae* [297]. Корреляция между локализацией когезина и конвергентно ориентированными парами генов также была обнаружена в клетках человека [298]. Обычно считается, что это отражает, как транскрибирующие РНК-полимеразы двигаются в противоположном экструзии направлении и препятствуют движению экструдера по ДНК, но прямых доказательств этого недостаточно. Кроме того, из этой гипотезы неявно следует, что одного активно транскрибируемого гена, движения РНК полимеразы на котором противоположны движению экструдера, было бы достаточно для образования петель.

В работе на динофлагеллятах [295] также было показано, что ингибирование транскрипции разрушает ТАДы, что указывает не только на роль РНК полимеразы 2 как барьера транскрипции, но и на роль суперспирализации ДНК вследствие активной транскрипции в качестве основной причины наблюдаемой пространственной организации. Более того, в работе [299] было показано повышенное накопление положительной суперспирализации и РНК-ДНК гибридов между конвергентными парами генов, что также может

свидетельствовать об их роли в формировании хроматиновых петель у *D. discoideum*.

Наблюдения, сделанные на хроматине *D. discoideum*, могут быть объяснены предположением, что КППГ действует как барьер для управляемой когезином экструзии петель. Когезины, движущиеся в обоих направлениях, останавливаются на вв-КППГ (и внутри них), таким образом уплотняя их в контактные домены. Когезины, захваченные КППГ, которые продолжают однонаправленную экструзию, образуют полосы на картах контактов. Слабо транскрибируемый ген представляет собой слабый барьер, пропускающий когезин. Это приводит к частичному разуплотнению контактного домена в-КППГ и потере полосы на карте контактов. Поскольку активно транскрибируемые КППГ с короткими межгенниками накапливают высокий уровень положительной суперспирализации у дрожжей [299], а суперспирализация может представлять собой барьер для движения когезина [300], самая высокая частота контактов между КППГ длинных генов с короткими межгенниками косвенно подтверждает предположение, что петли между КППГ образуются посредством экструзии, управляемой когезином.

В литературе [159] было показано, что фонтаны являются результатом загрузки когезина на энхансеры, связанные с ранним эмбриональным развитием, и что на более поздних стадиях в тех же локусах наблюдаются петли и ТАДы, которые являются чертой продолжающейся экструзии. На примере локуса ферритина мы проследили изменения хроматина эмбрионов *Danio rerio* до 20 часов после оплодотворения, где фонтаны ранних стадий превращаются в ТАДы и петли.

С использованием данных ChIP-seq для гистоновых модификаций и данных об открытости хроматина из работы [244], аннотированы цис-регуляторные последовательности, которые статистически значимо обогащены в основаниях петель в сравнении со случайными последовательностями той же длины. По совокупности полученных результатов можно предположить, что обнаруженные цис-регуляторные элементы в *D. discoideum* представляют не стадий-специфичный регуляторный элемент, а скорее некоторый инструмент поддержания стабильной экспрессии генов, в которых они расположены. Такое поведение регуляторных элементов согласуется с данными литературы [84], где отмечают преобладающую регуляцию

на близких расстояниях для большинства низших эукариот, что может быть связано с компактностью генома.

Однако открытым вопросом остается изменение структуры хроматина при изменении уровня экспрессии. Из полученных результатов на *D. discoideum*, где движущаяся РНК полимеразы, по-видимому, является барьером для когезина, противодействуя экстррузии (встречные гены) или замедляя движение когезина (гены, транскрипция которых идет в том же направлении, что и экстррузия) не удается сделать вывод о наличии подобной связи на уровне петель. Предполагается, что причин этому может быть несколько: 1) усреднения по множеству клеток (bulk) не позволяют детектировать изменения, которые происходят на уровне отдельных клеток; 2) изменения в экспрессии могут запаздывать в хроматине, т.е. скорость изменения хроматина происходит в течение нескольких часов после изменения транскрипции, как это было описано в [187]; 3) существует некоторый третий фактор, участвующий в поддержании стабильности петель, например, суперспирализация ДНК [301] или РНК-ДНК гибриды [299]. Однако нам удалось показать зависимость частоты контактов в контактном домене, формируемом на теле экспрессируемого гена, от уровня транскрипции последнего, а также меньшую частоту контактов со стороны гена с низкой экспрессией в конвергентной паре с высокой и низкой транскрипцией в сравнении с парой, где оба гена принадлежат к группе высокой экспрессии.

Еще одним открытым остается вопрос о роли фонтанов в формировании структуры хроматина губок *Petrosia ficiformis* и *Halisarca dujardini*. В их жизненном цикле отсутствует стадия очевидного аналога эмбрионального развития млекопитающих, а из имеющихся в литературе данных можно заключить, что фонтаны являются чертой именно раннего эмбрионального развития [159]. В рамках нашего исследования на губках не были детектированы выраженные ТАДы или петли, что может быть связано с низкой процессивностью когезина или низкой плотностью его загрузки. Однако точная роль фонтанов и механизм их образования в хроматине губок остается неизвестным.

Заключение

В настоящей работе было проведено исследование структуры хроматина почвенной амебы *Dictyostelium discoideum*, морской губки *Halisarca dujardini* и образцов с трисомиями человека *Homo sapiens* на различных уровнях организации, а также взаимосвязи этих структур с эпигенетикой и транскрипционной программой на примере петель *D. discoideum* и изменения локальной упаковки хроматина образцов с трисомиями. Все поставленные задачи были выполнены, а именно: анализ данных Hi-C на разных уровнях организации *D. discoideum* и *H. dujardini* свидетельствует в пользу структуры Рабля с межтеломерными и межцентромерными контактами, а также о компарментализации хроматина обоих организмов. В хроматине *H. dujardini* обнаружены хроматиновые фонтаны. В хроматине *D. discoideum* выявлены специфические для стадия развития точечные взаимодействия, обусловленные контактами генов длинных некодирующих РНК, у которых происходит изменение уровней экспрессии между стадиями. Значительная доля генома *D. discoideum* упакована в последовательные петли, высококонсервативные между стадиями развития, в формировании которых участвуют конвергентно расположенные экспрессируемые гены. Уровень экспрессии генов в конвергентной паре значительно меняет структуру контактов вокруг пары, что свидетельствует о барьерной функции генов, чья проницаемость зависит от их уровня экспрессии.

Сравнительный анализ хроматинового окружения локуса ферритина у двух представителей типа Porifera, *P. ficiformis* и *H. dujardini*, рыбы *D. rerio*, мыши *M. musculus* и человека *H. sapiens* показал, что в хроматине губок, но не *D. rerio*, основание хроматинового фонтана находится в непосредственной близости от генов консервативных ферритинов, а в хроматине *D. rerio* на стадии 20 часов после оплодотворения, а также у млекопитающих *M. musculus* и *H. sapiens* ген ферритина расположен внутри ТАДа, чья структура только незначительно меняется при изменении уровня экспрессии.

Наличие лишней хромосомы в ядре может систематически влиять на частоту контактов между другими хромосомами и превышает степень изменчивости между отдельными здоровыми донорами у образцов с трисомиями. На уровне локальных изменений упаковки хроматина для декомпактизо-

ванных регионов отмечена более частая принадлежность к В-компарменту, которая достигает 90 % для Tr18, что свидетельствует о преимущественной декомпактизации именно регионов гетерохроматина.

В нашей работе впервые проанализировано устройство хроматина у гаплоидной амебы *D. discoideum*, а также изучены взаимосвязи структур с эпигенетическими особенностями и экспрессией генов. Впервые аннотированы дальние контакты (зачастую транс-) в хроматине *D. discoideum* и показана их взаимосвязь с длинными некодирующими РНК, значительно меняющими свои уровни экспрессии между стадиями развития. Впервые обнаружены вытянутые петли и показана их связь с уровнями транскрипции вокруг оснований этих петель. Впервые выдвинута гипотеза о конвергентной транскрипции как о механизме формирования петель у *D. discoideum*. Впервые аннотированы энхансеры в геноме *D. discoideum* и показано их обогащение в основании петель. Впервые аннотированы фонтаны в хроматине морской губки *H. dujardini*. Впервые проанализировано устройство хроматина образцов с трисомиями по хромосомам 13, 16 и 18 у человека. Впервые показана связь изменения частот контактов хромосом с трисомиями по хромосомам 13, 16 и 18 с содержанием ламина-ассоциированных доменов на таких хромосомах.

Результаты нашей работы позволяют говорить о том, что: 1. Хромосомы *D. discoideum* и *H. dujardini* образуют структуру Рабля с обогащенными межтеломерными и межцентромерными взаимодействиями. 2. Хроматин *D. discoideum* образует множество последовательных петель с высокой консервативностью между стадиями и низким уровнем иерархичности, а их основания колокализуются с границами синтенных блоков. 3. Хроматин губки *H. dujardini* образует фонтаны; различные хроматиновые петли и топологически ассоциированные домены отсутствуют. 4. Наличие дополнительной копии хромосом вызывает как изменения в межхромосомных взаимодействиях, так и реорганизацию локальной упаковки хроматина, причём степень этих изменений варьирует для различных хромосом и, по-видимому, связана с изменениями взаимодействий с белками ламины.

Разработанная методология может быть применена для изучения устройства хроматина других беспозвоночных. Кроме того, настоящее исследование не только восполняет пробел об устройстве хроматина двух важных организмов, но и проливает свет на разнообразие устройства хро-

матина организмов в отсутствие белка CTCF. Результаты работы показали, что конвергентная пара генов может быть рассмотрена как барьер для экстррузии в других организмах. Полученные данные могут быть использованы в изучении разнообразия устройства хроматина организмов в отсутствие белка CTCF.

Основные результаты работы заключаются в следующем.

1. Хромосомы *D. discoideum* и *H. dujardini* образуют структуру Рабля с обогащенными межтеломерными и межцентромерными контактам.
2. Хроматин *H. dujardini* компартиментализован, и в нем присутствуют хроматиновые фонтаны.
3. Хроматин *D. discoideum* организован в преимущественно последовательные, неиерархические петли, внутренняя часть которых слабо изолирована от соседних локусов.
4. Основания петель представляют собой сайты конвергентной ориентации генов.
5. Сила оснований петель, отражающая эффективность барьерной функции для экстррузии, зависит от силы окружающей основания конвергентной транскрипции.
6. Конвергентные пары генов, расположенные в основаниях петель, образуют контактные домены, структура которых зависит от уровня транскрипции обоих генов в паре.
7. В геномах губок присутствуют гены типичных ферритинов (один или два паралога) и гены атипичного ферритина с аминокислотными заменами в консервативных доменах; основание хроматинового фонтана находится в непосредственной близости от генов консервативных ферритинов.
8. Трисомии человека по хромосомам 13, 16 и 18 приводят к изменению укладки хромосом как на уровне межхромосомных контактов, так и на уровне локальных взаимодействий.
9. При трисомии по трем изученным хромосомам значительно меняется взаимодействие внутри малых хромосом (хромосомы 16-22), а также взаимодействие малых и больших хромосом друг с другом по сравнению с нормальным кариотипом.

10. Локальные хроматиновые взаимодействия (на расстоянии не более 1 млн. п.н.) значительно меняются внутри малых хромосом по сравнению с контролем при трисомии по хромосомам 13 и 18, но не 16.

В заключение автор выражает благодарность и большую признательность научным руководителям Михаилу Сергеевичу Гельфанду и Екатерине Евгеньевне Храмеевой за помощь в работе, всестороннее участие и научное руководство. Глубокую признательность автор выражает Сергею Владимировичу Разину за доброе отношение, а также Сергею Владимировичу Ульянову за продуктивные и всесторонние обсуждения результатов работы, а также совместную подготовку материалов к публикации и написание статей. Автор благодарит коллектив лаборатории Е.Е. Храмеевой в Сколтехе и УНЦ «Биоинформатика» Института проблем передачи информации им. А.А.Харкевича РАН за помощь в работе и научные консультации, а также всех сотрудников и аспирантов Лаборатории структурно-функциональной организации хромосом Института биологии гена РАН за плодотворное сотрудничество и создание теплой атмосферы. Автор благодарит Александру Галицыну за помощь в расчёте показателя конвергентности, визуализации профиля сигнала вокруг петель и графика среднего сигнала хромосомы, а также обсуждения результатов работы, Илью Плетенева за помощь в анализе компартментов, дальних контактов и изучении структуры Рабля у *D. discoideum*; Александра Черкасова за подготовку библиотек ДНК и РНК для экспериментов Hi-C и RNA-seq, а также помощь в анализе компартментов, дальних контактов, изучении структуры Рабля *H. dujardini*; Илью Флямера за предварительный анализ данных Hi-C образцов с трисомиями, Николая Сафронова за обсуждения материалов работы и помощь в оформлении диссертации, а также авторов репозитория Russian-PhD-LaTeX-Dissertation-Template за помощь с оформлением работы по ГОСТу. Автор благодарен своим прошлым и нынешним студентам, Кириллу Ульянову, Василию Зубареву, Анастасии Каштановой, Евгению Егорову и Яне Дрозд, за их вдохновляющие вопросы и искренний интерес к науке.

Список литературы

1. *Barry, N. P.* Dictyostelium amoebae and neutrophils can swim / N. P. Barry, M. S. Bretscher // *Proceedings of the National Academy of Sciences*. — 2010. — Vol. 107, no. 25. — P. 11376—11380.
2. *Maniak, M.* Dictyostelium as a model for human lysosomal and trafficking diseases / M. Maniak // *Seminars in Cell & Developmental Biology*. — 2011. — Vol. 22, no. 1. — P. 114—119. — (GTPases in Intracellular Trafficking).
3. *Loomis, W. F.* Cell signaling during development of Dictyostelium / W. F. Loomis // *Developmental Biology*. — 2014. — Vol. 391, no. 1. — P. 1—16.
4. *Nichols, J. M.* Chemotaxis of a model organism: progress with Dictyostelium / J. M. Nichols, D. Veltman, R. R. Kay // *Current Opinion in Cell Biology*. — 2015. — Vol. 36. — P. 7—12. — (Cell adhesion and migration).
5. *Tosetti, N.* Amoebae as a tool to isolate new bacterial species, to discover new virulence factors and to study the host–pathogen interactions / N. Tosetti, A. Croxatto, G. Greub // *Microbial Pathogenesis*. — 2014. — Vol. 77. — P. 125—130.
6. Curcumin affects gene expression and reactive oxygen species via a PKA dependent mechanism in Dictyostelium discoideum / W. S. Swatson [et al.] // *PLOS ONE*. — 2017. — Vol. 12, no. 11. — e0187562.
7. Identifying an uptake mechanism for the antiepileptic and bipolar disorder treatment valproic acid using the simple biomedical model Dictyostelium / N. Terbach [et al.] // *Journal of Cell Science*. — 2011. — Vol. 124, no. 13. — P. 2267—2276.
8. The Dictyostelium model for mitochondrial disease / L. M. Francione [et al.] // *Seminars in Cell & Developmental Biology*. — 2011. — Vol. 22, no. 1. — P. 120—130. — (GTPases in Intracellular Trafficking).

9. Integrated actions of mTOR complexes 1 and 2 for growth and development of *Dictyostelium* / P. Jaiswal [et al.] // *International Journal of Developmental Biology*. — 2019. — Vol. 63, no. 8. — P. 521—527.
10. *Myre, M. A.* Functional Analysis of Proteins Involved in Neurodegeneration Using the Model Organism *Dictyostelium*: Alzheimer's, Huntington's, and Batten Disease / M. A. Myre, R. J. Huber, D. H. O'Day // *Molecular-Genetic and Statistical Techniques for Behavioral and Neural Research* / ed. by R. T. Gerlai. — San Diego : Academic Press, 2018. — P. 491—518.
11. The Parkinson's Disease-Associated Protein DJ-1 Protects *Dictyostelium* Cells from AMPK-Dependent Outcomes of Oxidative Stress / S. Chen [et al.] // *Cells*. — 2021. — Aug. — Vol. 10, no. 8. — P. 1874.
12. *Dictyostelium*, a microbial model for brain disease / S. J. Annesley [et al.] // *Biochimica et Biophysica Acta (BBA) - General Subjects*. — 2014. — Vol. 1840, no. 4. — P. 1413—1432. — (Frontiers of Mitochondrial Research).
13. Deep Phylogeny and Evolution of Sponges (Phylum Porifera) / G. Worheide [et al.] // *Adv Mar Biol*. — 2012. — Vol. 61. — P. 1—78. — pmid: 22560777.
14. The Demosponge *Amphimedon Queenslandica*: Reconstructing the Ancestral Metazoan Genome and Deciphering the Origin of Animal Multicellularity / B. M. Degnan [et al.] // *Cold Spring Harb Protoc*. — 2008. — Vol. 2008, no. 12. — pdb.emo108. — pmid: 21356734.
15. Dynamic and Widespread lncRNA Expression in a Sponge and the Origin of Animal Complexity / F. Gaiti [et al.] // *Molecular Biology and Evolution*. — 2015. — Vol. 32, no. 9. — P. 2367—2382.
16. The Mid-Developmental Transition and the Evolution of Animal Body Plans / M. Levin [et al.] // *Nature*. — 2016. — Mar. — Vol. 531, issue 7596, no. 7596. — P. 637—641.
17. Transdifferentiation and Mesenchymal-to-Epithelial Transition during Regeneration in Demospongiae (Porifera) / A. V. Ereskovsky [et al.] // *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*. — 2020. — Vol. 334, no. 1. — P. 37—58.
18. Pluripotency and the Origin of Animal Multicellularity / S. Sogabe [et al.] // *Nature*. — 2019. — June. — Vol. 570, issue 7762, no. 7762. — P. 519—522.

19. Whole-Body Regeneration in Sponges: Diversity, Fine Mechanisms, and Future Prospects / A. Ereskovsky [et al.] // *Genes*. — 2021. — Apr. — Vol. 12, issue 4, no. 4. — P. 506.
20. *Wilson, H. V.* On Some Phenomena of Coalescence and Regeneration in Sponges / H. V. Wilson // *Journal of Experimental Zoology*. — 1907. — Vol. 5, no. 2. — P. 245—258.
21. Transdifferentiation Is a Driving Force of Regeneration in *Halisarca Dujardini* (Demospongiae, Porifera) / I. E. Borisenko [et al.] // *PeerJ*. — 2015. — Vol. 3. — e1211.
22. Current Estimate of Down Syndrome Population Prevalence in the United States / A. P. Presson [et al.] // *J Pediatr*. — 2013. — Oct. — Vol. 163, no. 4. — P. 1163—1168. — pmid: 23885965.
23. Survival and Surgical Interventions for Children With Trisomy 13 and 18 / K. E. Nelson [et al.] // *JAMA*. — 2016. — Vol. 316, no. 4. — P. 420—428. — pmid: 27458947.
24. A Chromosome 21 Critical Region Does Not Cause Specific Down Syndrome Phenotypes / L. E. Olson [et al.] // *Science*. — 2004. — Vol. 306, no. 5696. — P. 687—690. — pmid: 15499018.
25. Down Syndrome / S. E. Antonarakis [et al.] // *Nat Rev Dis Primers*. — 2020. — Vol. 6, no. 1. — P. 9. — pmid: 32029743.
26. *Krivega, M.* Consequences of Chromosome Gain: A New View on Trisomy Syndromes / M. Krivega, C. M. Stiefel, Z. Storchova // *Am J Hum Genet*. — 2022. — Vol. 109, no. 12. — P. 2126—2140. — pmid: 36459979.
27. Sequence-specific protein interaction with a transcriptional enhancer involved in the autoregulated expression of cAMP receptor 1 in *Dictyostelium* / X. Mu [et al.] // *Development*. — 1998. — Vol. 125, no. 18. — P. 3689—3698.
28. *Hori, R.* Identification and characterization of multiple A/T-rich cisacting elements that control expression from *Dictyostelium* actin promoters: the *Dictyostelium* actin upstream activating sequence confers growth phase expression and has enhancer-like properties / R. Hori, R. A. Firtel // *Nucleic Acids Research*. — 1994. — Vol. 22, no. 23. — P. 5099—5111.

29. *Fosnaugh, K. L.* Enhancer Regions Responsible for Temporal and Cell-Type-Specific Expression of a Spore Coat Gene in Dictyostelium / K. L. Fosnaugh, W. F. Loomis // *Developmental Biology*. — 1993. — Vol. 157, no. 1. — P. 38—48.
30. Leaps and Lulls in the Developmental Transcriptome of Dictyostelium Discoideum / R. D. Rosengarten [et al.] // *BMC Genomics*. — 2015. — Vol. 16, no. 1. — P. 294.
31. *Kessin, R. H.* Dictyostelium: Evolution, Cell Biology, and the Development of Multicellularity / R. H. Kessin. — Cambridge University Press, 2001. — 318 p. — Google Books: rm7wHyUYSiAC.
32. Evolution of Multicellularity in Dictyostelia / Y. Kawabe [et al.] // *The International Journal of Developmental Biology*. — 2019. — Vol. 63, issue 8-9—10, no. 8-9—10. — P. 359—369.
33. Abundantly Expressed Class of Noncoding RNAs Conserved through the Multicellular Evolution of Dictyostelid Social Amoebas / J. Kjellin [et al.] // *Genome Res*. — 2021. — Vol. 31, no. 3. — P. 436—447. — pmid: 33479022.
34. The Multicellularity Genes of Dictyostelid Social Amoebas / G. Glockner [et al.] // *Nat Commun*. — 2016. — Vol. 7, issue 1, no. 1. — P. 12085.
35. When Dicty Met Myco, a (Not So) Romantic Story about One Amoeba and Its Intracellular Pathogen / E. Cardenal-Munoz [et al.] // *Frontiers in Cellular and Infection Microbiology*. — 2018. — Vol. 7.
36. Interactions and Cytotoxicity of Human Neurodegeneration- Associated Proteins Tau and α -Synuclein in the Simple Model Dictyostelium Discoideum / K. Mroczek [et al.] // *Frontiers in Cell and Developmental Biology*. — 2021. — Vol. 9.
37. *Hammond, W.* Insanity / W. Hammond // *Treatise on Diseases of the Nervous system*. — D. Appleton, New York, 1873.
38. *Meunier, H.* Pharmacodynamic properties of N-dipropylacetic acid / H. Meunier // *Therapie*. — 1963. — T. 18. — C. 435—438.
39. Loss of a prolyl oligopeptidase confers resistance to lithium by elevation of inositol (1,4,5) trisphosphate / R. Williams [et al.] // *The EMBO Journal*. — 1999. — Vol. 18, no. 10. — P. 2734—2745.

40. A common mechanism of action for three mood-stabilizing drugs / R. S. B. Williams [et al.] // *Nature*. — 2002. — May. — Vol. 417, no. 6886. — P. 292—295.
41. The genome of the social amoeba *Dictyostelium discoideum* / L. Eichinger [et al.] // *Nature*. — 2005. — May. — Vol. 435, no. 7038. — P. 43—57.
42. The Complex Repeats of *Dictyostelium Discoideum* / G. Glockner [et al.] // *Genome Res.* — 2001. — Vol. 11, no. 4. — P. 585—594.
43. Dynamics of a novel centromeric histone variant CenH3 reveals the evolutionary ancestral timing of centromere biogenesis / M. Dubin [et al.] // *Nucleic Acids Research*. — 2010. — Vol. 38, no. 21. — P. 7526—7537.
44. *Glockner, G.* Centromere sequence and dynamics in *Dictyostelium discoideum* / G. Glockner, A. J. Heidel // *Nucleic Acids Research*. — 2009. — Vol. 37, no. 6. — P. 1809—1816.
45. *Kaller, M.* Differential Effects of Heterochromatin Protein 1 Isoforms on Mitotic Chromosome Distribution and Growth in *Dictyostelium Discoideum* / M. Kaller, U. Euteneuer, W. Nellen // *Eukaryotic Cell*. — 2006. — Vol. 5, no. 3. — P. 530—543. — pmid: 16524908.
46. Chromatin organization changes between life stages of soil-living amoeba *Dictyostelium discoideum* / O. Tsoy [et al.] // 2019 IMB Conference - Chromosome Territories & Nuclear Architecture (2019 IMB Conference - Chromosome Territories & Nuclear Architecture). — Mainz, Germany, 10/2019.
47. The *Dictyostelium* Centrosome / R. Graf [et al.] // *Cells*. — 2021. — Oct. — Vol. 10, no. 10. — P. 2657.
48. The *Dictyostelium Discoideum* RNA-dependent RNA Polymerase RrpC Silences the Centromeric Retrotransposon DIRS-1 Post-Transcriptionally and Is Required for the Spreading of RNA Silencing Signals / S. Wiegand [et al.] // *Nucleic Acids Research*. — 2014. — Mar. — Vol. 42, no. 5. — P. 3330. — pmid: 24369430.
49. *Malicki, M.* Retrotransposon Domestication and Control in *Dictyostelium Discoideum* / M. Malicki, M. Iliopoulou, C. Hammann // *Front Microbiol.* — 2017. — T. 8. — pmid: 29051748.

50. Role of RNA Polymerase III Transcription Factors in the Selection of Integration Sites by the Dictyostelium Non-Long Terminal Repeat Retrotransposon TRE5-A / O. Siol [et al.] // *Mol Cell Biol.* — 2006. — Vol. 26, no. 22. — P. 8242—8251.
51. Convergent Evolution of tRNA Gene Targeting Preferences in Compact Genomes / T. Spaller [et al.] // *Mobile DNA.* — 2016. — Vol. 7.
52. Structure of Dictyostelium discoideum telomeres. Analysis of possible replication mechanisms / J. Rodriguez-Centeno [et al.] // *PLoS ONE.* — 2019. — Vol. 14, no. 9.
53. *Nabetani, A.* Alternative lengthening of telomeres pathway: Recombination-mediated telomere maintenance mechanism in human cells / A. Nabetani, F. Ishikawa // *The Journal of Biochemistry.* — 2011. — Vol. 149, no. 1. — P. 5—14.
54. Characterization of NE81, the first lamin-like nucleoskeleton protein in a unicellular organism / A. Kruger [и др.] // *Molecular Biology of the Cell.* — 2011. — Т. 23, № 2. — С. 360—370.
55. *Van Bortle, K.* Spinning the Web of Cell Fate / K. Van Bortle, V. G. Corces // *Cell.* — 2013. — Vol. 152, no. 6. — P. 1213—1217.
56. Nuclear lamina integrity is required for proper spatial organization of chromatin in *Drosophila* / S. V. Ulianov [et al.] // *Nature Communications.* — 2019. — Vol. 10, no. 1. — P. 1176.
57. Cell-type specific RNA-Seq reveals novel roles and regulatory programs for terminally differentiated Dictyostelium cells / K. Kin [et al.] // *BMC Genomics.* — 2018. — Vol. 19, no. 1. — P. 764.
58. *Rosengarten, R.* Transcriptional Regulators: Dynamic Drivers of Multicellular Formation, Cell Differentiation and Development / R. Rosengarten, B. Santhanam, M. Katoh-Kurasawa // *Dictyostelids: Evolution, Genomics and Cell Biology* / ed. by M. Romeralo, S. Baldauf, R. Escalante. — Berlin, Heidelberg : Springer, 2013. — P. 89—108.
59. Transcriptional Milestones in Dictyostelium Development / M. Katoh-Kurasawa [et al.] // *Genome Res.* — 2021. — Vol. 31, no. 8. — P. 1498—1511. — pmid: 34183452.

60. *Stevense, M.* Nuclear organization and transcriptional dynamics in *Dictyostelium* / M. Stevense, J. R. Chubb, T. Muramoto // *Development, Growth & Differentiation*. — 2011. — Vol. 53, no. 4. — P. 576—586.
61. *Kouzarides, T.* Chromatin Modifications and Their Function / T. Kouzarides // *Cell*. — 2007. — T. 128, № 4. — С. 693—705.
62. *Fukuzawa, M.* Control of prestalk-cell differentiation by transcription factors / M. Fukuzawa // *Development, Growth & Differentiation*. — 2011. — Vol. 53, no. 4. — P. 538—547.
63. The GATA transcription factor *GtaC* regulates early developmental gene expression dynamics in *Dictyostelium* / B. Santhanam [et al.] // *Nature Communications*. — 2015. — Vol. 6, no. 1. — P. 7551.
64. Nucleocytoplasmic Shuttling of a GATA Transcription Factor Functions as a Development Timer / H. Cai [et al.] // *Science*. — 2014. — Vol. 343, no. 6177.
65. *Yamada, Y.* The transcription factor *Spores Absent A* is a PKA dependent inducer of *Dictyostelium* sporulation / Y. Yamada, A. Cassidy, P. Schaap // *Scientific Reports*. — 2018. — Vol. 8, no. 1. — P. 6643.
66. *BzpF* is a CREB-like transcription factor that regulates spore maturation and stability in *Dictyostelium* / E. Huang [и др.] // *Developmental biology*. — 2011. — T. 358, № 1. — С. 137—146.
67. Restricted inactivation of serum response factor to the cardiovascular system / J. M. Miano [et al.] // *Proceedings of the National Academy of Sciences*. — 2004. — Vol. 101, no. 49. — P. 17132—17137.
68. *SrfB*, a member of the Serum Response Factor family of transcription factors, regulates starvation response and early development in *Dictyostelium* / M. Galardi-Castilla [et al.] // *Developmental Biology*. — 2008. — Vol. 316, no. 2. — P. 260—274.
69. *Escalante, R.* A Serum Response Factor homolog is required for spore differentiation in *Dictyostelium* / R. Escalante, L. Sastre // *Development*. — 1998. — Vol. 125, no. 19. — P. 3801—3808.
70. *Chang, W.-T.* Identification of the Cell Fate Gene *Stalky* in *Dictyostelium* / W.-T. Chang, P. C. Newell, J. D. Gross // *Cell*. — 1996. — Vol. 87, no. 3. — P. 471—481.

71. *Simpson, T. L.* The Cell Biology of Sponges / T. L. Simpson. — New York, NY : Springer, 1984.
72. *Ereskovsky, A. V.* The Comparative Embryology of Sponges / A. V. Ereskovsky. — Dordrecht : Springer Netherlands, 2010.
73. Genomics of Sponge-Associated *Streptomyces* Spp. Closely Related to *Streptomyces Albus J1074*: Insights into Marine Adaptation and Secondary Metabolite Biosynthesis Potential / E. Ian [et al.] // PLOS ONE. — 2014. — Vol. 9, no. 5. — e96719.
74. *Ereskovsky, A. V.* Morphological Evidence for Vertical Transmission of Symbiotic Bacteria in the Viviparous Sponge *Halisarca Dujardini* Johnston (Porifera, Demospongiae, Halisarcida) / A. V. Ereskovsky, E. Gonobobl-eva, A. Vishnyakov // Marine Biology. — 2005. — Vol. 146, no. 5. — P. 869—875.
75. *Schmittmann, L.* Individuality in the Immune Repertoire and Induced Response of the Sponge *Halichondria Panicea* / L. Schmittmann, S. Franzenburg, L. Pita // Frontiers in Immunology. — 2021. — T. 12.
76. *Fernandez-Valverde, S. L.* Bilaterian-like Promoters in the Highly Compact Amphimedon *Queenslandica* Genome / S. L. Fernandez-Valverde, B. M. Degnan // Sci Rep. — 2016. — Vol. 6, issue 1, no. 1. — P. 22496.
77. Landscape of Histone Modifications in a Sponge Reveals the Origin of Animal Cis-Regulatory Complexity / F. Gaiti [и др.] // eLife / под ред. R. Krumlauf. — 2017. — T. 6. — e22194.
78. A Comparative Genomics Multitool for Scientific Discovery and Conservation / D. P. Genereux [et al.] // Nature. — 2020. — Nov. — Vol. 587, issue 7833, no. 7833. — P. 240—245.
79. The Evolution of Invertebrate Gene Body Methylation / S. Sarda [и др.] // Molecular Biology and Evolution. — 2012. — T. 29, № 8. — С. 1907—1916.
80. *The Darwin Tree of Life Project Consortium.* Sequence Locally, Think Globally: The Darwin Tree of Life Project / The Darwin Tree of Life Project Consortium // Proceedings of the National Academy of Sciences. — 2022. — T. 119, № 4. — e2115642118.

81. *Lyko, F.* The DNA Methyltransferase Family: A Versatile Toolkit for Epigenetic Regulation / F. Lyko // *Nat Rev Genet.* — 2018. — Feb. — Vol. 19, issue 2, no. 2. — P. 81–92.
82. Extensive Conservation of Ancient Microsynteny across Metazoans Due to Cis-Regulatory Constraints / M. Irimia [et al.] // *Genome Res.* — 2012. — Vol. 22, no. 12. — P. 2356–2367. — pmid: 22722344.
83. Ancient Cis-Regulatory Constraints and the Evolution of Genome Architecture / M. Irimia [et al.] // *Trends in Genetics.* — 2013. — Vol. 29, no. 9. — P. 521–528.
84. Origin and Evolution of the Metazoan Non-Coding Regulatory Genome / F. Gaiti [и др.] // *Developmental Biology.* — 2017. — Т. 427, № 2. — С. 193–202. — (Changes in Developmental Controls Underlying the Evolution of Animal Body Plans).
85. *Sadler, K. C.* Epigenetics across the Evolutionary Tree: New Paradigms from Non-Model Animals / K. C. Sadler // *BioEssays.* — 2023. — Vol. 45, no. 1. — P. 2200036.
86. PIWI-interacting RNAs: Small RNAs with Big Functions / D. M. Ozata [et al.] // *Nat Rev Genet.* — 2019. — Feb. — Vol. 20, issue 2, no. 2. — P. 89–108.
87. Double-Stranded RNA-mediated Silencing of Genomic Tandem Repeats and Transposable Elements in the *D. Melanogaster* Germline / A. A. Aravin [et al.] // *Curr Biol.* — 2001. — Vol. 11, no. 13. — P. 1017–1027. — pmid: 11470406.
88. The Role of PIWI and the miRNA Machinery in *Drosophila* Germline Determination / H. B. Megosh [et al.] // *Curr Biol.* — 2006. — Vol. 16, no. 19. — P. 1884–1894. — pmid: 16949822.
89. PIWI Genes and piRNAs Are Ubiquitously Expressed in Mollusks and Show Patterns of Lineage-Specific Adaptation / J. Jehn [et al.] // *Commun Biol.* — 2018. — Vol. 1, issue 1, no. 1. — P. 1–11.
90. *Harrison, P. M.* The Ferritins: Molecular Properties, Iron Storage Function and Cellular Regulation / P. M. Harrison, P. Arosio // *Biochimica et Biophysica Acta (BBA) - Bioenergetics.* — 1996. — Т. 1275, № 3. — С. 161–203.

91. Structure, Function, and Evolution of Ferritins / S. C. Andrews [и др.] // *Journal of Inorganic Biochemistry*. — 1992. — Т. 47, № 1. — С. 161—174.
92. *Theil, E. C. Ferritin: Structure, Gene Regulation, and Cellular Function in Animals, Plants, and Microorganisms* / E. C. Theil // *Annu. Rev. Biochem.* — 1987. — Т. 56, № 1. — С. 289—315.
93. Ferritin and the Response to Oxidative Stress / К. Orino [и др.] // *Biochemical Journal*. — 2001. — Т. 357, № 1. — С. 241—247.
94. *Rucker, P. Role of H and L Subunits in Mouse Ferritin* / P. Rucker, F. M. Torti, S. V. Torti // *Journal of Biological Chemistry*. — 1996. — Vol. 271, no. 52. — P. 33352—33357. — pmid: 8969195.
95. Iron Redox Pathway Revealed in Ferritin via Electron Transfer Analysis / P. Chen [et al.] // *Sci Rep.* — 2020. — Vol. 10, issue 1, no. 1. — P. 4033.
96. *Yevenes, A. The Ferritin Superfamily* / A. Yevenes // *Macromolecular Protein Complexes: Structure and Function* / ed. by J. R. Harris, J. Marles-Wright. — Cham : Springer International Publishing, 2017. — P. 75—102. — (Subcellular Biochemistry).
97. Ferritin Light Chain Confers Protection Against Sepsis-Induced Inflammation and Organ Injury / A. Zarjou [и др.] // *Frontiers in Immunology*. — 2019. — Т. 10.
98. First Biochemical and Crystallographic Characterization of a Fast-Performing Ferritin from a Marine Invertebrate / E. De Meulenaere [и др.] // *Biochemical Journal*. — 2017. — Т. 474, № 24. — С. 4193—4206.
99. *Rawat, R. Evidence That Ferritin Is Associated with Light Production in the Mucus of the Marine Worm Chaetopterus* / R. Rawat, D. D. Deheyn // *Sci Rep.* — 2016. — Vol. 6, issue 1, no. 1. — P. 36854.
100. Crystallographic Characterization of Ferritin from *Sinonovacula constricta* / C. Su [и др.] // *Biochemical and Biophysical Research Communications*. — 2020. — Т. 524, № 1. — С. 217—223.
101. Cultivation of Primmorphs from the Marine Sponge *Suberites domuncula*: Morphogenetic Potential of Silicon and Iron / G. Le Pennec [и др.] // *Journal of Biotechnology*. — 2003. — Т. 100, № 2. — С. 93—108. — (Biotechnological Aspects of Marine Sponges).

102. *Guan, L. L.* Effect of Exogenous Siderophores on Iron Uptake Activity of Marine Bacteria under Iron-Limited Conditions / L. L. Guan, K. Kanoh, K. Kamino // *Applied and Environmental Microbiology*. — 2001. — Apr. — Т. 67, № 4. — С. 1710—1717.
103. Metaproteogenomic Analysis of a Community of Sponge Symbionts / M. Liu [et al.] // *ISME J*. — 2012. — Aug. — Vol. 6, issue 8, no. 8. — P. 1515—1525.
104. *Ilbert, M.* Insight into the Evolution of the Iron Oxidation Pathways / M. Ilbert, V. Bonnefoy // *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. — 2013. — Т. 1827, № 2. — С. 161—175. — (The Evolutionary Aspects of Bioenergetic Systems).
105. Evolution of the Acute Phase Response: Iron Release by Echinoderm (*Asterias Forbesi*) Coelomocytes, and Cloning of an Echinoderm Ferritin Molecule / G. Beck [и др.] // *Developmental & Comparative Immunology*. — 2002. — Т. 26, № 1. — С. 11—26.
106. *Neves, J. V.* Transferrin and Ferritin Response to Bacterial Infection: The Role of the Liver and Brain in Fish / J. V. Neves, J. M. Wilson, P. N. S. Rodrigues // *Developmental & Comparative Immunology*. — 2009. — Т. 33, № 7. — С. 848—857.
107. Iron Metabolism in the Social Amoeba *Dictyostelium Discoideum*: A Role for Ferric Chelate Reductases / B. Peracino [и др.] // *European Journal of Cell Biology*. — 2022. — Т. 101, № 3. — С. 151230.
108. Hexasomy 13q31.3q34 Due to Two Marker Chromosomes with Inverted Duplication in a Fetus with Increased Nuchal Translucency / A. Stembalska [et al.] // *Birth Defects Res A Clin Mol Teratol*. — 2015. — Apr. — Vol. 103, no. 4. — P. 255—259. — pmid: 25852029.
109. Frequency of Chromosomal Abnormalities in Products of Conception / T. M. A. Teles [et al.] // *Rev Bras Ginecol Obstet*. — 2017. — Mar. — Vol. 39, no. 3. — P. 110—114. — pmid: 28297732.
110. The Genomic Basis of Sporadic and Recurrent Pregnancy Loss: A Comprehensive in-Depth Analysis of 24,900 Miscarriages / J. Finley [et al.] // *Reprod Biomed Online*. — 2022. — July. — Vol. 45, no. 1. — P. 125—134. — pmid: 35523710.

111. *Hassold, T.* The Origin of Human Aneuploidy: Where We Have Been, Where We Are Going / T. Hassold, H. Hall, P. Hunt // *Hum Mol Genet.* — 2007. — Vol. 16 Spec No. 2. — R203—208. — pmid: 17911163.
112. Down Syndrome-A Narrative Review with a Focus on Anatomical Features / A. Arumugam [et al.] // *Clin Anat.* — 2016. — July. — Vol. 29, no. 5. — P. 568—577. — pmid: 26599319.
113. *Hong, D. S.* Cognitive and Neurological Aspects of Sex Chromosome Aneuploidies / D. S. Hong, A. L. Reiss // *Lancet Neurol.* — 2014. — Mar. — Vol. 13, no. 3. — P. 306—318. — pmid: 24556008.
114. Molecular Mapping of the Edwards Syndrome Phenotype to Two Noncontiguous Regions on Chromosome 18 / L. Boghosian-Sell [et al.] // *Am J Hum Genet.* — 1994. — Sept. — Vol. 55, no. 3. — P. 476—483. — pmid: 8079991.
115. *Laurent, A. P.* Gain of Chromosome 21 in Hematological Malignancies: Lessons from Studying Leukemia in Children with Down Syndrome / A. P. Laurent, R. S. Kotecha, S. Malinge // *Leukemia.* — 2020. — Aug. — Vol. 34, no. 8. — P. 1984—1999. — pmid: 32433508.
116. Domains of Genome-Wide Gene Expression Dysregulation in Down's Syndrome / A. Letourneau [et al.] // *Nature.* — 2014. — Apr. — Vol. 508, no. 7496. — P. 345—350.
117. Single-Cell Chromatin Accessibility Landscape of Human Umbilical Cord Blood in Trisomy 18 Syndrome / X. Qiu [et al.] // *Hum Genomics.* — 2021. — Vol. 15, no. 1. — P. 40. — pmid: 34193281.
118. The Genome-Wide Impact of Trisomy 21 on DNA Methylation and Its Implications for Hematopoiesis / I. S. Muskens [et al.] // *Nat Commun.* — 2021. — Vol. 12, no. 1. — P. 821. — pmid: 33547282.
119. DNA Methylation Changes in Down Syndrome Derived Neural iPSCs Uncover Co-Dysregulation of ZNF and HOX3 Families of Transcription Factors / L. Laan [et al.] // *Clin Epigenetics.* — 2020. — Vol. 12, no. 1. — P. 9. — pmid: 31915063.
120. *Zheng, H.* The Role of 3D Genome Organization in Development and Cell Differentiation / H. Zheng, W. Xie // *Nature Reviews Molecular Cell Biology.* — 2019. — Sept. — Vol. 20, issue 9, no. 9. — P. 535—550.

121. *Maass, P. G.* Interchromosomal Interactions: A Genomic Love Story of Kissing Chromosomes / P. G. Maass, A. R. Barutcu, J. L. Rinn // *J Cell Biol.* — 2019. — Vol. 218, no. 1. — P. 27–38.
122. Spatial Organization of Chromosome Territories in the Interphase Nucleus of Trisomy 21 Cells / S. Kemeny [et al.] // *Chromosoma.* — 2018. — June. — Vol. 127, no. 2. — P. 247–259. — pmid: 29238858.
123. Down-Syndrome-Induced Senescence Disrupts the Nuclear Architecture of Neural Progenitors / H. S. Meharena [et al.] // *Cell Stem Cell.* — 2022. — Vol. 29, no. 1. — 116–130.e7. — pmid: 34995493.
124. Single Chromosome Aneuploidy Induces Genome-Wide Perturbation of Nuclear Organization and Gene Expression / R. Braun [et al.] // *Neoplasia.* — 2019. — Apr. — Vol. 21, no. 4. — P. 401–412. — pmid: 30909073.
125. Genetic Dosage and Position Effect of Small Supernumerary Marker Chromosome (sSMC) in Human Sperm Nuclei in Infertile Male Patient / M. Olszewska [et al.] // *Sci Rep.* — 2015. — Vol. 5. — P. 17408. — pmid: 26616419.
126. Centromeric Association of Small Supernumerary Marker Chromosomes with Their Sister-Chromosomes Detected by Three Dimensional Molecular Cytogenetics / E. Klein [et al.] // *Mol Cytogenet.* — 2012. — Vol. 5. — P. 15. — pmid: 22413994.
127. Chromatin Organization by an Interplay of Loop Extrusion and Compartmental Segregation / J. Nuebler [et al.] // *PNAS.* — 2018. — Vol. 115, no. 29. — E6697–E6706. — pmid: 29967174.
128. Two Independent Modes of Chromatin Organization Revealed by Cohesin Removal / W. Schwarzer [и др.] // *Nature.* — 2017. — T. 551, № 7678. — C. 51–56. — pmid: 29094699.
129. Cohesin Loss Eliminates All Loop Domains / S. S. P. Rao [et al.] // *Cell.* — 2017. — Vol. 171, no. 2. — 305–320.e24. — pmid: 28985562.
130. A Mechanism of Cohesin-Dependent Loop Extrusion Organizes Zygotic Genome Architecture / J. Gassler [и др.] // *The EMBO Journal.* — 2017. — T. 36, № 24. — C. 3600–3618.
131. *Mirny, L. A.* Two Major Mechanisms of Chromosome Organization / L. A. Mirny, M. Imakaev, N. Abdennur // *Current Opinion in Cell Biology.* — 2019. — T. 58. — C. 142–152. — (Cell Nucleus).

132. *Marchal, C.* Control of DNA Replication Timing in the 3D Genome / C. Marchal, J. Sima, D. M. Gilbert // *Nat Rev Mol Cell Biol.* — 2019. — Dec. — Vol. 20, issue 12, no. 12. — P. 721—737.
133. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome / E. Lieberman-Aiden [et al.] // *Science.* — 2009. — Vol. 326, no. 5950. — P. 289—293.
134. Replication Timing Maintains the Global Epigenetic State in Human Cells / K. N. Klein [et al.] // *Science.* — 2021. — Vol. 372, no. 6540. — P. 371—378. — pmid: 33888635.
135. Cohesin Residency Determines Chromatin Loop Patterns / L. Costantino [et al.] // *eLife* / ed. by A. L. Marston, J. K. Tyler. — 2020. — Vol. 9. — e59889.
136. Cohesin-Dependent Globules and Heterochromatin Shape 3D Genome Architecture in *S. Pombe* / T. Mizuguchi [и др.] // *Nature.* — 2014. — Т. 516, № 7531. — С. 432—435. — pmid: 25307058.
137. Real-time imaging of DNA loop extrusion by condensin / M. Ganji [и др.] // *Science.* — 2018. — Т. 360, № 6384. — С. 102—105.
138. Invariant TAD Boundaries Constrain Cell-Type-Specific Looping Interactions between Promoters and Distal Elements around the CFTR Locus / E. M. Smith [et al.] // *The American Journal of Human Genetics.* — 2016. — Vol. 98, no. 1. — P. 185—201. — pmid: 26748519.
139. *Spielmann, M.* Structural Variation in the 3D Genome / M. Spielmann, D. G. Lupianez, S. Mundlos // *Nat Rev Genet.* — 2018. — July. — Vol. 19, issue 7, no. 7. — P. 453—467.
140. Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding / T.-H. S. Hsieh [et al.] // *Molecular Cell.* — 2020. — Vol. 78, no. 3. — 539—553.e8.
141. Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions / J. R. Dixon [et al.] // *Nature.* — 2012. — May. — Vol. 485, issue 7398, no. 7398. — P. 376—380.
142. Comparative Hi-C Reveals That CTCF Underlies Evolution of Chromosomal Domain Architecture / M. Vietri Rudan [et al.] // *Cell Reports.* — 2015. — Vol. 10, no. 8. — P. 1297—1309. — pmid: 25732821.

143. Single-Nucleus Hi-C Reveals Unique Chromatin Reorganization at Oocyte-to-Zygote Transition / I. M. Flyamer [et al.] // *Nature*. — 2017. — Vol. 544, no. 7648. — P. 110—114.
144. Nanoscale 3D DNA Tracing in Single Human Cells Visualizes Loop Extrusion Directly in Situ / K. S. Beckwith [et al.]. — 2023. — preprint.
145. Dynamics of CTCF- and Cohesin-Mediated Chromatin Looping Revealed by Live-Cell Imaging / M. Gabriele [и др.] // *Science*. — 2022. — Т. 376, № 6592. — С. 496—501.
146. Order and Stochasticity in the Folding of Individual *Drosophila* Genomes / S. V. Ulianov [et al.] // *Nat Commun*. — 2021. — Vol. 12, issue 1, no. 1. — P. 1—17.
147. Form and function of topologically associating genomic domains in budding yeast / U. Eser [et al.] // *Proceedings of the National Academy of Sciences*. — 2017. — Vol. 114, no. 15. — E3061—E3070.
148. Sub-nucleosomal Genome Structure Reveals Distinct Nucleosome Folding Motifs / M. Ohno [et al.] // *Cell*. — 2019. — Vol. 176, no. 3. — 520—534.e25.
149. Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C / T.-H. S. Hsieh [и др.] // *Cell*. — 2015. — Т. 162, № 1. — С. 108—119.
150. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping / S. S. P. Rao [et al.] // *Cell*. — 2014. — Vol. 159, no. 7. — P. 1665—1680.
151. ATP-dependent DNA Binding, Unwinding, and Resection by the Mre11/Rad50 Complex / Y. Liu [et al.] // *EMBO J*. — 2016. — Vol. 35, no. 7. — P. 743—758. — pmid: 26717941.
152. Hansen, A. S. CTCF as a Boundary Factor for Cohesin-Mediated Loop Extrusion: Evidence for a Multi-Step Mechanism / A. S. Hansen // *Nucleus*. — 2020. — Т. 11, № 1. — С. 132—148. — pmid: 32631111.
153. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization / E. P. Nora [et al.] // *Cell*. — 2017. — Vol. 169, no. 5. — 930—944.e22.

154. Topologically Associating Domains and Chromatin Loops Depend on Cohesin and Are Regulated by CTCF, WAPL, and PDS5 Proteins / G. Wutz [и др.] // *The EMBO Journal*. — 2017. — Т. 36, № 24. — С. 3573—3599.
155. Wapl Is an Essential Regulator of Chromatin Structure and Chromosome Segregation / A. Tedeschi [et al.] // *Nature*. — 2013. — Sept. — Vol. 501, issue 7468, no. 7468. — P. 564—568.
156. Rapid Depletion of CTCF and Cohesin Proteins Reveals Dynamic Features of Chromosome Architecture / N. Q. Liu [et al.]. — 2021. — preprint.
157. Chromatin Jets Define the Properties of Cohesin-Driven in Vivo Loop Extrusion / Y. Guo [et al.] // *Molecular Cell*. — 2022. — Vol. 82, no. 20. — 3769—3780.e5.
158. Cohesin Forms Fountains at Active Enhancers in *C. Elegans* / B. Isiaka [et al.]. — 2023. — preprint.
159. Extrusion Fountains Are Hallmarks of Chromosome Organization Emerging upon Zygotic Genome Activation / A. Galitsyna [et al.]. — 2023. — preprint.
160. Chromatin Architecture Transitions from Zebrafish Sperm through Early Embryogenesis / C. L. Wike [и др.] // *Genome Res*. — 2021. — ИЮНЬ. — Т. 31, № 6. — С. 981—994. — pmid: 34006569.
161. *Krefting, J.* Evolutionary Stability of Topologically Associating Domains Is Associated with Conserved Gene Regulation / J. Krefting, M. A. Andrade-Navarro, J. Ibn-Salem // *BMC Biology*. — 2018. — Т. 16, № 1. — С. 87.
162. *Kim, E.* Looping the Genome with SMC Complexes / E. Kim, R. Barth, C. Dekker // *Annual Review of Biochemistry*. — 2023. — Т. 92, № 1. — С. 15—41. — pmid: 37137166.
163. Highly Rearranged Chromosomes Reveal Uncoupling between Genome Topology and Gene Expression / Y. Ghavi-Helm [et al.] // *Nat Genet*. — 2019. — Aug. — Vol. 51, issue 8, no. 8. — P. 1272—1282.
164. Developmentally Regulated Shh Expression Is Robust to TAD Perturbations / I. Williamson [и др.] // *Development*. — 2019. — Т. 146, № 19. — dev179523.

165. Live-Cell Imaging Reveals Enhancer-Dependent Sox2 Transcription in the Absence of Enhancer Proximity / J. M. Alexander [и др.] // *eLife* / под ред. R. H. Singer, K. Struhl, Z. Liu. — 2019. — Т. 8. — e41769.
166. Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation / N. S. Benabdallah [и др.] // *Molecular Cell*. — 2019. — Т. 76, № 3. — 473—484.e7.
167. Building Regulatory Landscapes Reveals That an Enhancer Can Recruit Cohesin to Create Contact Domains, Engage CTCF Sites and Activate Distant Genes / N. J. Rinzema [et al.] // *Nat Struct Mol Biol*. — 2022. — June. — Vol. 29, issue 6, no. 6. — P. 563—574.
168. *Horsfield, J. A.* Full Circle: A Brief History of Cohesin and the Regulation of Gene Expression / J. A. Horsfield // *The FEBS Journal*. — 2023. — Vol. 290, no. 7. — P. 1670—1687.
169. DNA Loop Extrusion by Human Cohesin / I. F. Davidson [и др.] // *Science*. — 2019. — Т. 366, № 6471. — С. 1338—1345.
170. *Hirano, T.* Condensin-Based Chromosome Organization from Bacteria to Vertebrates / T. Hirano // *Cell*. — 2016. — Vol. 164, no. 5. — P. 847—857. — pmid: 26919425.
171. Organization of the Mitotic Chromosome / N. Naumova [и др.] // *Science*. — 2013. — Т. 342, № 6161. — С. 948—953.
172. A pathway for mitotic chromosome formation / J. H. Gibcus [и др.] // *Science*. — 2018. — Т. 359, № 6376. — eaao6135.
173. Cohesin Loss Eliminates All Loop Domains / S. S. P. Rao [и др.] // *Cell*. — 2017. — Т. 171, № 2. — 305—320.e24.
174. DNA-loop Extruding Condensin Complexes Can Traverse One Another / E. Kim [et al.] // *Nature*. — 2020. — Mar. — Vol. 579, issue 7799, no. 7799. — P. 438—442.
175. Cohesin Mediates DNA Loop Extrusion and Sister Chromatid Cohesion by Distinct Mechanisms / K. Nagasaka [et al.]. — 2022. — preprint.
176. *Liu, Y.* CTCF–CTCF Loops and Intra-TAD Interactions Show Differential Dependence on Cohesin Ring Integrity / Y. Liu, J. Dekker // *Nat Cell Biol*. — 2022. — Oct. — Vol. 24, issue 10, no. 10. — P. 1516—1527.

177. The Structural Basis for Cohesin–CTCF-anchored Loops / Y. Li [et al.] // Nature. — 2020. — Feb. — Vol. 578, issue 7795, no. 7795. — P. 472–476.
178. SMC Motor Proteins Extrude DNA Asymmetrically and Contain a Direction Switch / R. Barth [et al.]. — 2023. — preprint.
179. *Banigan, E. J.* Loop Extrusion: Theory Meets Single-Molecule Experiments / E. J. Banigan, L. A. Mirny // Current Opinion in Cell Biology. — 2020. — T. 64. — C. 124–138. — (Cell Nucleus).
180. Live-Cell Imaging Reveals a Stable Cohesin-Chromatin Interaction after but Not before DNA Replication / D. Gerlich [et al.] // Current Biology. — 2006. — Vol. 16, no. 15. — P. 1571–1578. — pmid: 16890534.
181. Recent Evidence That TADs and Chromatin Loops Are Dynamic Structures / A. S. Hansen [и др.] // Nucleus. — 2018. — Т. 9, № 1. — С. 20–32. — pmid: 29077530.
182. CTCF Is a DNA-tension-dependent Barrier to Cohesin-Mediated Loop Extrusion / I. F. Davidson [et al.] // Nature. — 2023. — Apr. — Vol. 616, issue 7958, no. 7958. — P. 822–827.
183. The Chromatin Insulator CTCF and the Emergence of Metazoan Diversity / P. Heger [и др.] // Proceedings of the National Academy of Sciences. — 2012. — Т. 109, № 43. — С. 17507–17512.
184. Cohesin-Dependent Chromosome Loop Extrusion Is Limited by Transcription and Stalled Replication Forks / K. Jeppsson [и др.] // Science Advances. — 2022. — Т. 8, № 23. — eabn7063.
185. Sister Chromatid Cohesion Halts DNA Loop Expansion / N. Bastie [et al.] // Molecular Cell. — 2024. — Vol. 84, no. 6. — P. 1139–1148. — pmid: 38452765.
186. RNA Polymerases as Moving Barriers to Condensin Loop Extrusion / H. B. Brandao [и др.] // Proceedings of the National Academy of Sciences. — 2019. — Т. 116, № 41. — С. 20489–20499.
187. Transcription Shapes 3D Chromatin Organization by Interacting with Loop Extrusion / E. J. Banigan [et al.] // Proceedings of the National Academy of Sciences. — 2023. — Vol. 120, no. 11. — e2210480120.

188. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping / S. S. P. Rao [и др.] // *Cell*. — 2014. — Т. 159, № 7. — С. 1665—1680.
189. *Yaffe, E.* Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture / E. Yaffe, A. Tanay // *Nature Genetics*. — 2011. — Nov. — Vol. 43, no. 11. — P. 1059—1065.
190. *Belaghzal, H.* Hi-C 2.0: An Optimized Hi-C Procedure for High-Resolution Genome-Wide Mapping of Chromosome Conformation / H. Belaghzal, J. Dekker, J. H. Gibcus // *Methods*. — 2017. — Vol. 123. — P. 56—65. — pmid: 28435001.
191. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome / E. Lieberman-Aiden [et al.] // *Science*. — 2009. — Vol. 326, no. 5950. — P. 289—293. — pmid: 19815776.
192. Hi-C 3.0: Improved Protocol for Genome-Wide Chromosome Conformation Capture / D. L. Lafontaine [et al.] // *Curr Protoc*. — 2021. — July. — Vol. 1, no. 7. — e198. — pmid: 34286910.
193. Micro-C XL: assaying chromosome conformation from the nucleosome to the entire genome / T.-H. S. Hsieh [et al.] // *Nature Methods*. — 2016. — Dec. — Vol. 13, no. 12. — P. 1009—1011.
194. *Belaghzal, H.* Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation / H. Belaghzal, J. Dekker, J. H. Gibcus // *Methods (San Diego, Calif.)* — 2017. — Т. 123. — С. 56—65.
195. Using DNase Hi-C techniques to map global and local three-dimensional genome architecture at high resolution / W. Ma [и др.] // *Methods (San Diego, Calif.)* — 2018. — Т. 142. — С. 59—73.
196. Systematic Evaluation of Chromosome Conformation Capture Assays / B. Akgol Oksuz [et al.] // *Nat Methods*. — 2021. — Sept. — Vol. 18, issue 9, no. 9. — P. 1046—1055.
197. Comparison of computational methods for Hi-C data analysis / M. Forcato [et al.] // *Nature Methods*. — 2017. — July. — Vol. 14, no. 7. — P. 679—685.

198. Ultrastructural Details of Mammalian Chromosome Architecture / N. Krietenstein [et al.] // *Molecular Cell*. — 2020. — Vol. 78, no. 3. — 554—565.e7. — pmid: 32213324.
199. *Schmitt, A. D.* Genome-wide mapping and analysis of chromosome architecture / A. D. Schmitt, M. Hu, B. Ren // *Nature Reviews Molecular Cell Biology*. — 2016. — Dec. — Vol. 17, no. 12. — P. 743—755.
200. Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries / Y. Gong [et al.] // *Nature Communications*. — 2018. — Vol. 9, no. 1. — P. 542.
201. HiFive: a tool suite for easy and efficient HiC and 5C data analysis / M. E. Sauria [и др.] // *Genome Biology*. — 2015. — Т. 16, № 1. — С. 237.
202. HIPPIE: a high-throughput identification pipeline for promoter interacting enhancer elements / Y.-C. Hwang [и др.] // *Bioinformatics*. — 2015. — Т. 31, № 8. — С. 1290—1292.
203. Binless normalization of Hi-C data provides significant interaction and difference detection independent of resolution / Y. G. Spill [et al.] // *Nature Communications*. — 2019. — Vol. 10, no. 1. — P. 1938.
204. HiCNorm: removing biases in Hi-C data via Poisson regression / M. Hu [и др.] // *Bioinformatics*. — 2012. — Сент. — Т. 28, № 23. — С. 3131—3133. — eprint: <https://academic.oup.com/bioinformatics/article-pdf/28/23/3131/18529988/bts570.pdf>.
205. *Schmid, M. W.* HiCdat: a fast and easy-to-use Hi-C data analysis tool / M. W. Schmid, S. Grob, U. Grossniklaus // *BMC Bioinformatics*. — 2015. — Т. 16, № 1. — С. 277.
206. *Lajoie, B. R.* The Hitchhiker's guide to Hi-C analysis: practical guidelines / B. R. Lajoie, J. Dekker, N. Kaplan // *Methods (San Diego, Calif.)* — 2015. — Т. 72. — С. 65—75.
207. Iterative correction of Hi-C data reveals hallmarks of chromosome organization / M. Imakaev [et al.] // *Nature Methods*. — 2012. — Oct. — Vol. 9, no. 10. — P. 999—1003.
208. Cumulative Contact Frequency of a Chromatin Region Is an Intrinsic Property Linked to Its Function / M. D. Samborskaia [et al.] // *PeerJ*. — 2020. — Vol. 8. — e9566. — pmid: 32864204.

209. *Love, M. I.* Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2 / M. I. Love, W. Huber, S. Anders // *Genome Biology*. — 2014. — Т. 15, № 12. — С. 550.
210. *Wu, H.-J.* A computational strategy to adjust for copy number in tumor Hi-C data / H.-J. Wu, F. Michor // *Bioinformatics*. — 2016. — Т. 32, № 24. — С. 3695—3701.
211. OneD: increasing reproducibility of Hi-C samples with abnormal karyotypes / E. Vidal [и др.] // *Nucleic Acids Research*. — 2018. — Т. 46, № 8. — e49.
212. Robust Hi-C Maps of Enhancer-Promoter Interactions Reveal the Function of Non-coding Genome in Neural Development and Diseases / L. Lu [и др.] // *Molecular Cell*. — 2020. — Т. 79, № 3. — 521—534.e15.
213. A high-resolution map of the three-dimensional chromatin interactome in human cells / F. Jin [et al.] // *Nature*. — 2013. — Nov. — Vol. 503, no. 7475. — P. 290—294.
214. Topological domains in mammalian genomes identified by analysis of chromatin interactions / J. R. Dixon [et al.] // *Nature*. — 2012. — May. — Vol. 485, no. 7398. — P. 376—380.
215. HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient / T. Yang [и др.] // *Genome Research*. — 2017. — Ноябрь. — Т. 27, № 11. — С. 1939—1949.
216. HiC-spector: a matrix library for spectral and reproducibility analysis of Hi-C contact maps / K.-K. Yan [и др.] // *Bioinformatics*. — 2017. — Т. 33, № 14. — С. 2199—2201.
217. GenomeDISCO: a concordance score for chromosome conformation capture experiments using random walks on contact map graphs / O. Ursu [и др.] // *Bioinformatics*. — 2018. — Т. 34, № 16. — С. 2701—2707.
218. QuASAR: quantitative allele-specific analysis of reads / C. T. Harvey [и др.] // *Bioinformatics*. — 2015. — Т. 31, № 8. — С. 1235—1242.
219. *Ay, F.* Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts / F. Ay, T. L. Bailey, W. S. Noble // *Genome Research*. — 2014. — Vol. 24, no. 6. — P. 999—1011.

220. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C / B. Mifsud [et al.] // *Nature Genetics*. — 2015. — June. — Vol. 47, no. 6. — P. 598–606.
221. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities / S. Heinz [et al.] // *Molecular Cell*. — 2010. — Vol. 38, no. 4. — P. 576–589.
222. *Lun, A. T.* diffHic: a Bioconductor package to detect differential genomic interactions in Hi-C data / A. T. Lun, G. K. Smyth // *BMC Bioinformatics*. — 2015. — T. 16. — C. 258.
223. Analysis of Hi-C data using SIP effectively identifies loops in organisms from *C. elegans* to mammals / M. J. Rowley [et al.] // *Genome Research*. — 2020. — gr.257832.119.
224. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments / N. C. Durand [и др.] // *Cell Systems*. — 2016. — Т. 3, № 1. — С. 95–98.
225. Computational Processing and Quality Control of Hi-C, Capture Hi-C and Capture-C Data / P. Hansen [и др.] // *Genes*. — 2019. — Т. 10, № 7. — С. 548.
226. *Kim, K.* covNorm: An R Package for Coverage Based Normalization of Hi-C and Capture Hi-C Data / K. Kim, I. Jung // *Computational and Structural Biotechnology Journal*. — 2021. — Vol. 19. — P. 3149–3159. — pmid: 34141136.
227. Chromosome conformation elucidates regulatory relationships in developing human brain / H. Won [et al.] // *Nature*. — 2016. — Oct. — Vol. 538, no. 7626. — P. 523–527.
228. LASCA: loop and significant contact annotation pipeline / A. V. Luzhin [et al.] // *Scientific Reports*. — 2021. — Vol. 11, no. 1. — P. 6361.
229. *Bicciato, S.* Hi-C Data Analysis. Vol. 2301 / S. Bicciato, F. Francesco. — Humana New York, NY, 2022. — XIII, 354. — (Methods in Molecular Biology).
230. HiGlass: Web-Based Visual Exploration and Analysis of Genome Interaction Maps / P. Kerpedjiev [et al.] // *Genome Biology*. — 2018. — Vol. 19, no. 1. — P. 125.

231. Galaxy HiCExplorer 3: a web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization / J. Wolff [et al.] // *Nucleic Acids Research*. —
232. *Ramirez, F.* HiCBrowser : A Simple Web Browser to Visualize Hi-C and Other Genomic Tracks / F. Ramirez. — The deepTools ecosystem, 03.07.2023.
233. open2c/cooltools: v0.4.1 / S. Venev [и др.]. — 2021.
234. *Mirny, L. A.* The Fractal Globule as a Model of Chromatin Architecture in the Cell / L. A. Mirny // *Chromosome Res.* — 2011. — Vol. 19, no. 1. — P. 37—51.
235. Condensin-driven remodelling of X chromosome topology during dosage compensation / E. Crane [et al.] // *Nature*. — 2015. — July. — Vol. 523, no. 7559. — P. 240—244.
236. The Long Noncoding RNA Transcriptome of *Dictyostelium discoideum* Development / R. D. Rosengarten [et al.] // *G3: Genes, Genomes, Genetics*. — 2017. — Vol. 7, no. 2. — P. 387—398.
237. Liquid Chromatin Hi-C Characterizes Compartment-Dependent Chromatin Interaction Dynamics / H. Belaghzal [и др.] // *Nature Genetics*. — 2021. — T. 53. — C. 1—12.
238. LASCA: Loop and Significant Contact Annotation Pipeline / A. V. Luzhin [et al.] // *Sci Rep.* — 2021. — Vol. 11, issue 1, no. 1. — P. 6361.
239. Computer vision for pattern detection in chromosome contact maps / C. Matthey-Doret [et al.] // *Nature Communications*. — 2020. — Vol. 11, no. 1. — P. 1—11.
240. Computer Vision for Pattern Detection in Chromosome Contact Maps / C. Matthey-Doret [et al.] // *Nat Commun.* — 2020. — Vol. 11, issue 1, no. 1. — P. 1—11.
241. Leaps and lulls in the developmental transcriptome of *Dictyostelium discoideum* / R. D. Rosengarten [и др.] // *BMC Genomics*. — 2015. — T. 16, № 1. — C. 294.

242. *Eisenberg, E.* Human housekeeping genes, revisited / E. Eisenberg, E. Y. Levanon // Trends in genetics: TIG. — 2013. — Октябрь. — Т. 29, № 10. — С. 569—574.
243. Role of epigenetics in unicellular to multicellular transition in *Dictyostelium* / S. Y. Wang [et al.] // Genome Biology. — 2021. — Vol. 22, no. 1. — P. 134.
244. Role of Epigenetics in Unicellular to Multicellular Transition in *Dictyostelium* / S. Y. Wang [и др.] // Genome Biology. — 2021. — Т. 22, № 1. — С. 134.
245. The Putative "Nucleation Site" in Human H-chain Ferritin Is Not Required for Mineralization of the Iron Core / F. Bou-Abdallah [et al.] // Biochemistry. — 2004. — Vol. 43, no. 14. — P. 4332—4337. — pmid: 15065877.
246. *Behera, R. K.* Moving Fe²⁺ from Ferritin Ion Channels to Catalytic OH Centers Depends on Conserved Protein Cage Carboxylates / R. K. Behera, E. C. Theil // Proceedings of the National Academy of Sciences. — 2014. — Т. 111, № 22. — С. 7925—7930.
247. SIREs: Searching for Iron-Responsive Elements / M. Campillos [et al.] // Nucleic Acids Res. — 2010. — July. — Vol. 38, Web Server issue. — W360—367. — pmid: 20460462.
248. *Eisenberg, E.* Human Housekeeping Genes, Revisited / E. Eisenberg, E. Y. Levanon // Trends Genet. — 2013. — Октябрь. — Vol. 29, no. 10. — P. 569—574. — pmid: 23810203.
249. *Bonev, B.* Organization and function of the 3D genome / B. Bonev, G. Cavalli // Nature Reviews Genetics. — 2016. — Ноябрь. — Vol. 17, no. 11. — P. 661—678.
250. Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in *Drosophila* / J. Brennecke [et al.] // Cell. — 2007. — Vol. 128, no. 6. — P. 1089—1103.
251. Studying RNA–DNA Interactome by Red-C Identifies Noncoding RNAs Associated with Various Chromatin Types and Reveals Transcription Dynamics / A. A. Gavrilov [и др.] // Nucleic Acids Research. — 2020. — Т. 48, № 12. — С. 6699—6714.

252. Comparative Genomics of the Social Amoebae *Dictyostelium Discoideum* and *Dictyostelium Purpureum* / R. Sugang [et al.] // *Genome Biology*. — 2011. — Vol. 12, no. 2. — R20.
253. *Flyamer, I. M.* Coolpup.Py: Versatile Pile-up Analysis of Hi-C Data / I. M. Flyamer, R. S. Illingworth, W. A. Bickmore // *Bioinformatics*. — 2020. — Vol. 36, no. 10. — P. 2980—2985.
254. deepTools2: A next Generation Web Server for Deep-Sequencing Data Analysis / F. Ramirez [et al.] // *Nucleic Acids Res.* — 2016. — Vol. 44, W1. — W160—W165.
255. High-Resolution Mapping of the Spatial Organization of a Bacterial Chromosome / T. B. Le [и др.] // *Science*. — 2013. — Т. 342, № 6159. — С. 731—734. — pmid: 24158908.
256. Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C / T.-H. S. Hsieh [et al.] // *Cell*. — 2015. — Vol. 162, no. 1. — P. 108—119. — pmid: 26119342.
257. *Takemata, N.* Multi-Scale Architecture of Archaeal Chromosomes / N. Takemata, S. D. Bell // *Molecular Cell*. — 2021. — Т. 81, № 3. — 473—487.e6.
258. Formation of Chromosomal Domains by Loop Extrusion / G. Fudenberg [et al.] // *Cell Rep.* — 2016. — Vol. 15, no. 9. — P. 2038—2049. — pmid: 27210764.
259. Enhancer Evolution across 20 Mammalian Species / D. Villar [et al.] // *Cell*. — 2015. — Vol. 160, no. 3. — P. 554—566. — pmid: 25635462.
260. *Small, S.* Transcriptional Enhancers in *Drosophila* / S. Small, D. N. Arnosti // *Genetics*. — 2020. — Т. 216, № 1. — С. 1—26. — pmid: 32878914.
261. Parental-to-Embryo Switch of Chromosome Organization in Early Embryogenesis / S. Collombet [et al.] // *Nature*. — 2020. — Apr. — Vol. 580, issue 7801, no. 7801. — P. 142—146.
262. *Eckersley-Maslin, M. A.* Dynamics of the Epigenetic Landscape during the Maternal-to-Zygotic Transition / M. A. Eckersley-Maslin, C. Alda-Catalinas, W. Reik // *Nature Reviews Molecular Cell Biology*. — 2018. — July. — Vol. 19, issue 7, no. 7. — P. 436—450.

263. *Tadros, W.* The Maternal-to-Zygotic Transition: A Play in Two Acts / W. Tadros, H. D. Lipshitz // *Development*. — 2009. — Vol. 136, no. 18. — P. 3033–3042. — pmid: 19700615.
264. Dynamic Epigenomic Landscapes during Early Lineage Specification in Mouse Embryos / Y. Zhang [et al.] // *Nature Genetics*. — 2018. — Jan. — Vol. 50, no. 1. — P. 96–105.
265. Enhancer–Promoter Interactions and Transcription Are Largely Maintained upon Acute Loss of CTCF, Cohesin, WAPL or YY1 / T.-H. S. Hsieh [et al.] // *Nat Genet*. — 2022. — Dec. — Vol. 54, issue 12, no. 12. — P. 1919–1932.
266. A Compendium of Chromatin Contact Maps Reveal Spatially Active Regions in the Human Genome / A. D. Schmitt [и др.] // *Cell reports*. — 2016. — Т. 17, № 8. — С. 2042–2059. — pmid: 27851967.
267. *Thul, P. J.* The Human Protein Atlas: A Spatial Map of the Human Proteome / P. J. Thul, C. Lindskog // *Protein Sci*. — 2018. — Vol. 27, no. 1. — P. 233–244. — pmid: 28940711.
268. *Cremer, T.* Chromosome Territories, Nuclear Architecture and Gene Regulation in Mammalian Cells / T. Cremer, C. Cremer // *Nat Rev Genet*. — 2001. — Apr. — Vol. 2, issue 4, no. 4. — P. 292–301.
269. *Stack, S. M.* Chromosome Polarization and Nuclear Rotation in *Allium cepa* Roots / S. M. Stack, C. R. Clarke // *Cytologia*. — 1974. — Т. 39, № 3. — С. 553–560.
270. Cell cycle-dependent specific positioning and clustering of centromeres and telomeres in fission yeast. / H. Funabiki [и др.] // *Journal of Cell Biology*. — 1993. — Т. 121, № 5. — С. 961–976.
271. *Agard, D. A.* Three-Dimensional Architecture of a Polytene Nucleus / D. A. Agard, J. W. Sedat // *Nature*. — 1983. — Apr. — Vol. 302, issue 5910, no. 5910. — P. 676–681.
272. Yeast Nuclei Display Prominent Centromere Clustering That Is Reduced in Nondividing Cells and in Meiotic Prophase / Q.-w. Jin [и др.] // *Journal of Cell Biology*. — 1998. — Т. 141, № 1. — С. 21–29.

273. 3D Genomics across the Tree of Life Reveals Condensin II as a Determinant of Architecture Type / C. Hoencamp [et al.] // *Science*. — 2021. — Vol. 372, no. 6545. — P. 984–989.
274. Chromosome-Level Organization of the Regulatory Genome in the *Drosophila* Nervous System / G. Mohana [et al.] // *Cell*. — 2023. — Vol. 186, no. 18. — 3826–3844.e26. — pmid: 37536338.
275. Massively Multiplex Single-Cell Hi-C / V. Ramani [et al.] // *Nat Methods*. — 2017. — Mar. — Vol. 14, issue 3, no. 3. — P. 263–266.
276. *Grob, S.* Hi-C Analysis in *Arabidopsis* Identifies the KNOT, a Structure with Similarities to the flamenco Locus of *Drosophila* / S. Grob, M. W. Schmid, U. Grossniklaus // *Molecular Cell*. — 2014. — T. 55, № 5. — C. 678–693.
277. *Grob, S.* Invasive DNA elements modify the nuclear architecture of their insertion site by KNOT-linked silencing in *Arabidopsis thaliana* / S. Grob, U. Grossniklaus // *Genome Biology*. — 2019. — T. 20. — C. 120.
278. Genome-Wide Hi-C Analyses in Wild-Type and Mutants Reveal High-Resolution Chromatin Interactions in *Arabidopsis* / S. Feng [et al.] // *Molecular Cell*. — 2014. — Vol. 55, no. 5. — P. 694–707. — pmid: 25132175.
279. *Grob, S.* Hi-C Analysis in *Arabidopsis* Identifies the KNOT, a Structure with Similarities to the Flamenco Locus of *Drosophila* / S. Grob, M. W. Schmid, U. Grossniklaus // *Molecular Cell*. — 2014. — Vol. 55, no. 5. — P. 678–693. — pmid: 25132176.
280. 3D Chromatin Architecture of Large Plant Genomes Determined by Local A/B Compartments / P. Dong [et al.] // *Mol Plant*. — 2017. — Vol. 10, no. 12. — P. 1497–1509. — pmid: 29175436.
281. Prominent Topologically Associated Domains Differentiate Global Chromatin Packing in Rice from *Arabidopsis* / C. Liu [et al.] // *Nat Plants*. — 2017. — Sept. — Vol. 3, no. 9. — P. 742–748. — pmid: 28848243.
282. Evolutionarily Conserved Principles Predict 3D Chromatin Organization / M. J. Rowley [et al.] // *Mol Cell*. — 2017. — Vol. 67, no. 5. — 837–852.e7. — pmid: 28826674.

283. Systemic Loss and Gain of Chromatin Architecture throughout Zebrafish Development / L. J. Kaaij [et al.] // *Cell Rep.* — 2018. — Vol. 24, no. 1. — 1—10.e4. — pmid: 29972771.
284. Principles of 3D Chromosome Folding and Evolutionary Genome Reshuffling in Mammals / L. Alvarez-Gonzalez [et al.] // *Cell Rep.* — 2022. — Vol. 41, no. 12. — P. 111839. — pmid: 36543130.
285. *Chakraborty, A.* The Role of 3D Genome Organization in Disease: From Compartments to Single Nucleotides / A. Chakraborty, F. Ay // *Semin Cell Dev Biol.* — 2019. — Vol. 90. — P. 104—113. — pmid: 30017907.
286. *Razin, S. V.* Genome-Directed Cell Nucleus Assembly / S. V. Razin, S. V. Ulianov // *Biology.* — 2022. — May. — Vol. 11, issue 5, no. 5. — P. 708.
287. Single-Cell Dynamics of Genome-Nuclear Lamina Interactions / J. Kind [et al.] // *Cell.* — 2013. — Vol. 153, no. 1. — P. 178—192. — pmid: 23523135.
288. Promoter-Intrinsic and Local Chromatin Features Determine Gene Repression in LADs / C. Leemans [et al.] // *Cell.* — 2019. — Vol. 177, no. 4. — 852—864.e14. — pmid: 30982597.
289. Emerging Evidence of Chromosome Folding by Loop Extrusion / G. Fudenberg [et al.] // *Cold Spring Harbor Symposia on Quantitative Biology.* — 2017. — Vol. 82. — P. 45—55.
290. HOTTIP-dependent R-loop Formation Regulates CTCF Boundary Activity and TAD Integrity in Leukemia / H. Luo [et al.] // *Molecular Cell.* — 2022. — Vol. 82, no. 4. — 833—851.e11. — pmid: 35180428.
291. Jpx RNA Regulates CTCF Anchor Site Selection and Formation of Chromosome Loops / H. J. Oh [et al.] // *Cell.* — 2021. — Vol. 184, no. 25. — 6157—6173.e24. — pmid: 34856126.
292. RNA polymerases as moving barriers to condensin loop extrusion / H. B. Brandao [и др.] // *Proceedings of the National Academy of Sciences.* — 2019. — Т. 116, № 41. — С. 20489—20499.
293. MCM complexes are barriers that restrict cohesin-mediated loop extrusion / B. J. H. Dequeker [et al.] // *Nature.* — 2022. — June. — Vol. 606, no. 7912. — P. 197—203.

294. Cohesin residency determines chromatin loop patterns / L. Costantino [и др.] // *eLife* / под ред. A. L. Marston, J. K. Tyler, A. L. Marston. — 2020. — Т. 9. — e59889.
295. Transcription-dependent domain-scale three-dimensional genome organization in the dinoflagellate *Breviolum minutum* / G. K. Marinov [et al.] // *Nature Genetics*. — 2021. — May. — Vol. 53, no. 5. — P. 613–617.
296. Genetic and Spatial Organization of the Unusual Chromosomes of the Dinoflagellate *Symbiodinium Microadriaticum* / A. Nand [et al.] // *Nat Genet*. — 2021. — May. — Vol. 53, issue 5, no. 5. — P. 618–629.
297. Cohesin relocation from sites of chromosomal loading to places of convergent transcription / A. Lengronne [et al.] // *Nature*. — 2004. — July. — Vol. 430, no. 6999. — P. 573–578.
298. Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl / G. A. Busslinger [et al.] // *Nature*. — 2017. — Apr. — Vol. 544, no. 7651. — P. 503–507.
299. Negative Supercoil at Gene Boundaries Modulates Gene Topology / Y. J. Achar [et al.] // *Nature*. — 2020. — Jan. — Vol. 577, no. 7792. — P. 701–705.
300. Transcription-Mediated Supercoiling Regulates Genome Folding and Loop Formation / M. V. Neguembor [et al.] // *Mol Cell*. — 2021. — Vol. 81, no. 15. — 3065–3081.e12. — pmid: 34297911.
301. Transcription-Dependent Domain-Scale Three-Dimensional Genome Organization in the Dinoflagellate *Breviolum Minutum* / G. K. Marinov [et al.] // *Nat Genet*. — 2021. — May. — Vol. 53, issue 5, no. 5. — P. 613–617.