

# Пространственная организация хроматина

Александра Галицына  
аспирант Сколтеха (лаба М.С. Гельфанда)  
ФББ МГУ выпуск 2017

[agalitzina@gmail.com](mailto:agalitzina@gmail.com)

# План лекций

## Лекция 1:

- Введение в пространственную организацию хроматина
  - Общие принципы пространственной организации хроматина и подходы к ее исследованию: микроскопия и фиксация конформации хромосом (3C).
  - Краткая история развития экспериментальных методов, основанных на 3C и секвенировании (Hi-C, 4C, 5C).
  - Уровни организации хроматина по Hi-C
- От теории к практике:
  - Особенности картирования данных ДНК-ДНК взаимодействий
  - Фильтрация данных Hi-C, бинирование, получение контактных карт
  - Итеративная коррекция

## Лекция 2:

- Поиск особенностей карт контактов (ТАДы, компартменты, обогащенные взаимодействия).
- Общепринятые пакеты для обработки, форматы хранения, базы данных и браузеры Hi-C экспериментов.
- Подходы к интеграции данных Hi-C с другими геномными разметками: браузер HiGlass, построение усредненных профилей и усредненных карт.
- Специальные случаи
  - Применение Hi-C для сборки геномов
  - Single-cell Hi-C
  - РНК-ДНК взаимодействия

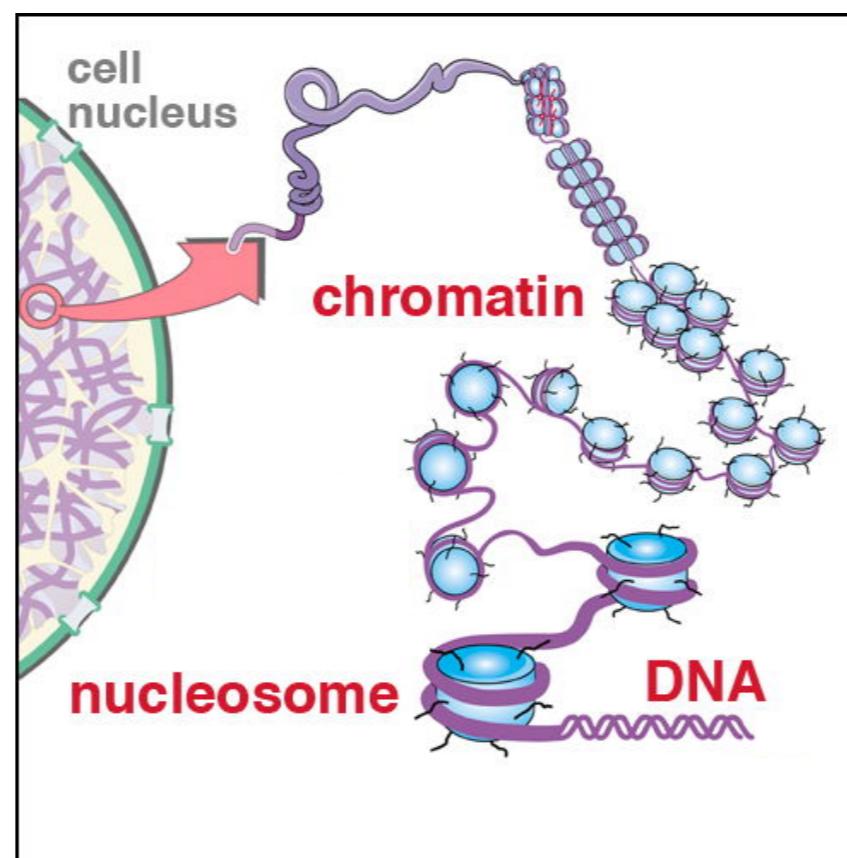
# Введение

- Общие принципы пространственной организации хроматина и подходы к ее исследованию: микроскопия и фиксация конформации хромосом (3C).
- Краткая история развития экспериментальных методов, основанных на 3C и секвенировании (Hi-C, 4C, 5C).
- Уровни организации хроматина по Hi-C.

# Введение: пространственная организация генома эукариот

Факторы хроматина → Структура → Функция

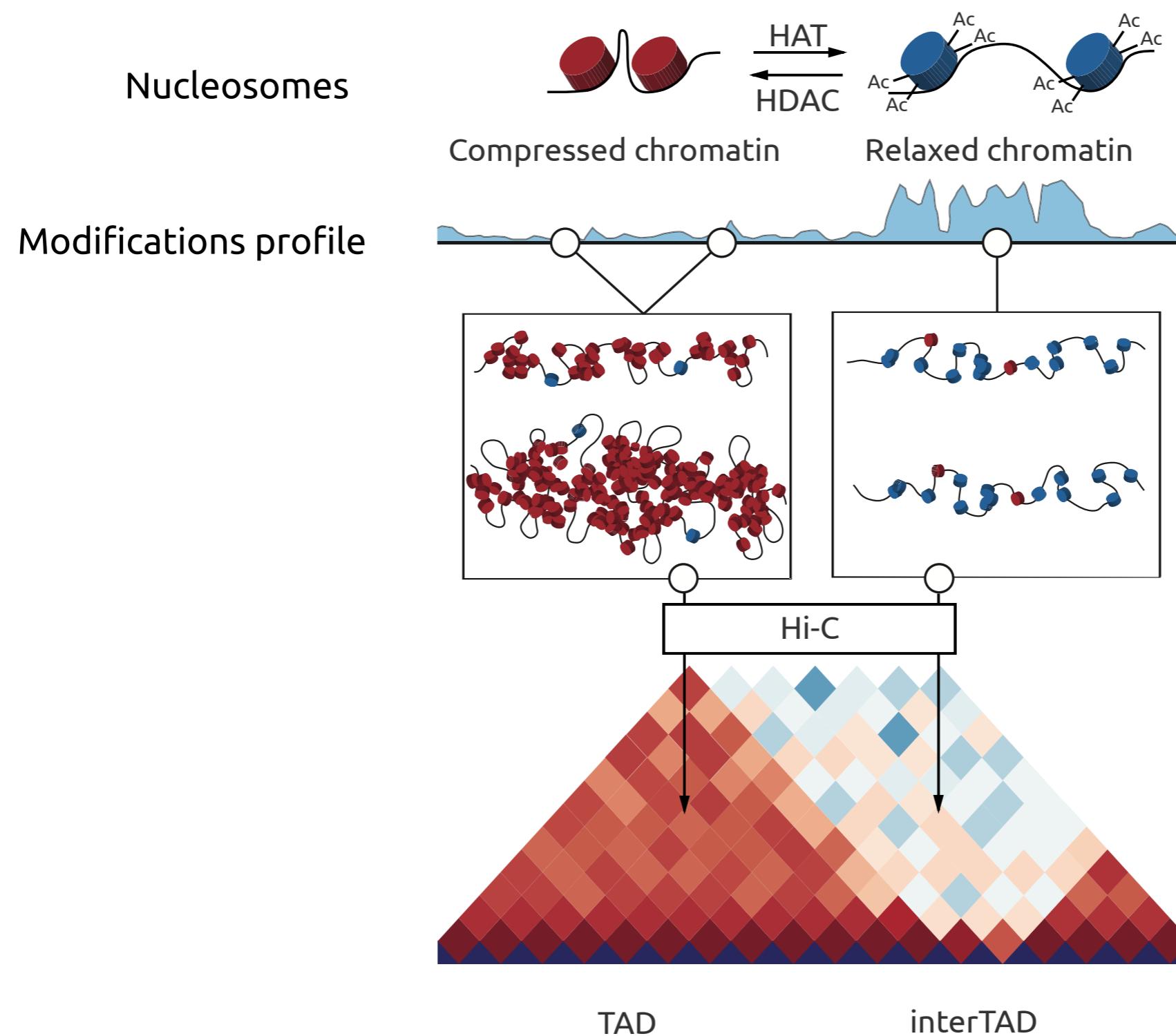
Модификации гистонов  
Связывание транскрипционных факторов  
Некодирующие РНК  
Модификации нуклеотидов  
Связывание с оболочкой ядра



Репликация  
Рекомбинация  
Регуляция  
Транскрипция

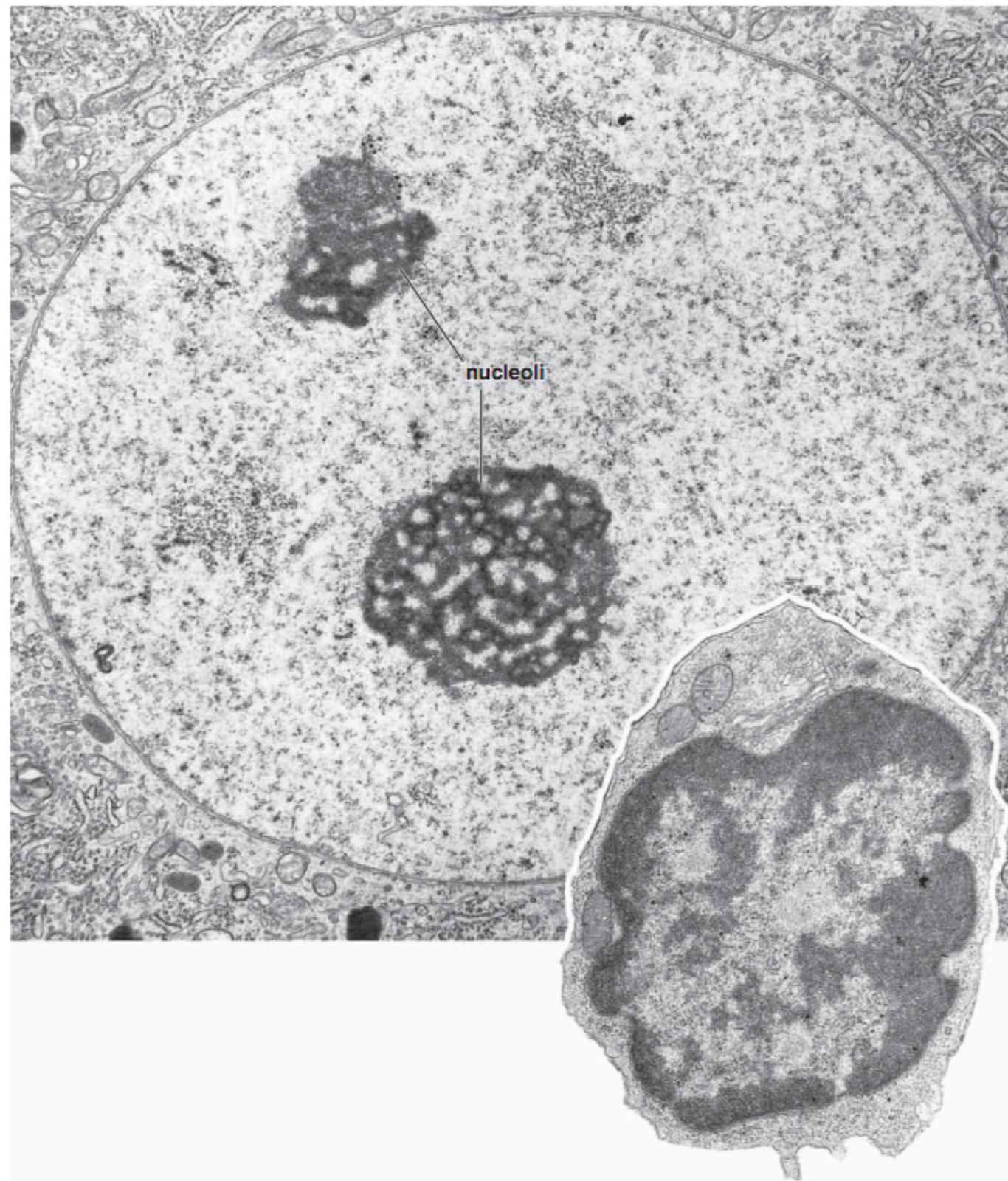
компактизация в  $10^4\text{-}10^5$  раз

# Пример взаимосвязи эпигенетики и структуры



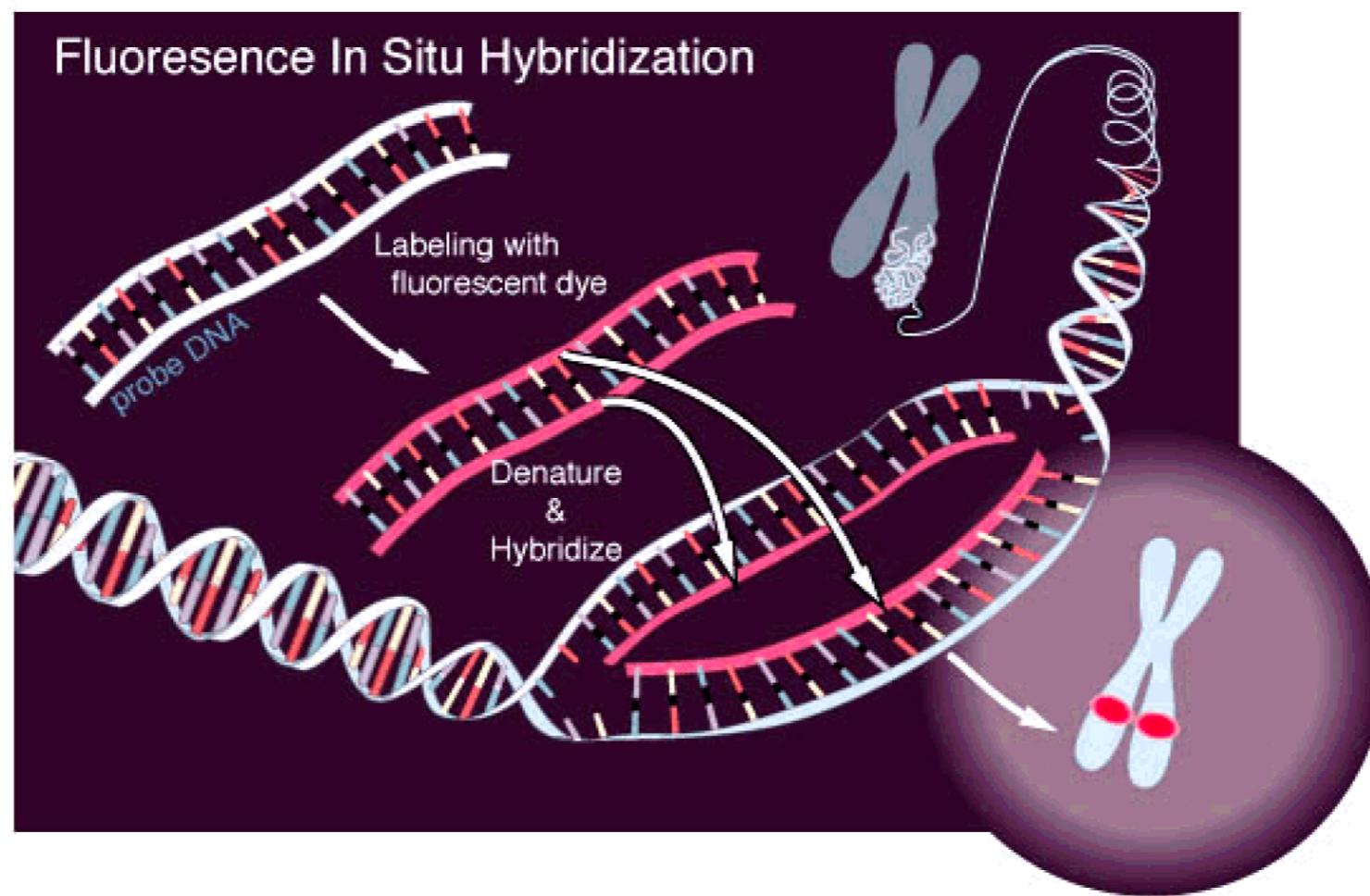
Как измерить ДНК-ДНК взаимодействия?

# Микроскопия ядра



Ros 2006 "Histology Atlas with Correlated Cell and Molecular Biology"

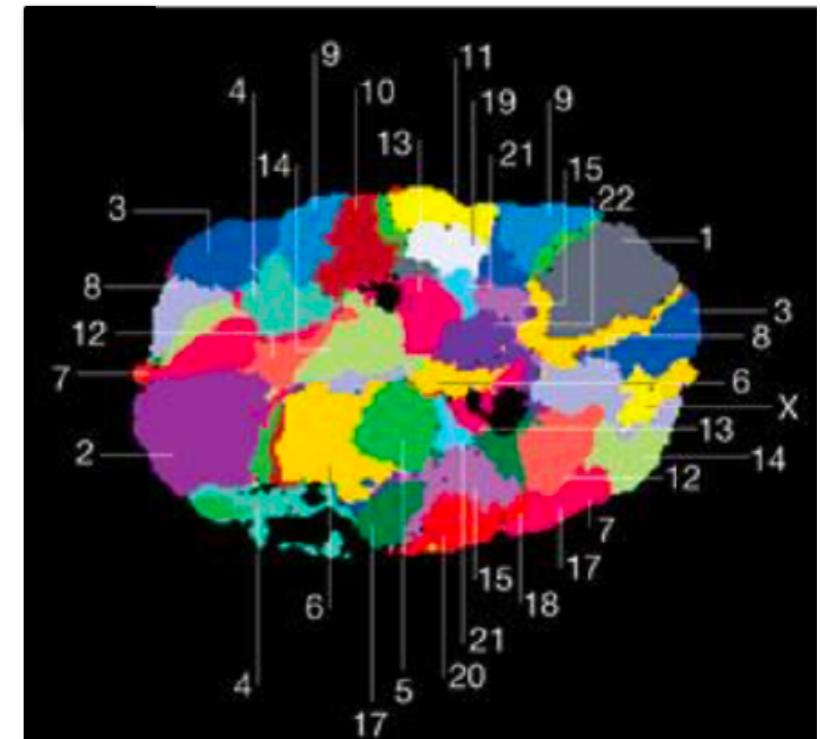
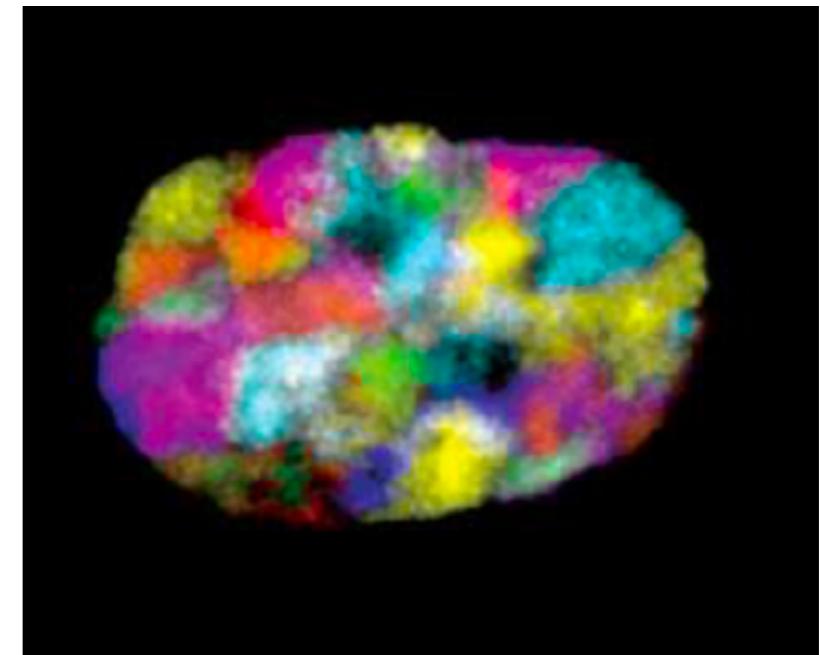
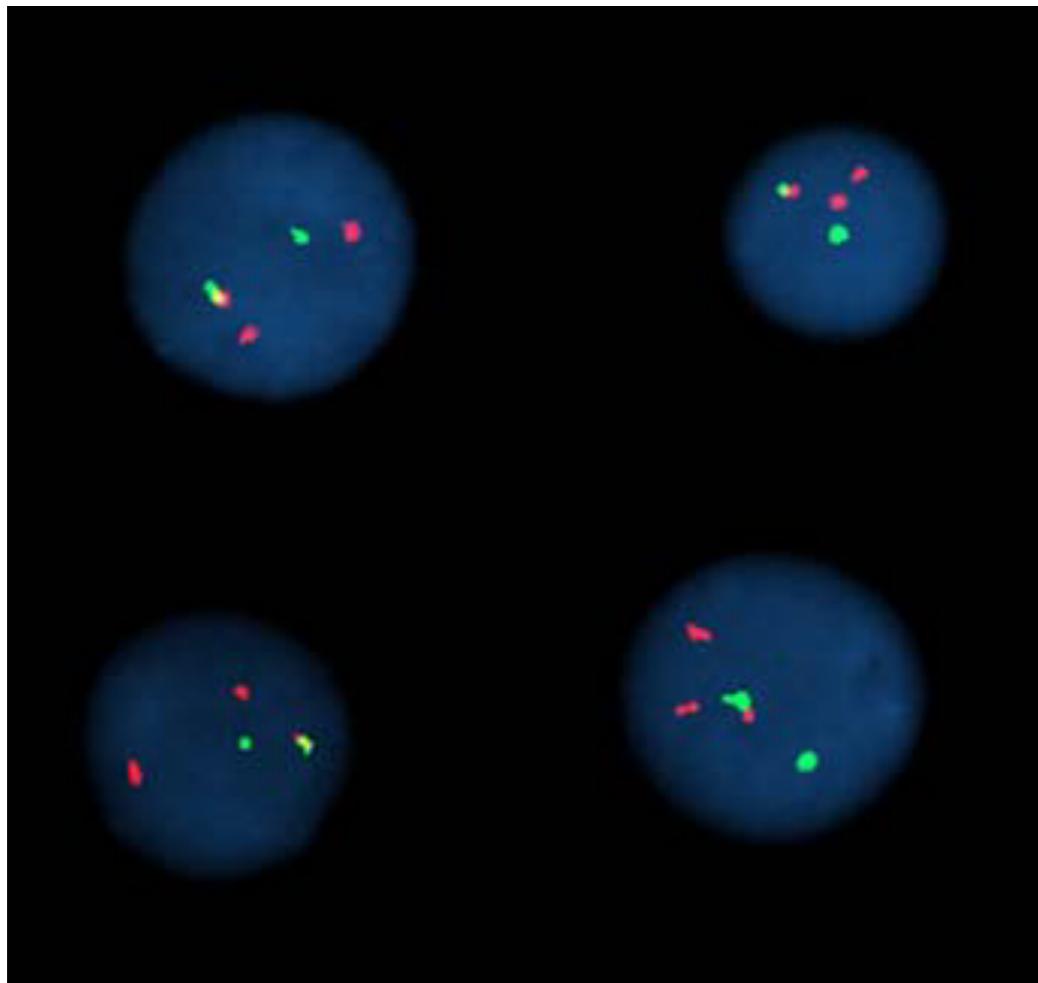
# FISH-МИКРОСКОПИЯ



# FISH-МИКРОСКОПИЯ

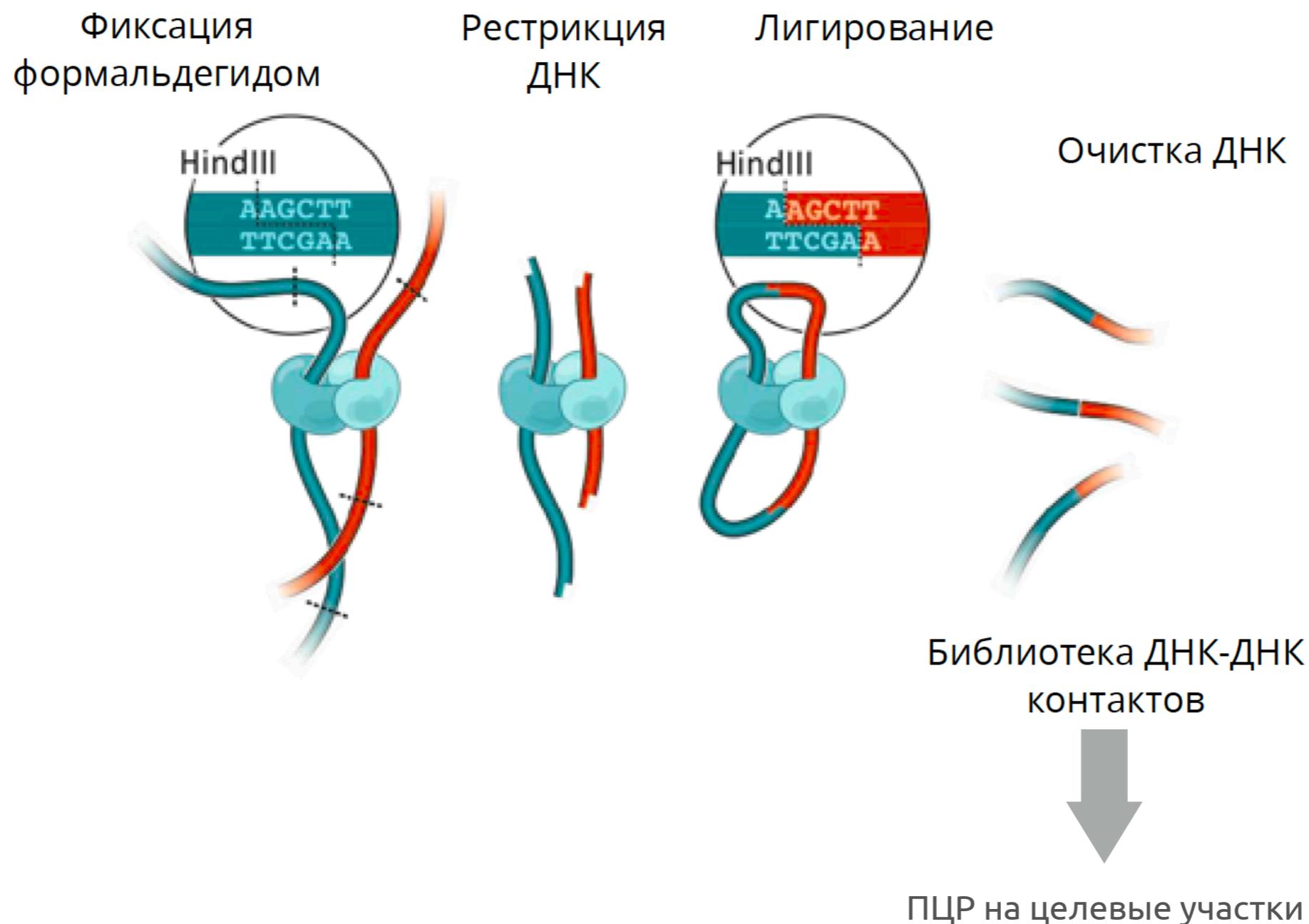
Fluorescent *in situ* hybridization  
с метками на хромосомы:

Fluorescent *in situ* hybridization  
для двух меток:



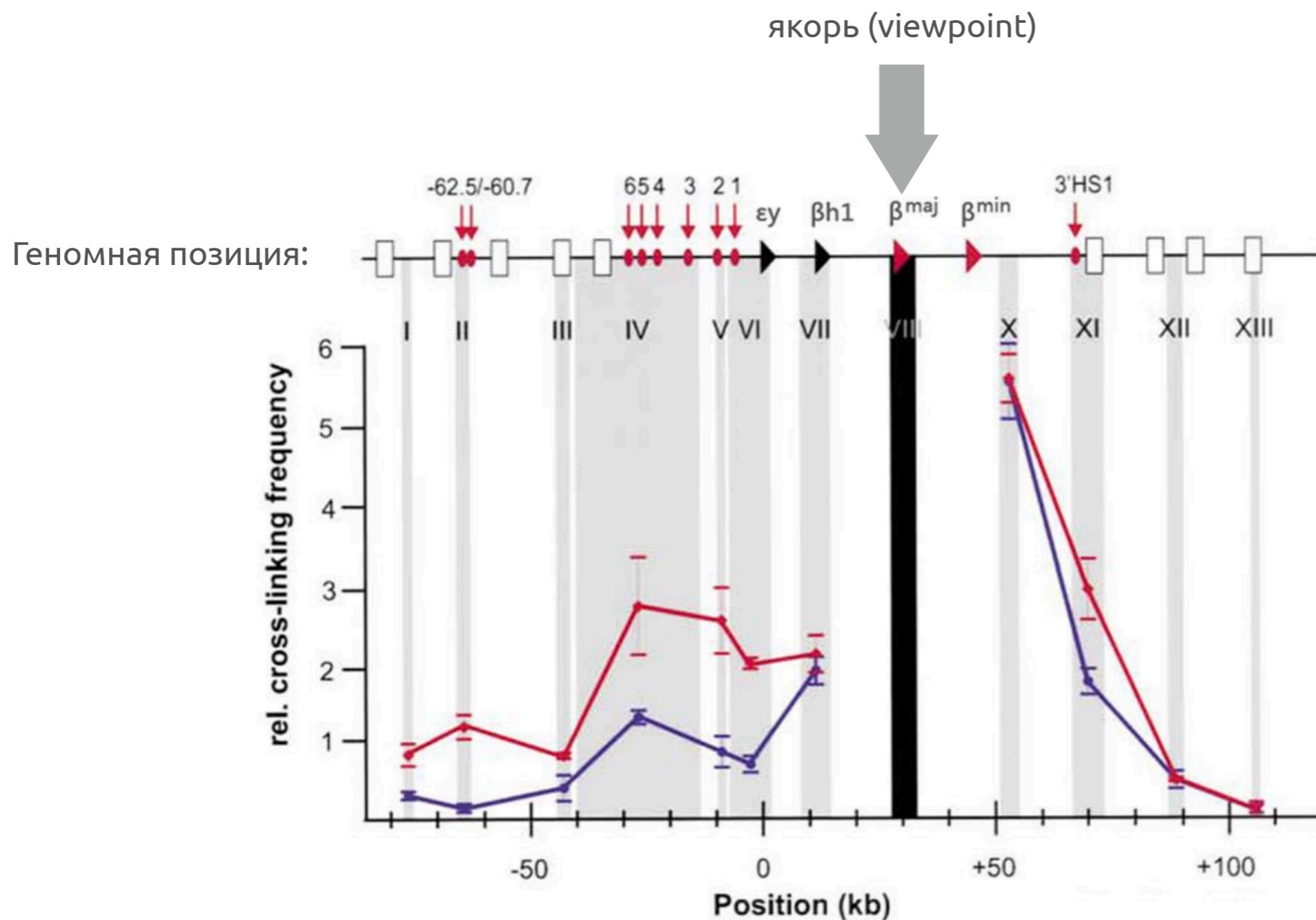
# Метод фиксации конформации хромосом

- Chromosomes conformation capture (3C):



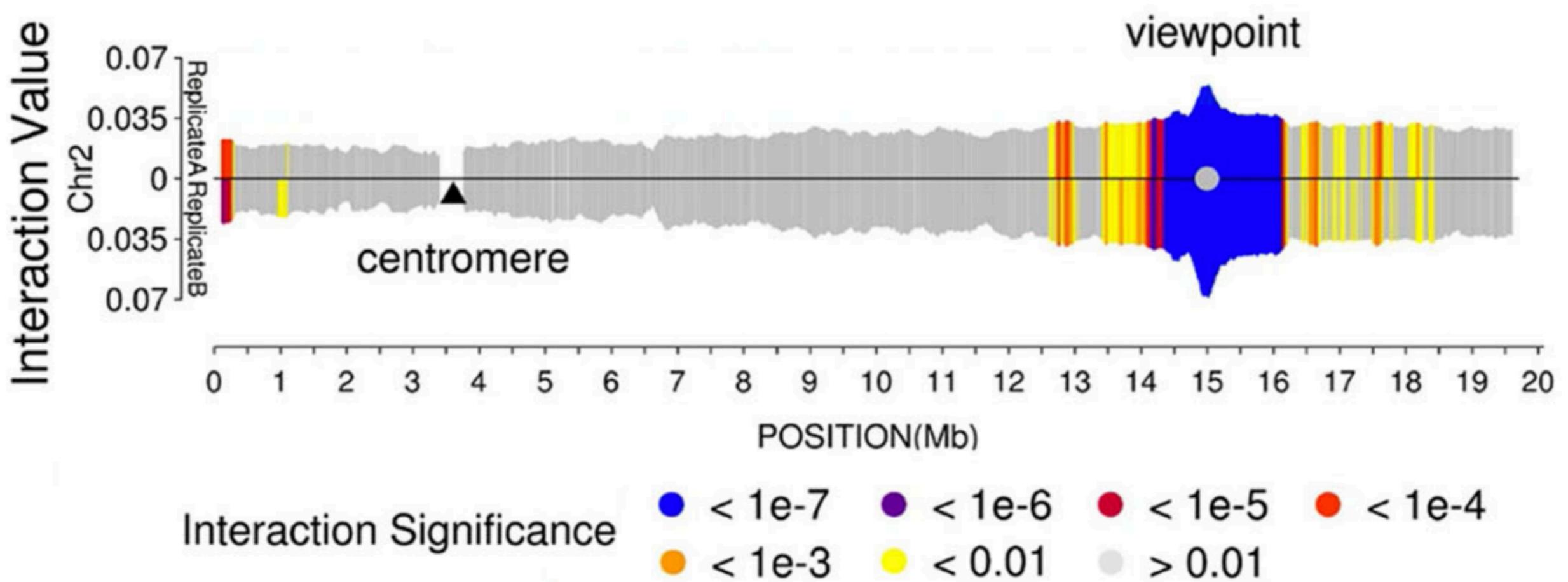
# Пример ЗС-анализа

- Регион бета-глобина млекопитающих:



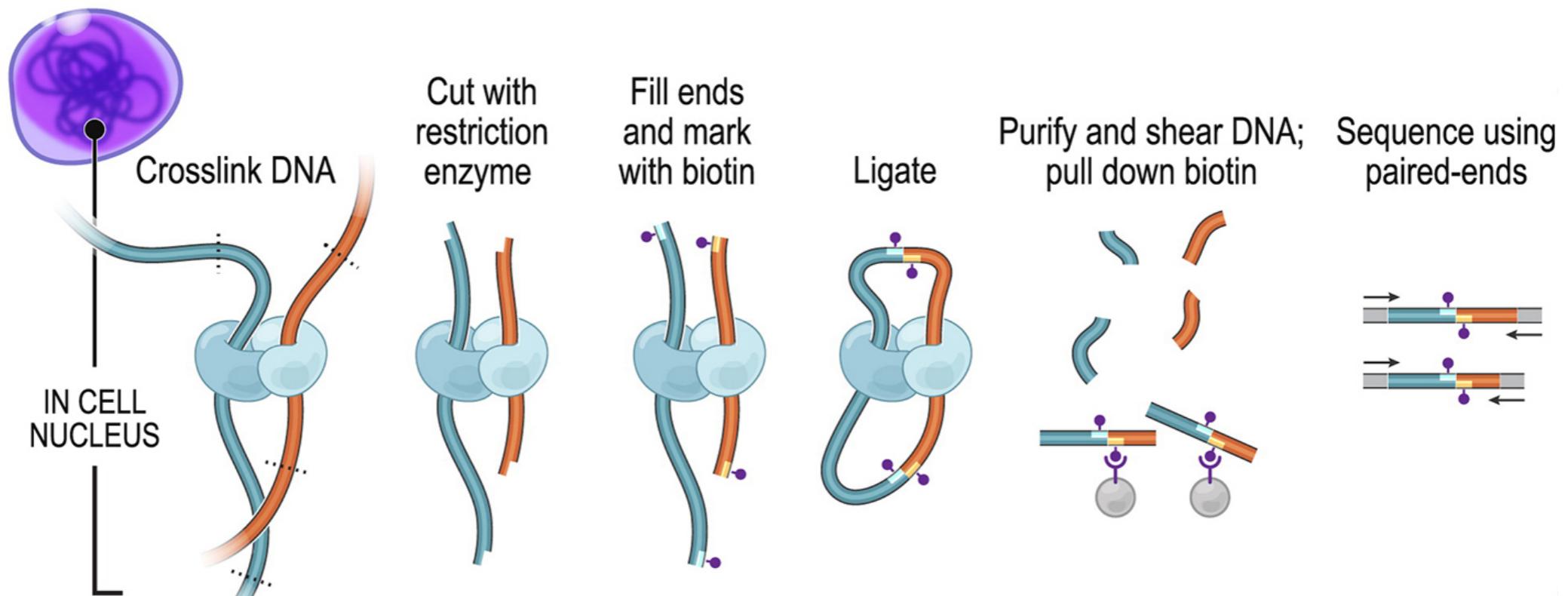
## 4C: один против всех (one-vs-all)

- Регион бета-глобина млекопитающих:

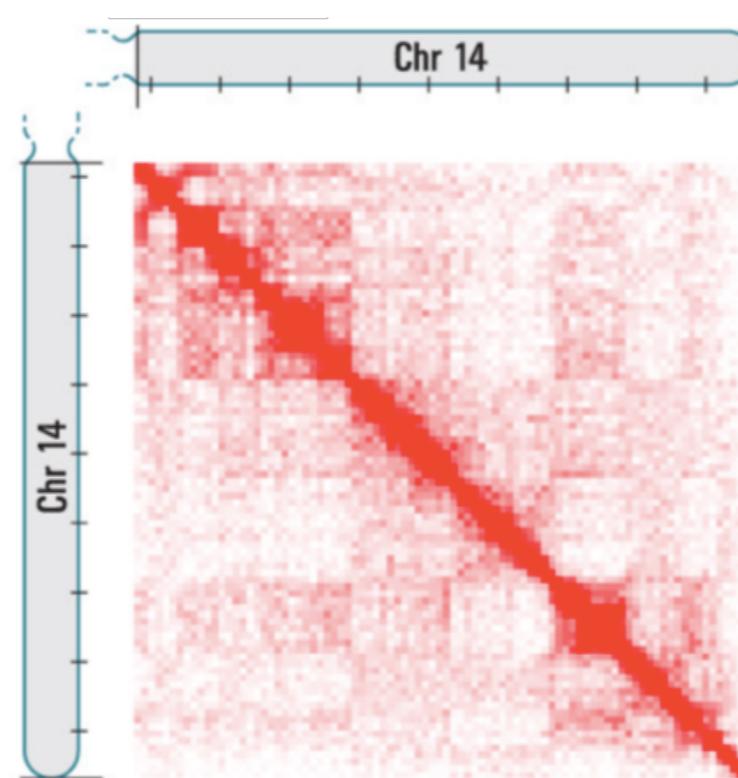


# Hi-C: high-throughput chromosomes conformation capture

Протокол:



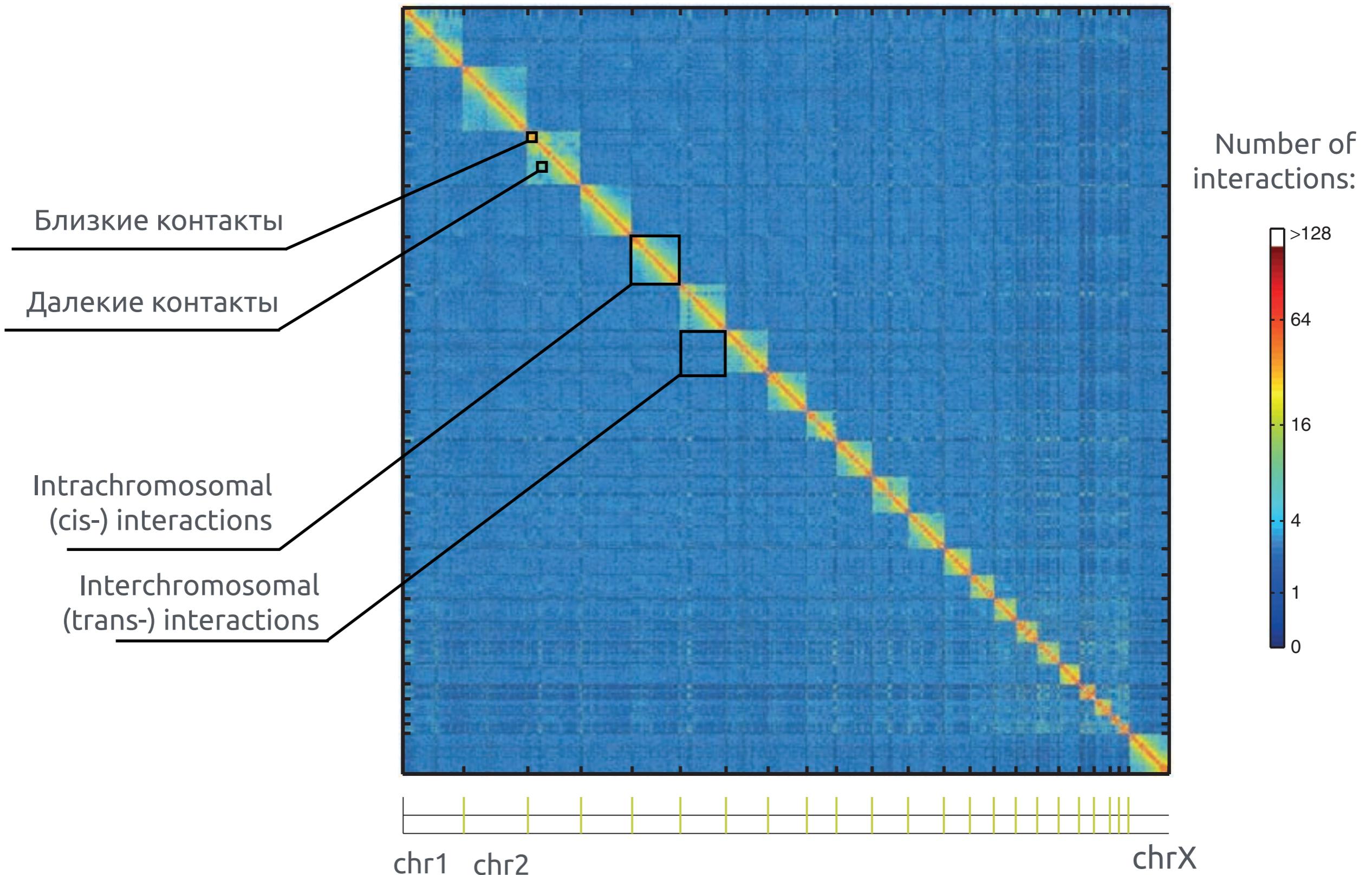
Карта ДНК-ДНК взаимодействий:



# Разнообразие методов семейства 3C:

Type of probing	Assay abbreviation	Full assay name	Year
1 vs 1	3C	Chromosome conformation capture	2002
1 vs Many/All	Multiplexed 3C-seq	Multiplexed chromosome conformation capture sequencing	2011
	Open-ended 3C	Open-ended chromosome conformation capture	2006
	4C	Chromosome conformation capture-on-chip	2006
	ACT	Associated chromosome trap	2006
	e4C	Enhanced chromosome conformation capture-on-chip	2010
	3C-DSL	Chromosome conformation capture combined with DNA selection and ligation	2011
	4C-seq	Chromosome conformation capture-on-chip combined with high-throughput sequencing	2011
	4C	Circular chromosome conformation capture	2012
	TLA	Targeted locus amplification	2014
Many vs Many	5C	Chromosome conformation capture carbon copy	2006
	ChIA-PET	Chromatin interaction analysis paired-end tag sequencing	2009
Many vs All	Capture-3C	Chromosome conformation capture coupled with oligonucleotide capture technology	2014
	Capture-HiC	Hi-C coupled with oligonucleotide capture technology	2014
All vs All	GCC	Genome conformation capture	2009
	Hi-C	Genome-wide chromosome conformation capture	2009
	ELP	Genome-wide chromosome conformation capture with enrichment of ligation products	2010
	TCC	Tethered conformation capture	2012
	Single-cell Hi-C	Single-cell genome-wide chromosome conformation capture	2013
	In situ Hi-C	Genome-wide chromosome conformation capture with in situ ligation	2014
	DNase Hi-C	Genome-wide chromosome conformation capture with DNase I digestion	2015
	Micro-C	Genome-wide chromosome conformation capture with micrococcal nuclease digestion	2015
	GAM	Genome Architecture Mapping	2017

# Контактная карта ДНК-ДНК взаимодействий:

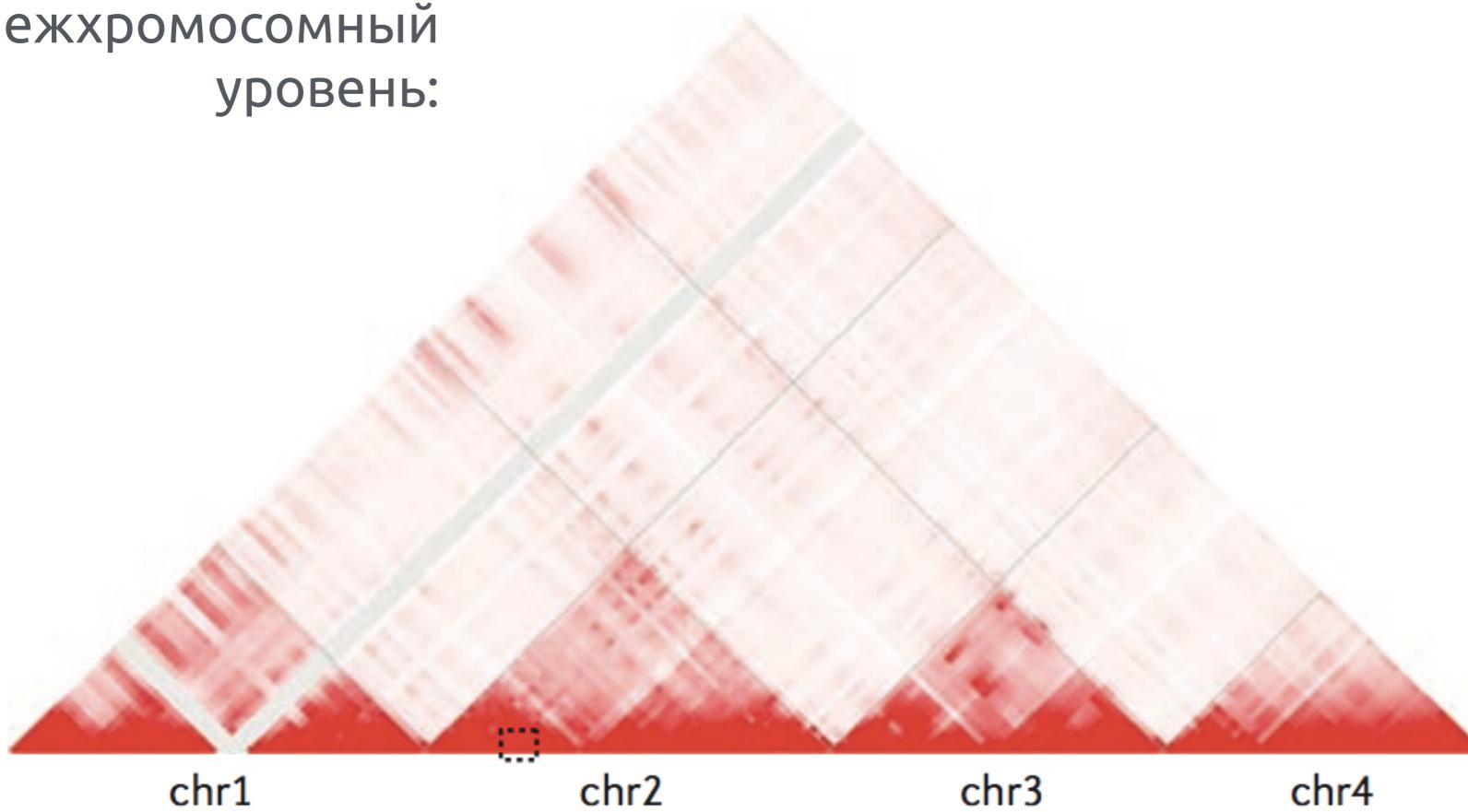


Adopted from Imakaev et al. *Nature Methods* 2012

# Хромосомные территории

- Самый высокий уровень организации карты
- Транс-контакты (межхромосомные) достаточно редки.
- Цис-контакты (внутрихромосомные) более частые.

Межхромосомный  
уровень:

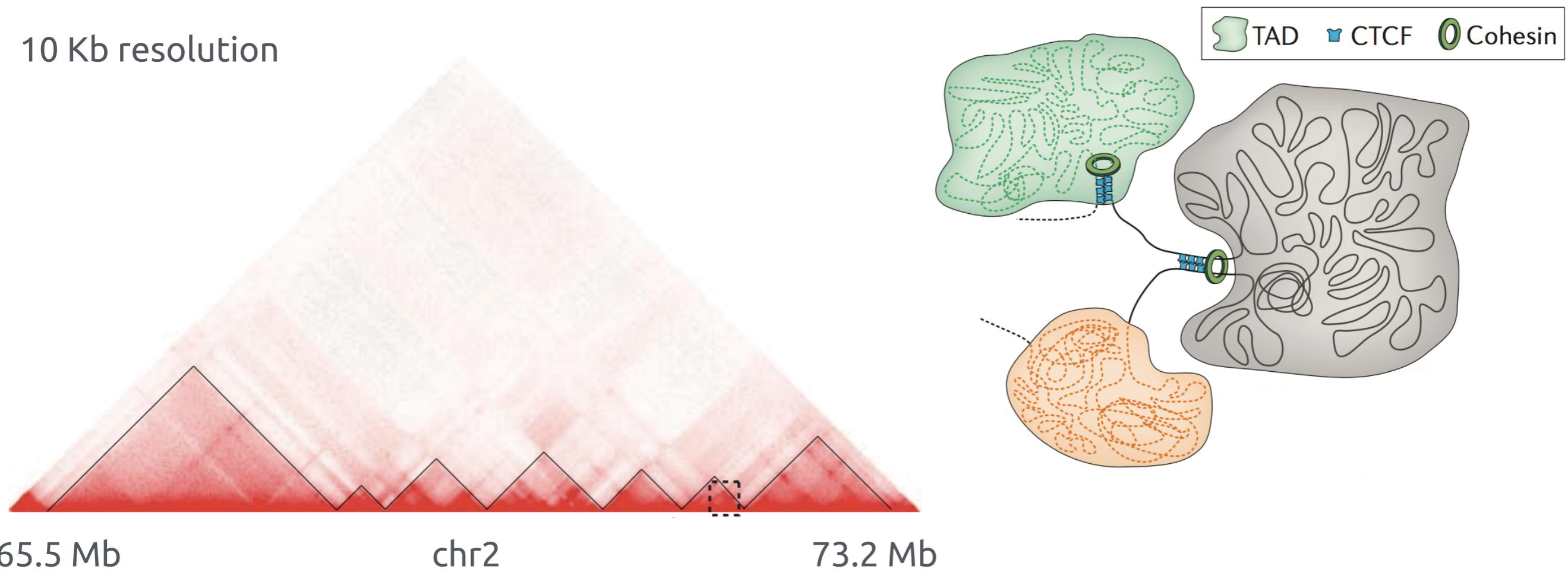


Map is rotated by 90°, upper triangle visualized



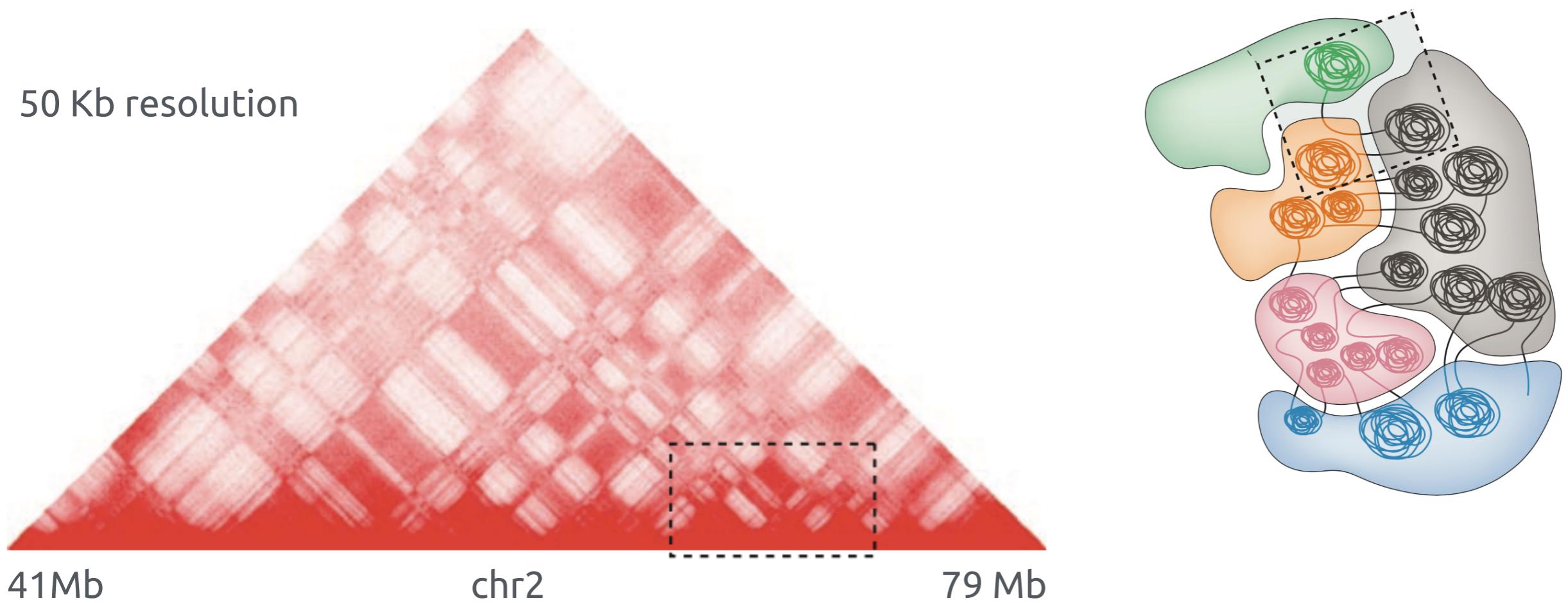
# Топологически Ассоциированные Домены (ТАДы)

- Внутри хромосом хроматин организован в локальные компактные структуры - ТАДы



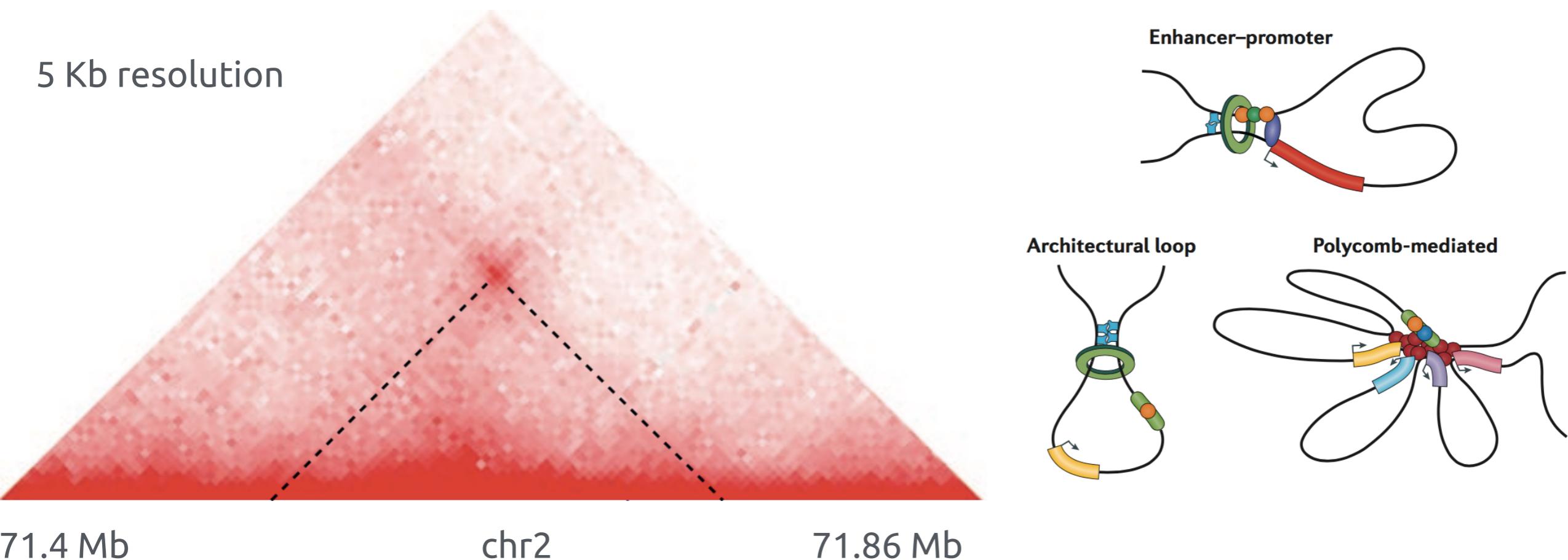
# Компартменты

- ТАДы взаимодействуют на далеких геномных расстояниях
- Активные ТАДы взаимодействуют с активными, неактивные - с неактивными.
- Такие структуры называются компартментами хроматина

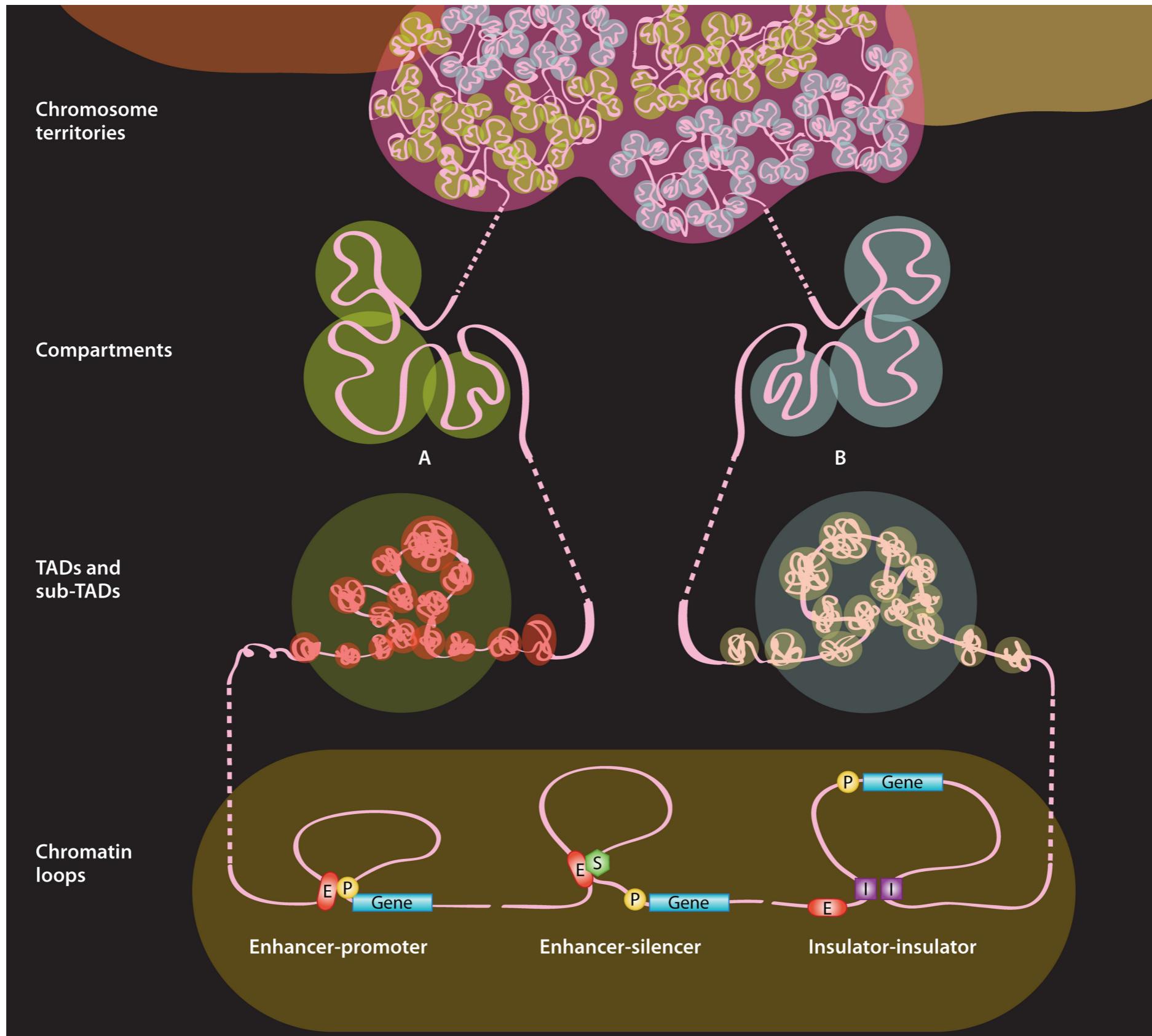


# Петли (обогащенные взаимодействия) хроматина

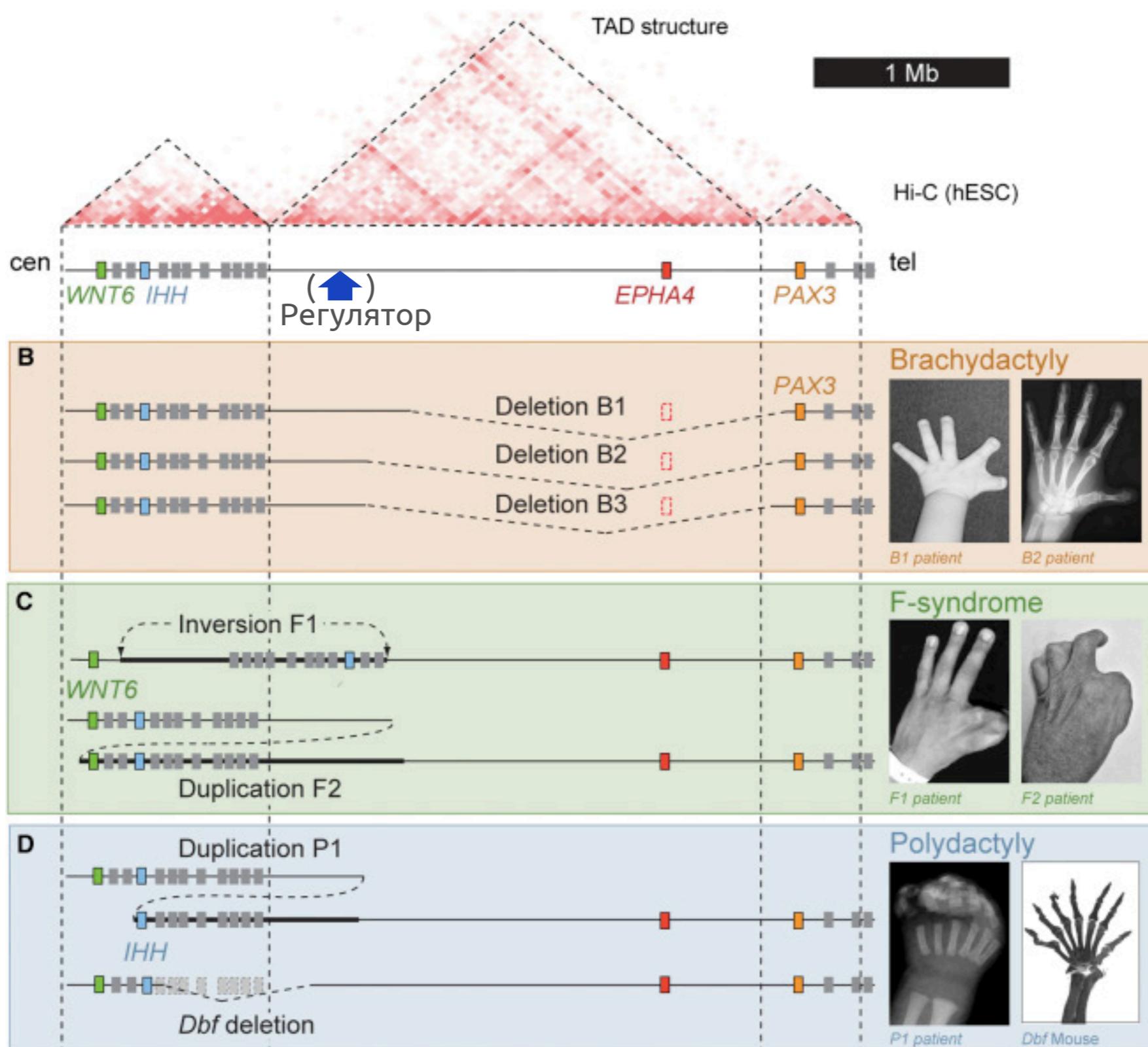
- У генов зачастую есть регуляторы, которые расположены на некотором отдалении от промотора.
- Регуляторы входят в контакт с регулируемыми генами, образуя петли.
- Возможны архитектурные петли для достижения лучшей компактизации, не связанные с регуляцией.



# "The Zoo" of chromatin features



# Почему это важно?

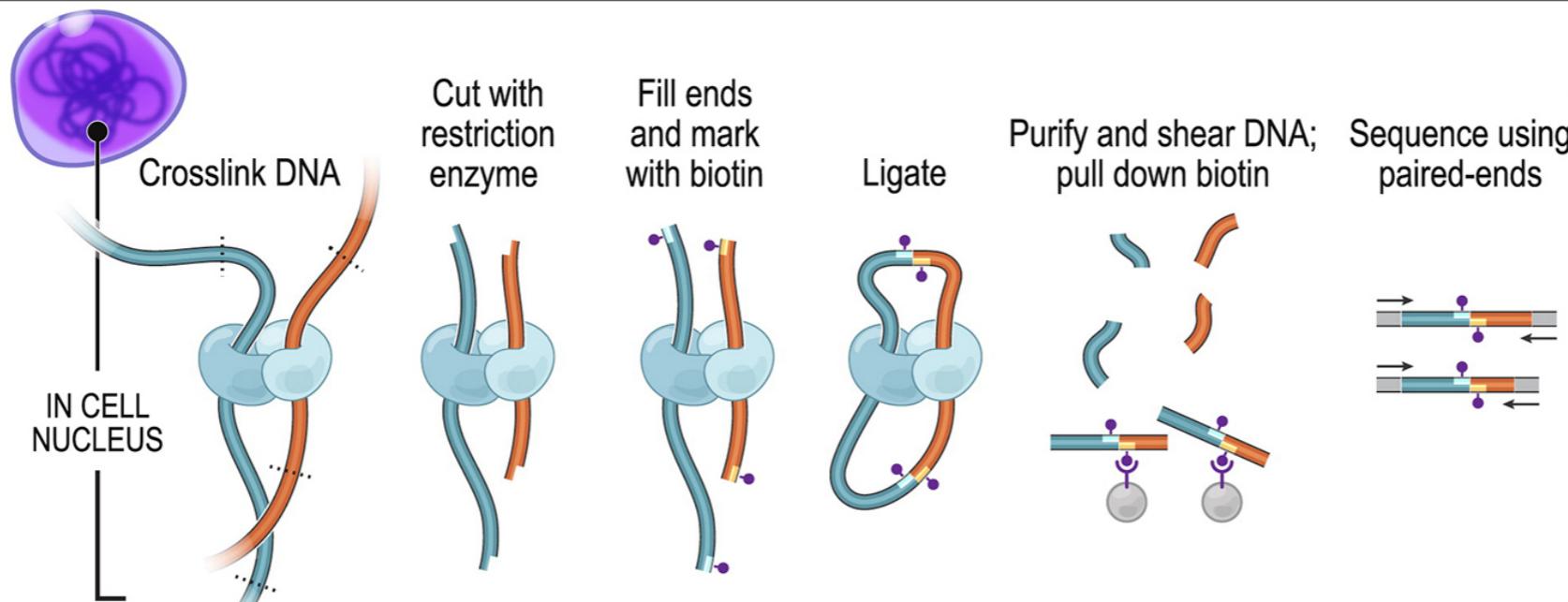


# От теории к практике

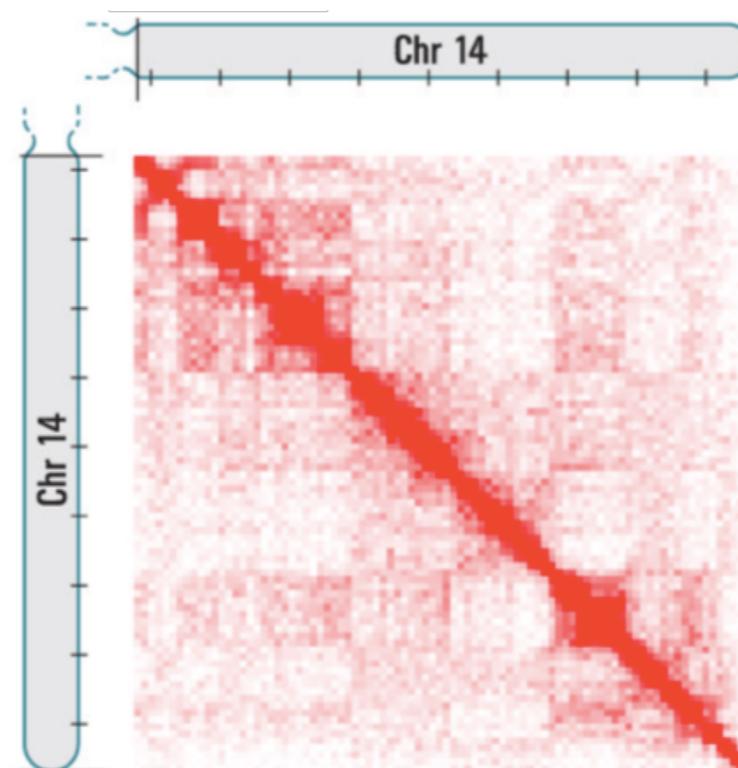
1. Особенности картирования данных ДНК-ДНК взаимодействий
2. Фильтрация данных Ні-С, бинирование, получение контактных карт
3. Итеративная коррекция

# Постановка задачи

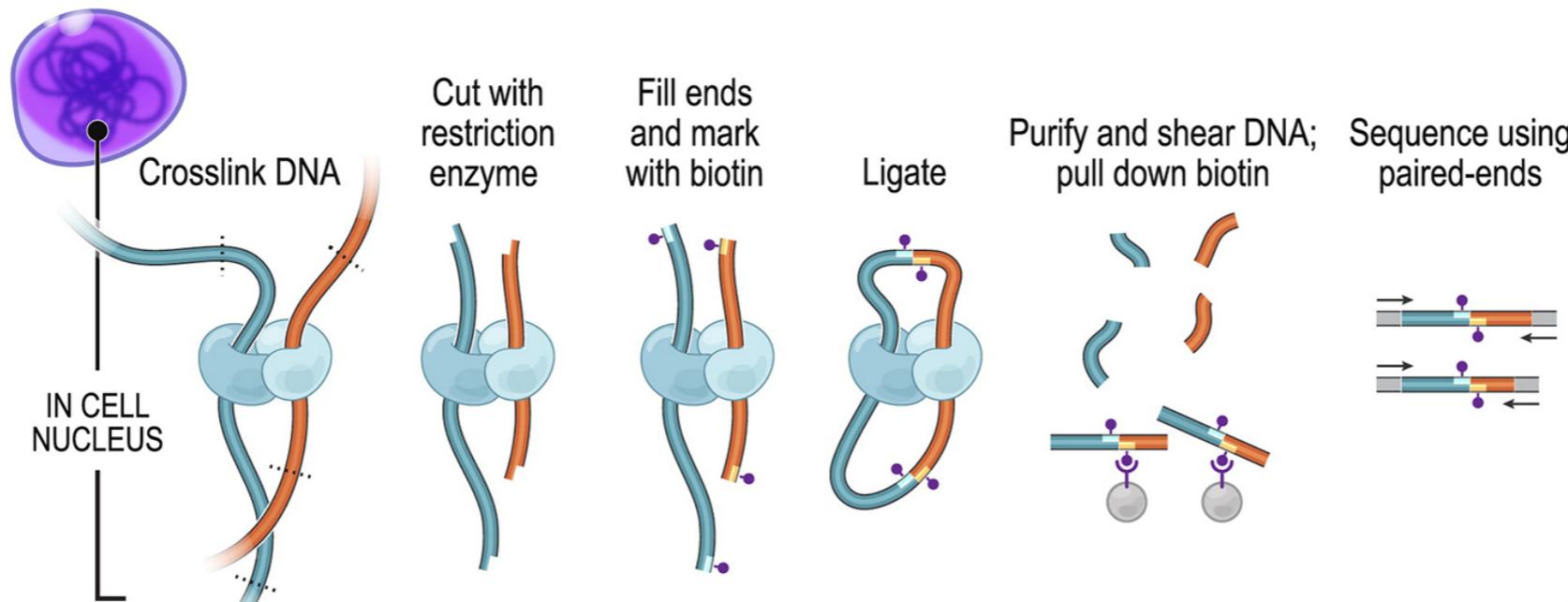
Протокол:



Карта ДНК-ДНК взаимодействий:



# Последовательность обработки Hi-C



## 1. Картирование ридов

- Получение файла с парами геномных картирований

## 2. Фильтрация контактов

- Удаление ПЦР-дупликатов
- Фильтрация по рестриктным фрагментам

## 3. Получение карты контактов

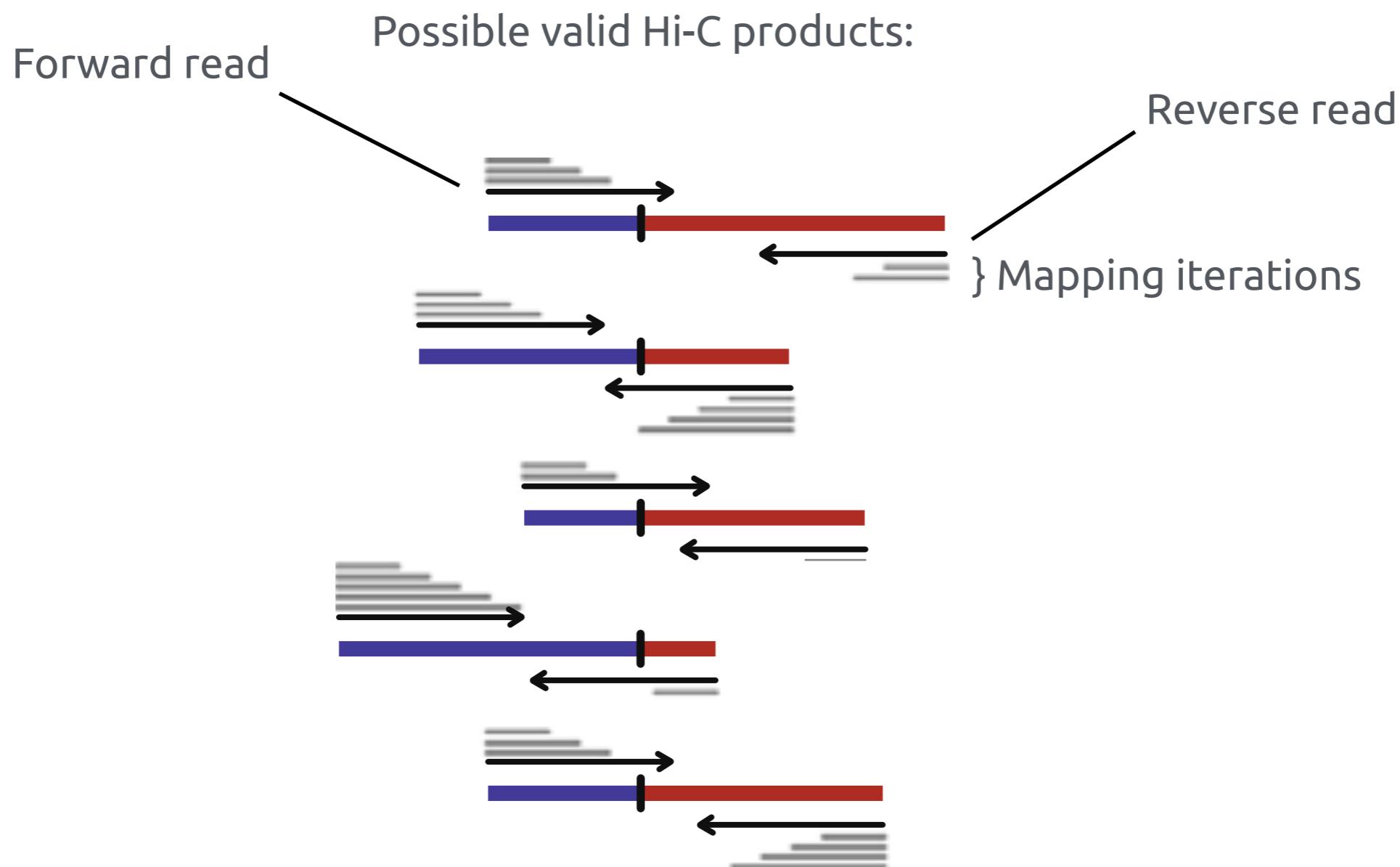
- Бинирование

## 4. Балансировка и нормализация карты

## 5. Поиск особенностей (features) карт (ТАДы, компартменты, петли)

# Картирование ридов

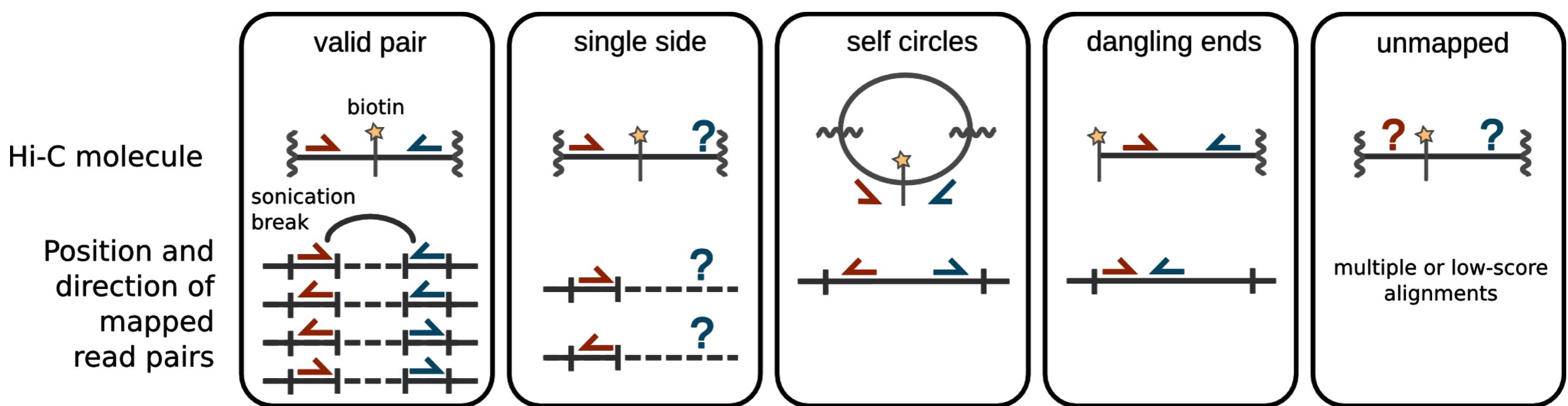
- Итеративное картирование или картирование с возможностью разбиения ридов (split read alignment, e.g. bwa mem)



Adopted from Lajoie et al., The Hitchhiker's guide to Hi-C analysis: Practical guidelines.  
Methods 2015

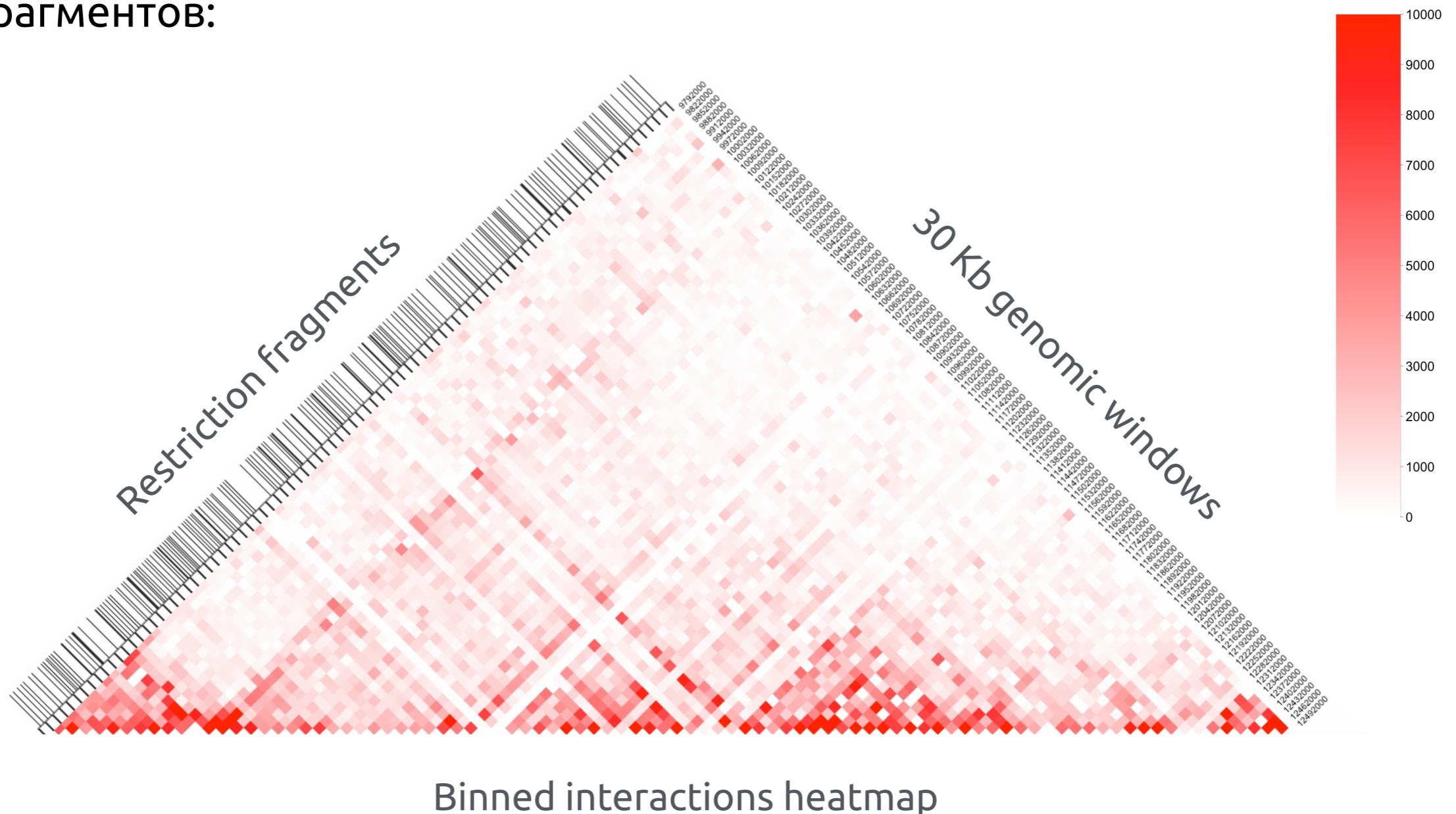
# Фильтрация на уровне рестриктных фрагментов

- Возможные ситуации картирования Hi-C:



# Бинирование

- Бины генома - это последовательные геномные окна одинакового размера.
- Каждый сайт рестрикции приписывается своему бину генома, по каждой паре бинов берется сумма контактов попадающих в них пар рестриктных фрагментов:

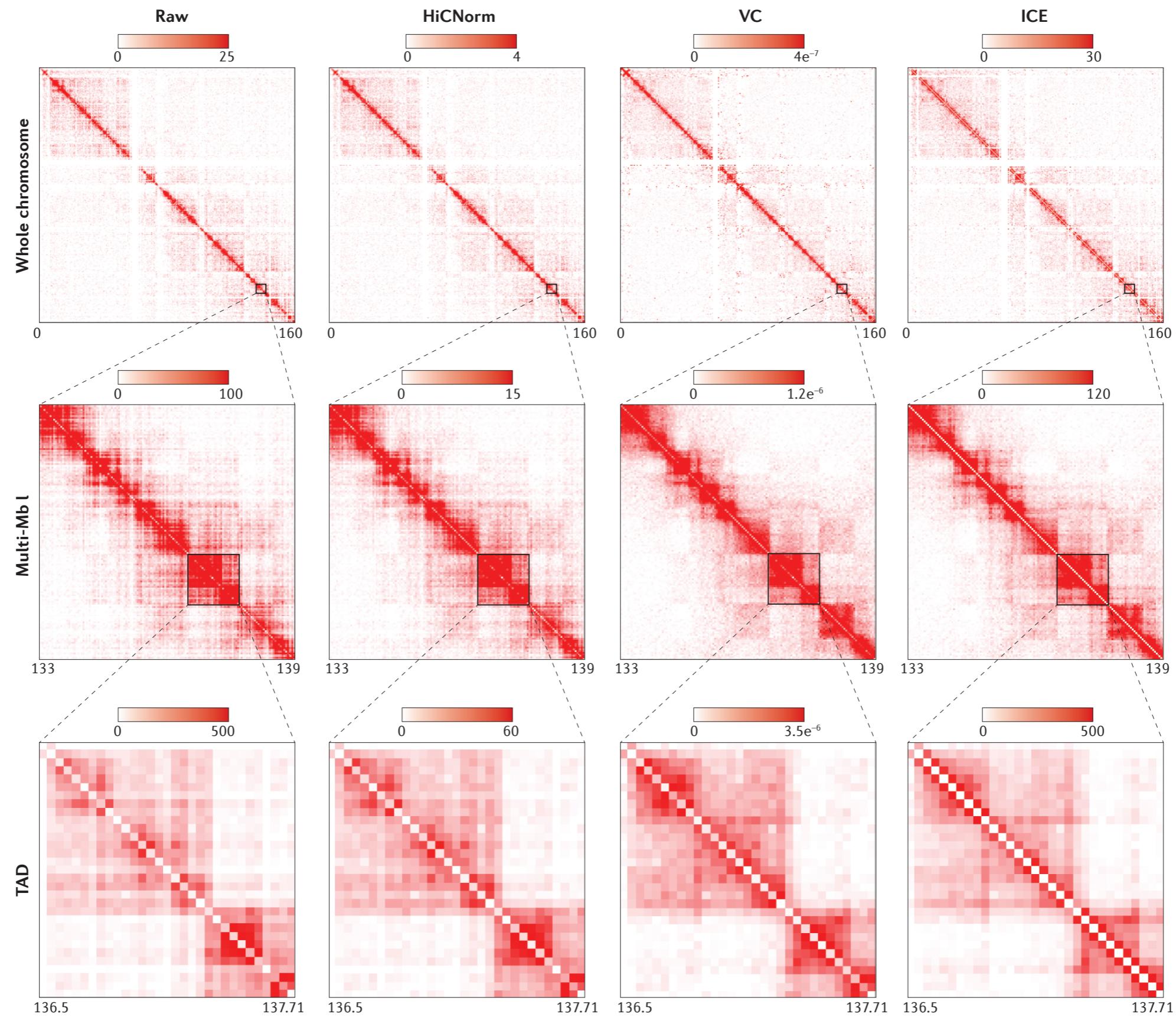


# Балансировка карты

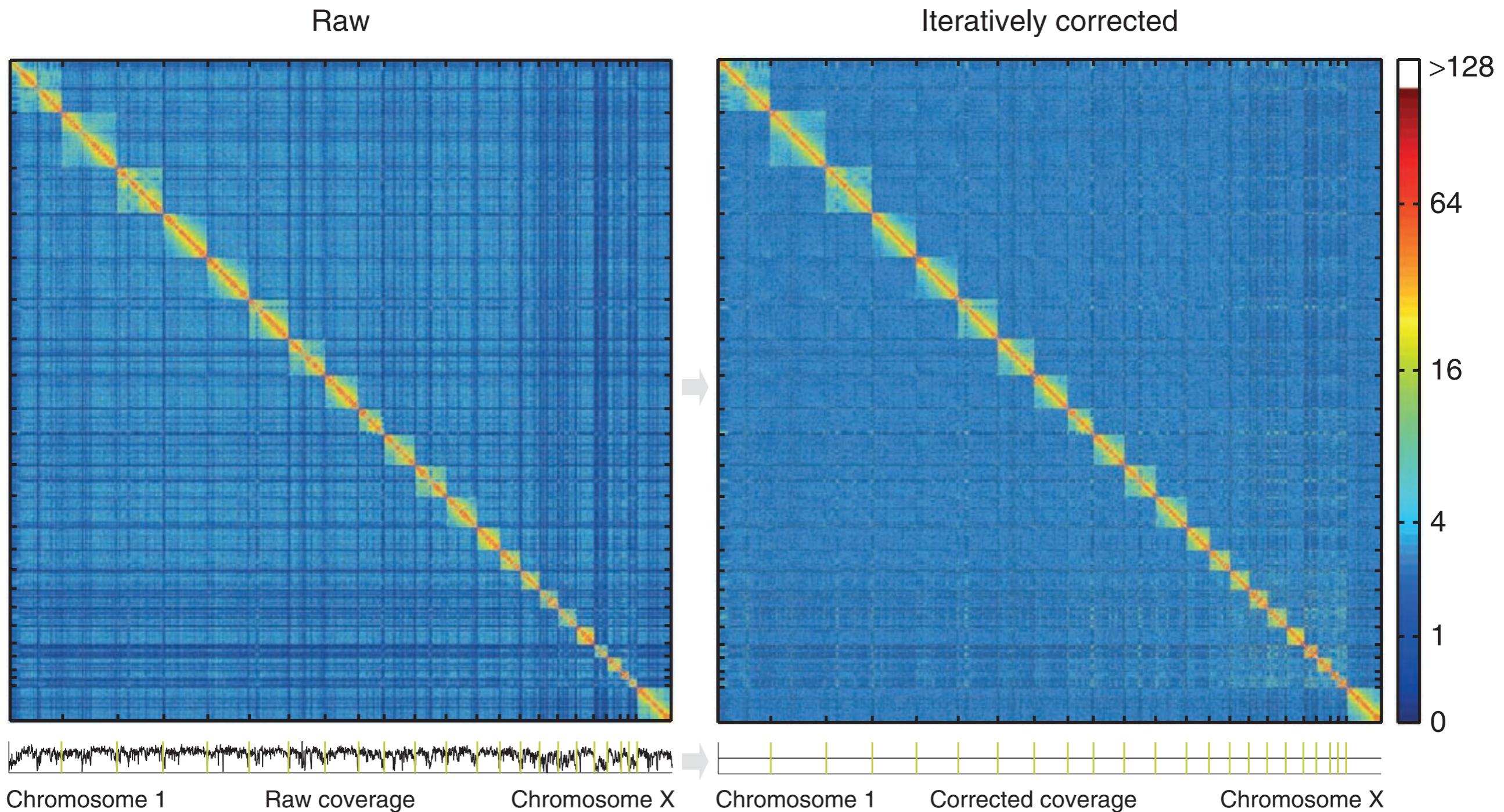
- Балансировка необходима для коррекции систематических экспериментальных погрешностей в эксперименте.
- Два основных подхода для балансировки:

Approach	Type	Model assumption	Implementation	Computational speed
Yaffe and Tanay	Explicit	Restriction enzyme fragment lengths, GC content and sequence mappability are three major systematic biases in Hi-C	Perl and R	Slow
HiCNorm			R	Fast
Iterative correction (ICE)			Python	Fast
Knight and Ruiz	Implicit	All the bias is captured by the sequencing coverage of each bin, equal visibility	JAVA	Fast
HiC-Pro			Python and R	Very fast

# Примеры результатов разных балансировок

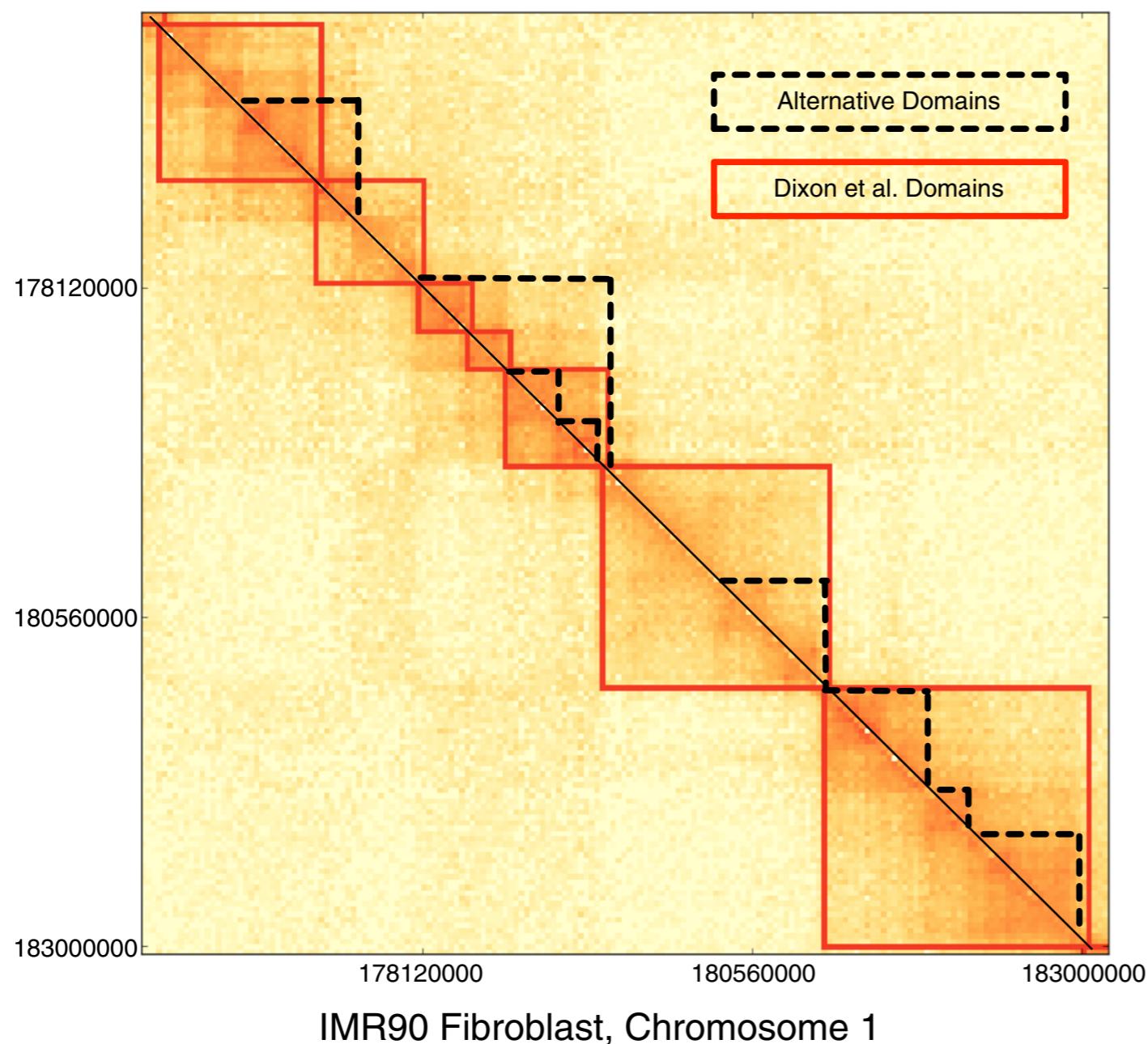


# Итеративная коррекция



# Поиск ТАДов (TADs calling)

- ТАДы иерархичны, не существует однозначного подхода для их определения:



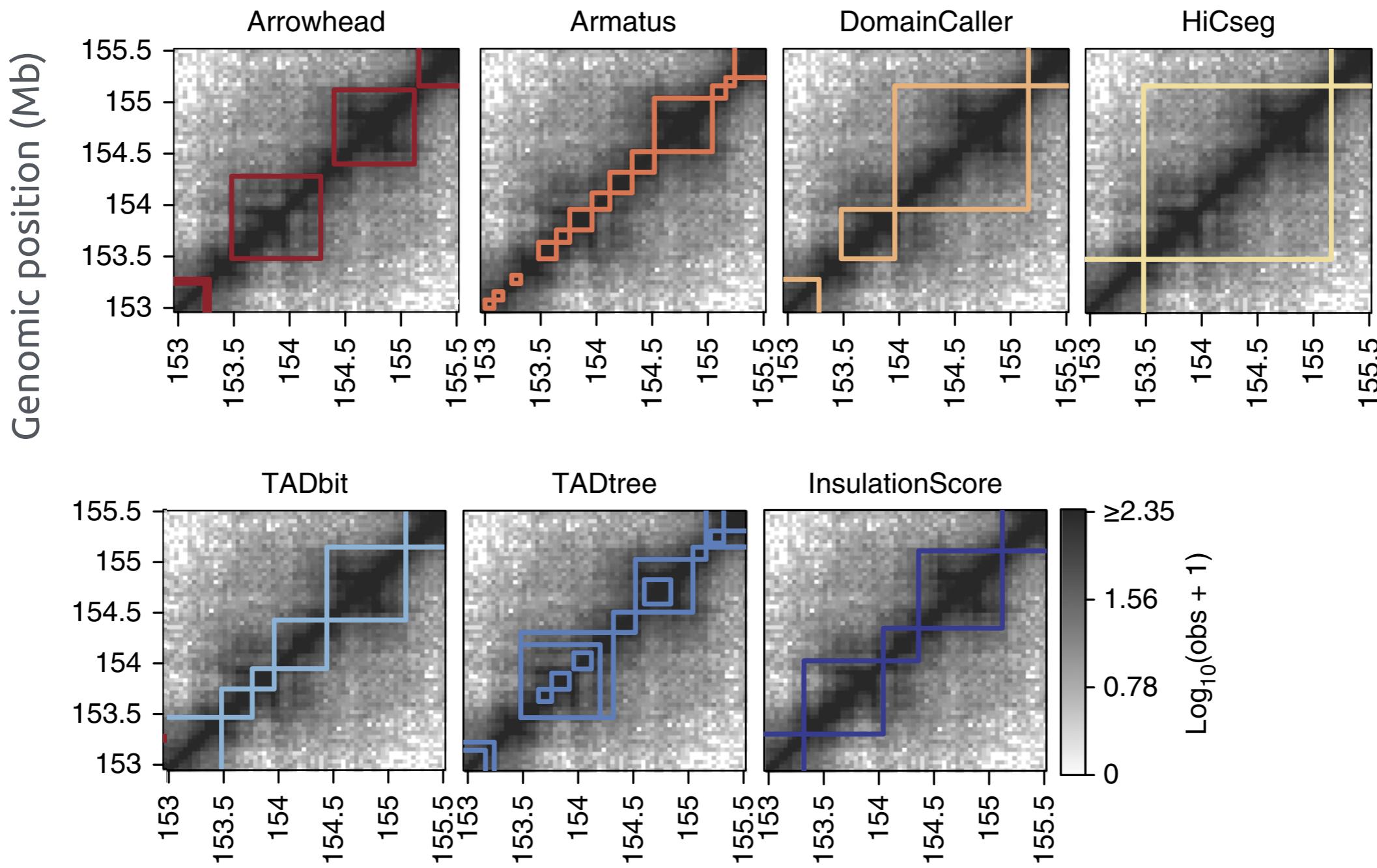
Агматус - программа для поиска ТАДов по данным матриц контактов Hi-C.

Агматус выдает несколько разметок ТАДов для разных характерных размеров ТАДов.

ТАДы иерархичны: более крупные ТАДы могут быть разбиты на ряд более мелких.

# Поиск ТАДов

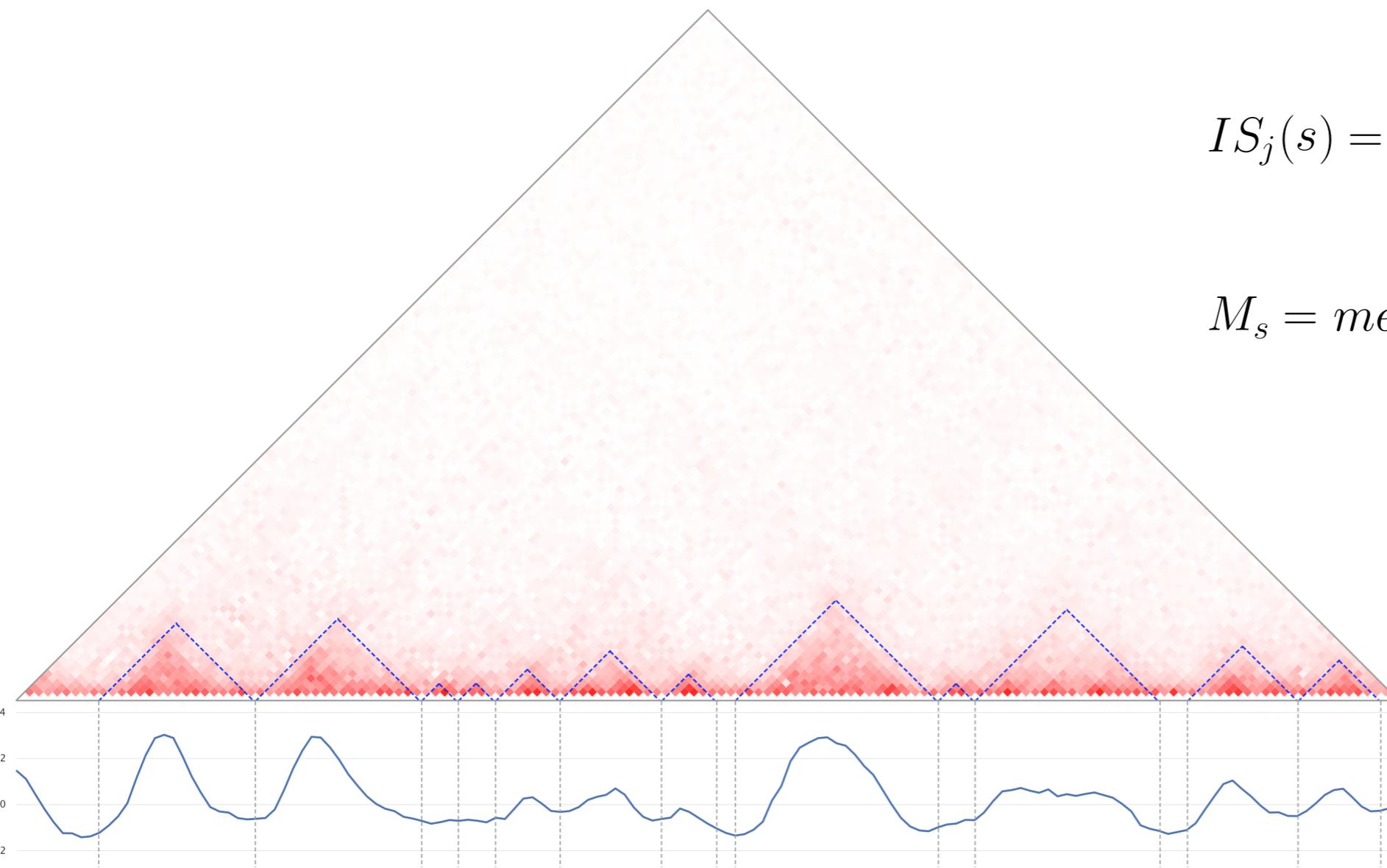
- A recent comparison of multiple TADs calling tools:



# Поиск ТАДов

- Insulation score - это интуитивный и несложный метод поиска границ ТАДов:

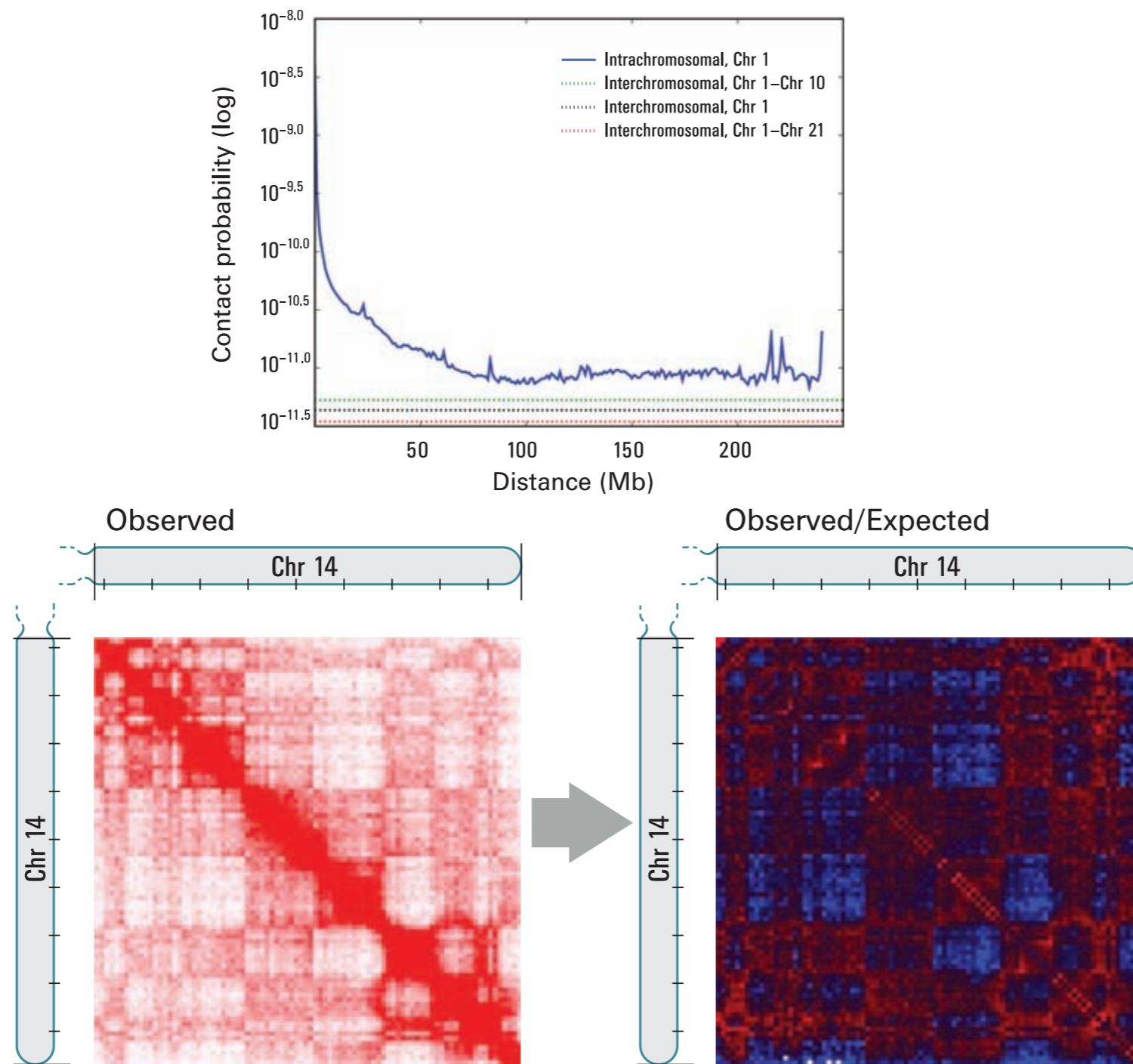
1. Подсчет Insulation score для каждого бина генома,
2. Поиск локальных минимумов профиля и соединение в ТАДы:



$$IS_j(s) = \log_2 \left( \frac{\sum_{k=j-s/2}^{k=j+s/2} C_{k,k+s}}{M_s} \right)$$
$$M_s = mean_s \left( \sum_{k=j-s/2}^{k=j+s/2} C_{k,k+s} \right)$$

# ПОИСК КОМПАРТМЕНТОВ

- Method from Lieberman-Aiden, 2009:
  - Normalization of interaction matrix by expected interactions:



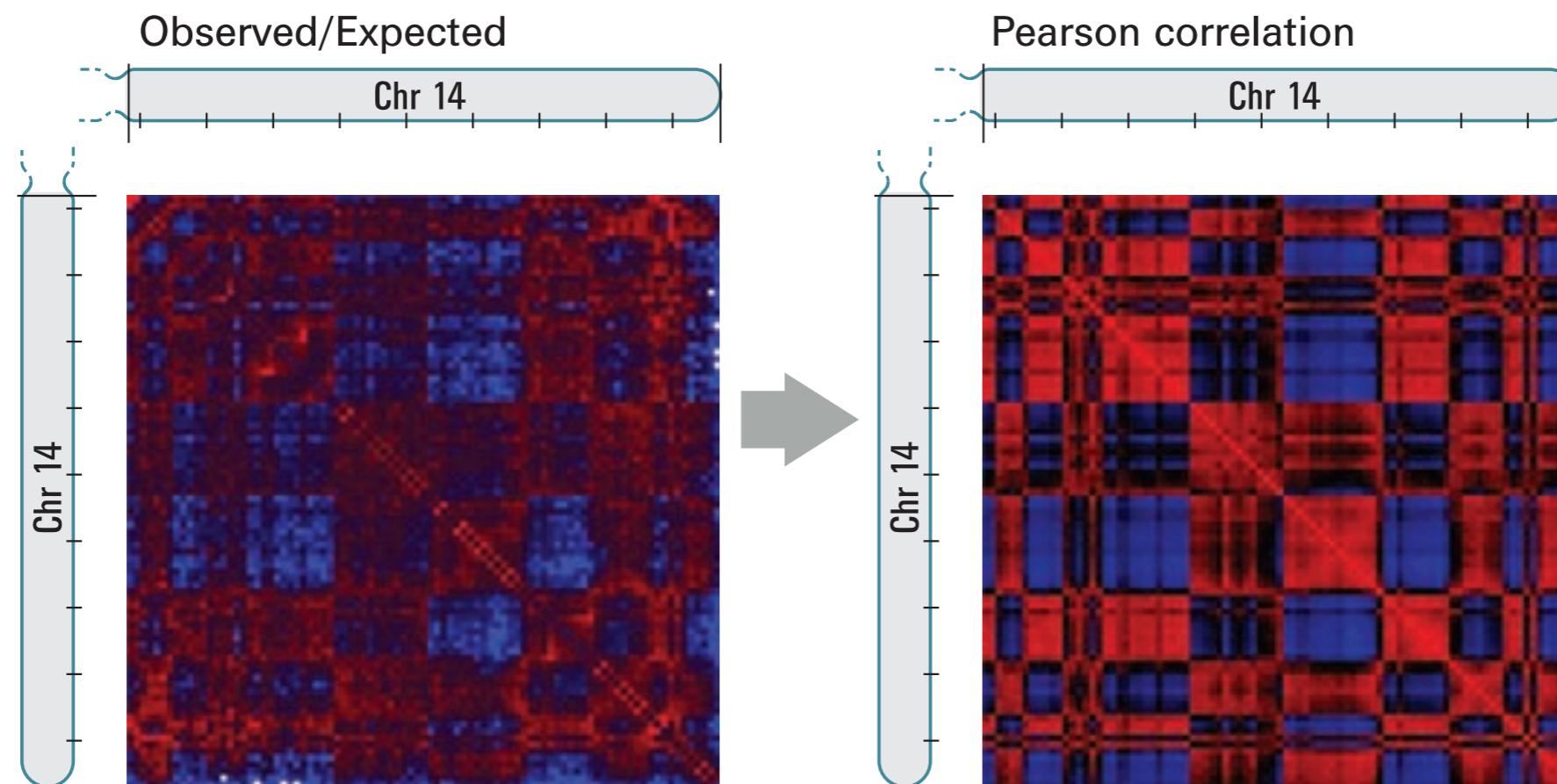
Lieberman-Aiden et al. *Nature* 2009

# Compartments calling

- Method from 2009:

②

## Calculation of Pearson correlation

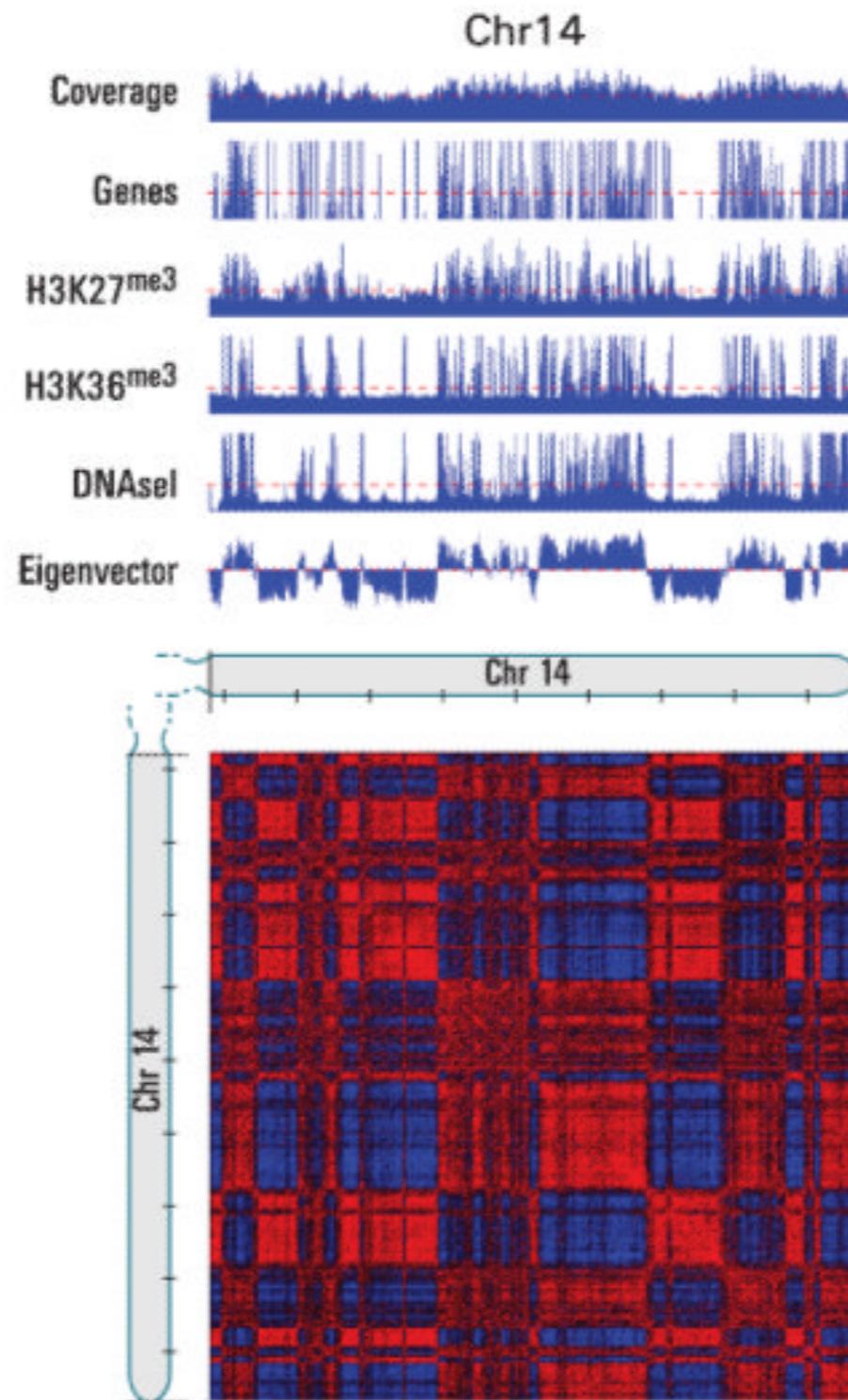


# Compartments calling

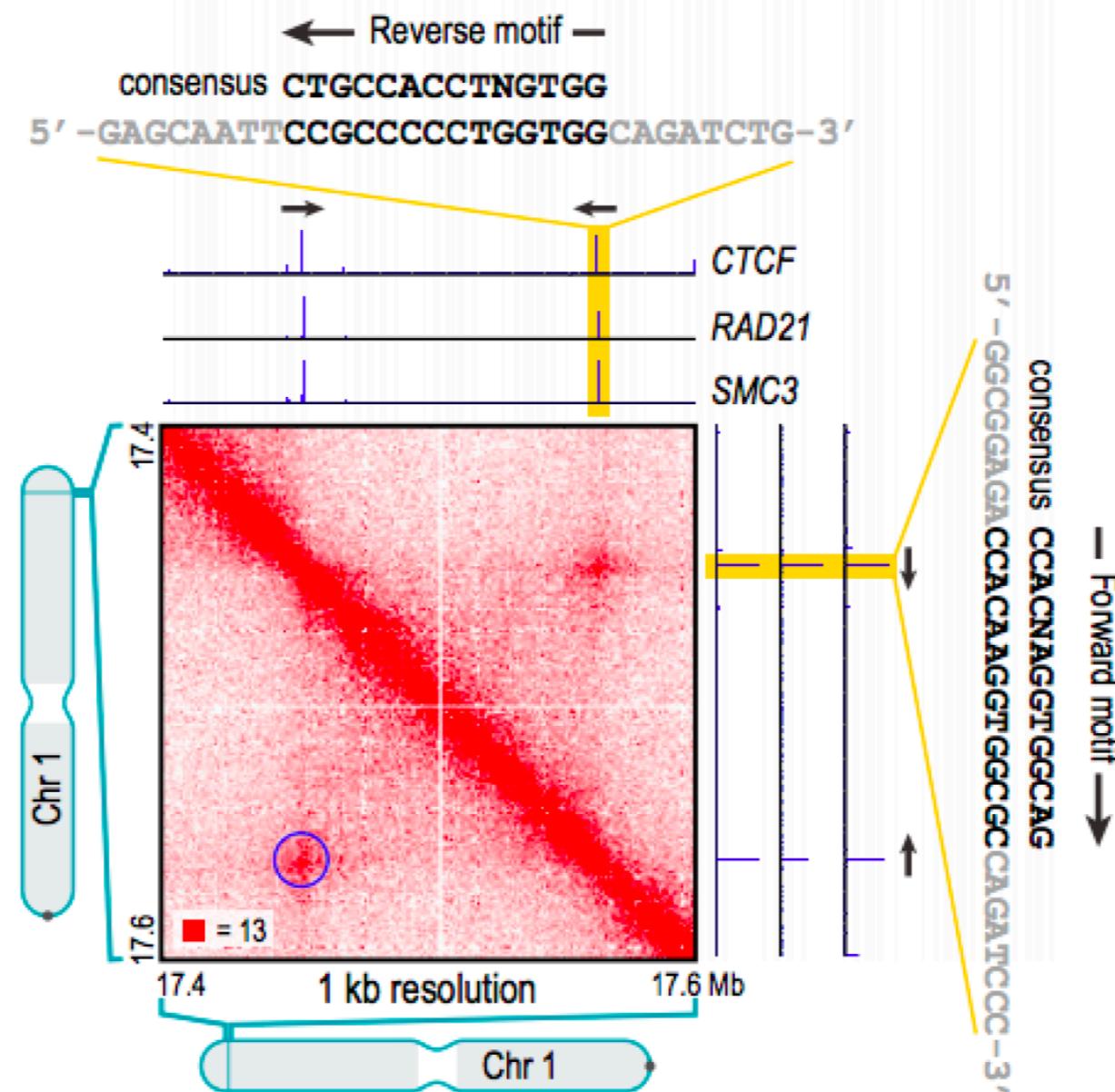
- Eigenvector decomposition:

3

- Eigenvector expansion (PCA, principal component analysis)

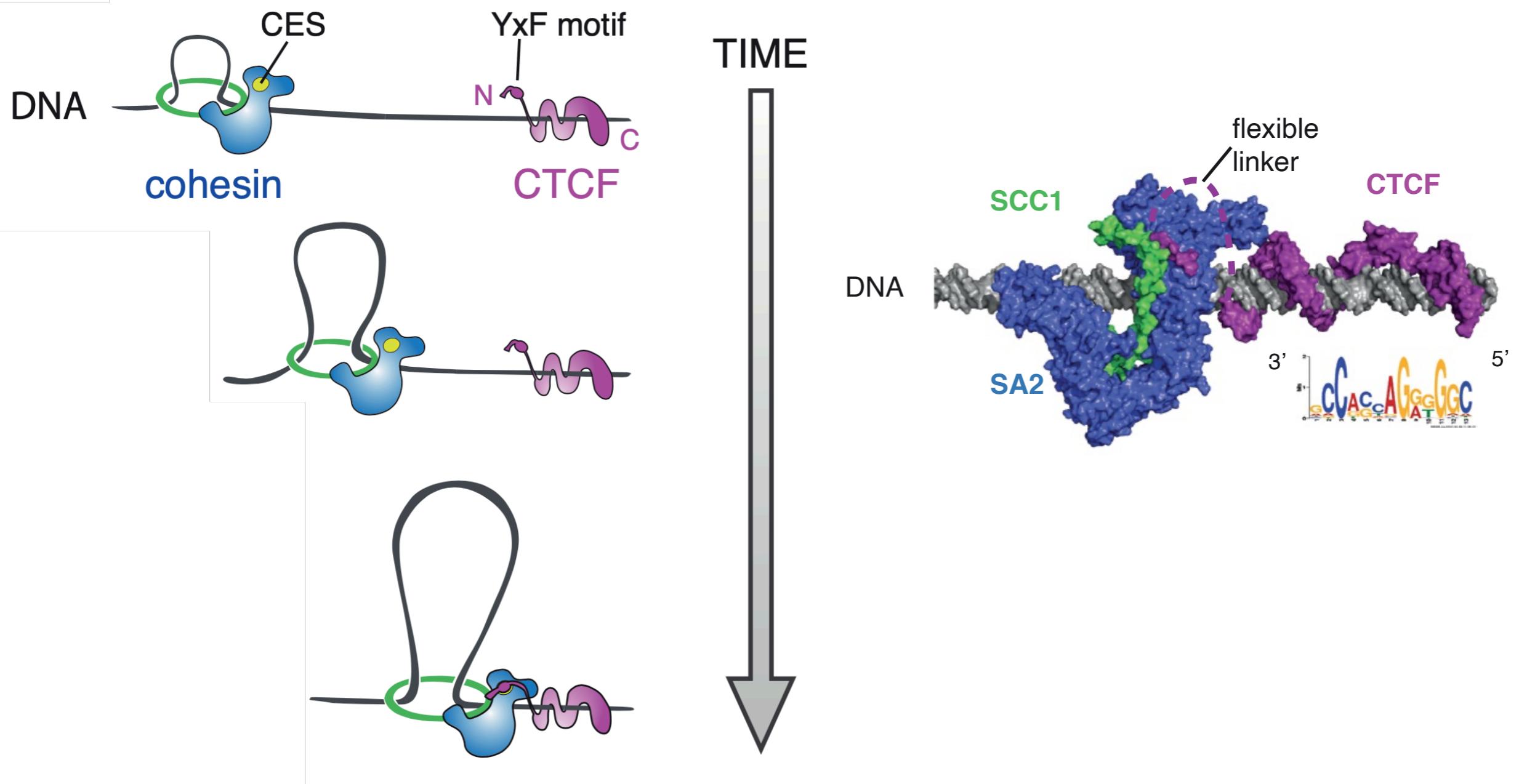


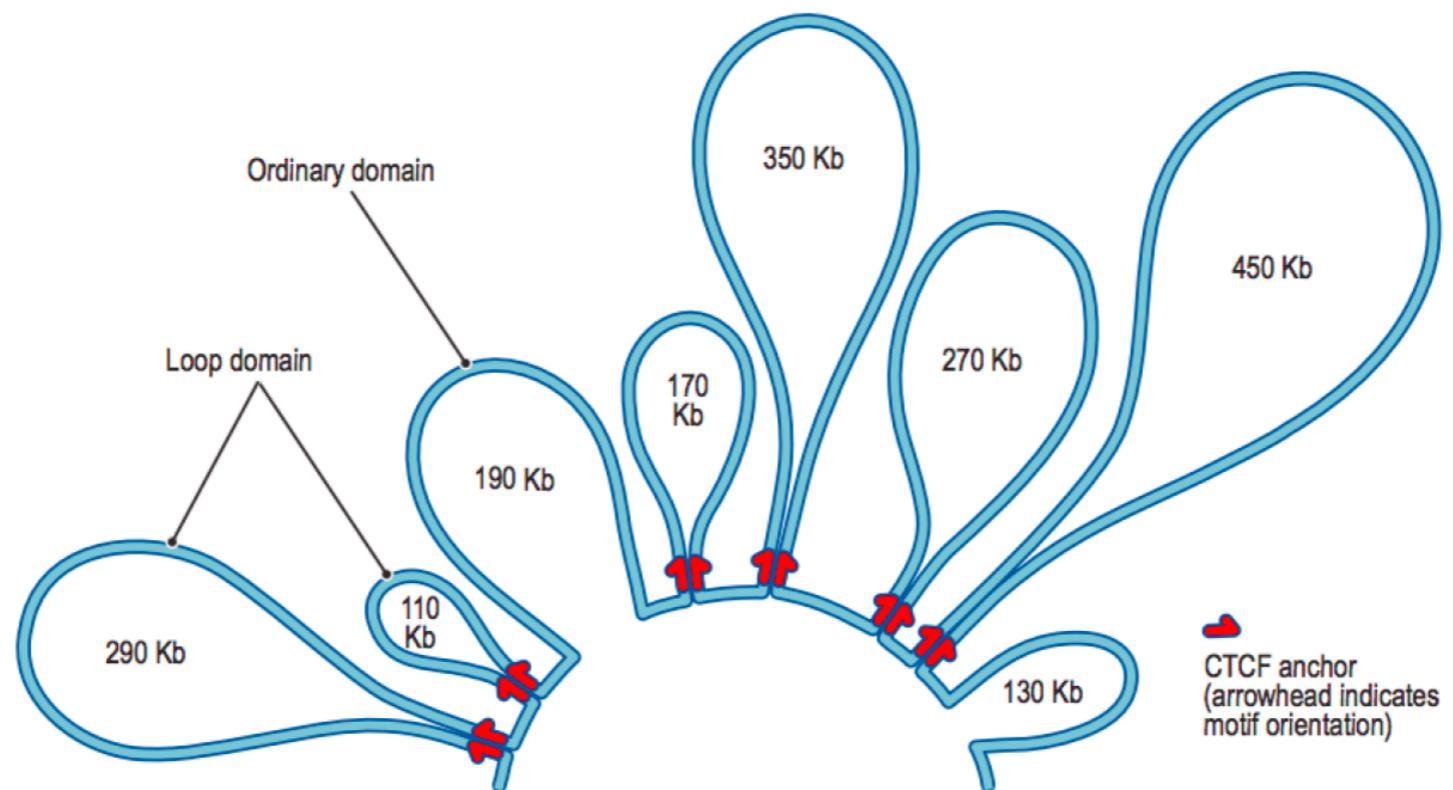
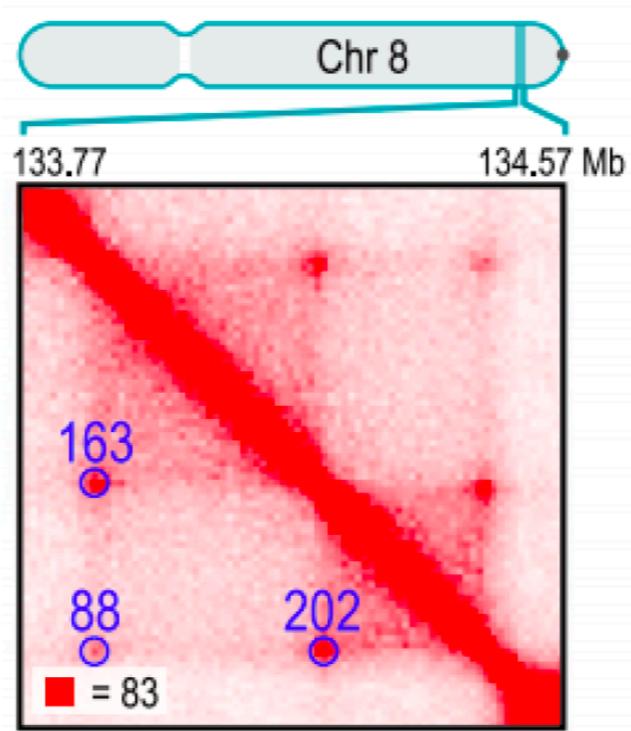
# Обогащенные взаимодействия Ні-С



# Экструзия петель

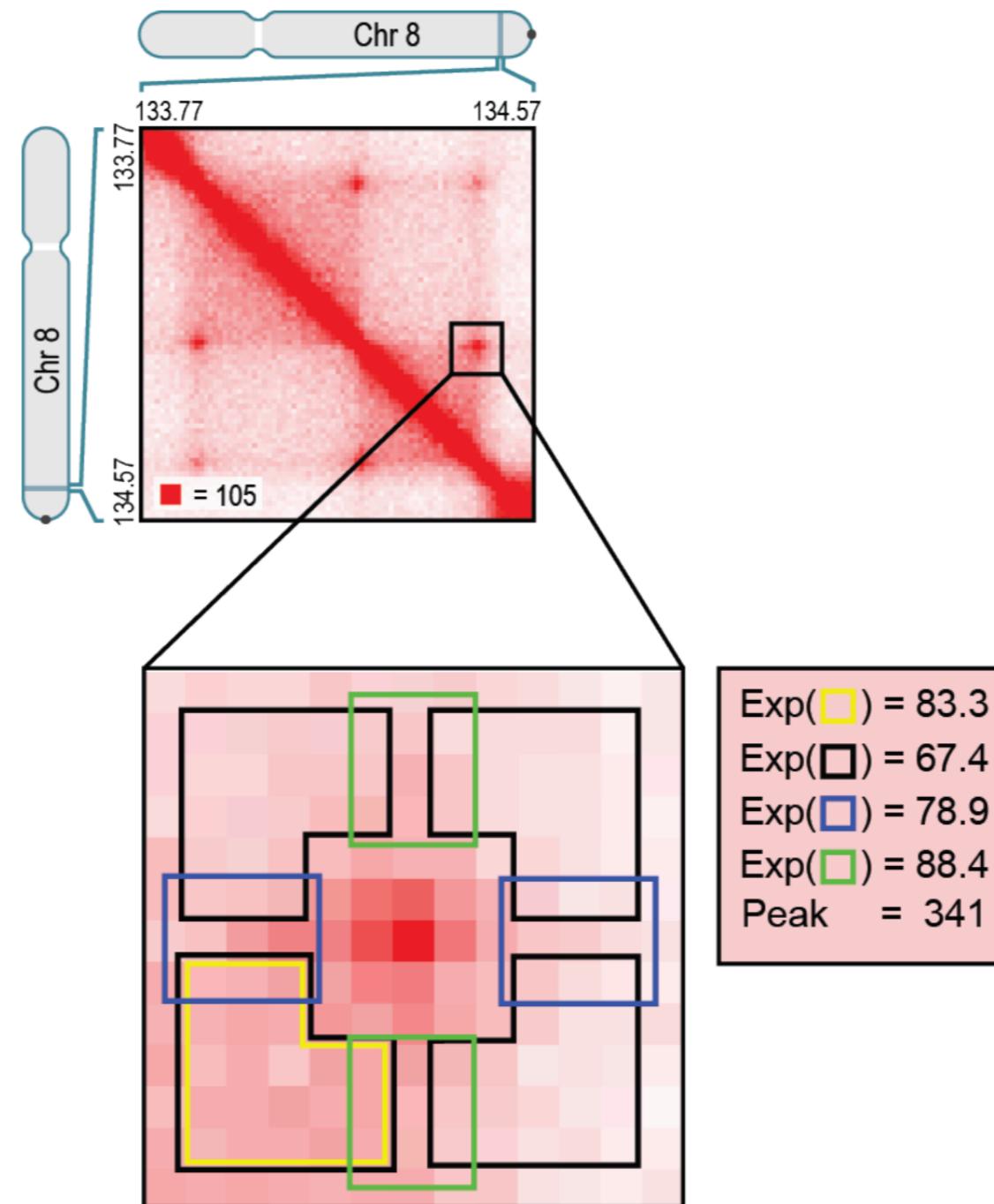
- Механизм образования петель хроматина:
- CTCF - граничный элемент
- когезин - экструдер, выпячивающий белок





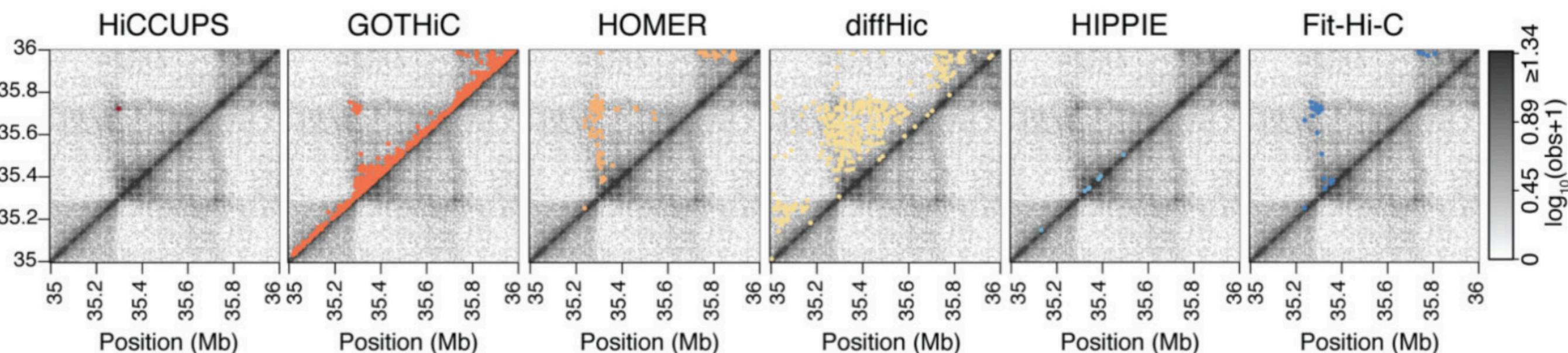
# Поиск обогащенных взаимодействий: HiCCUPS

- Hi-C Computational Unbiased Peak Search:

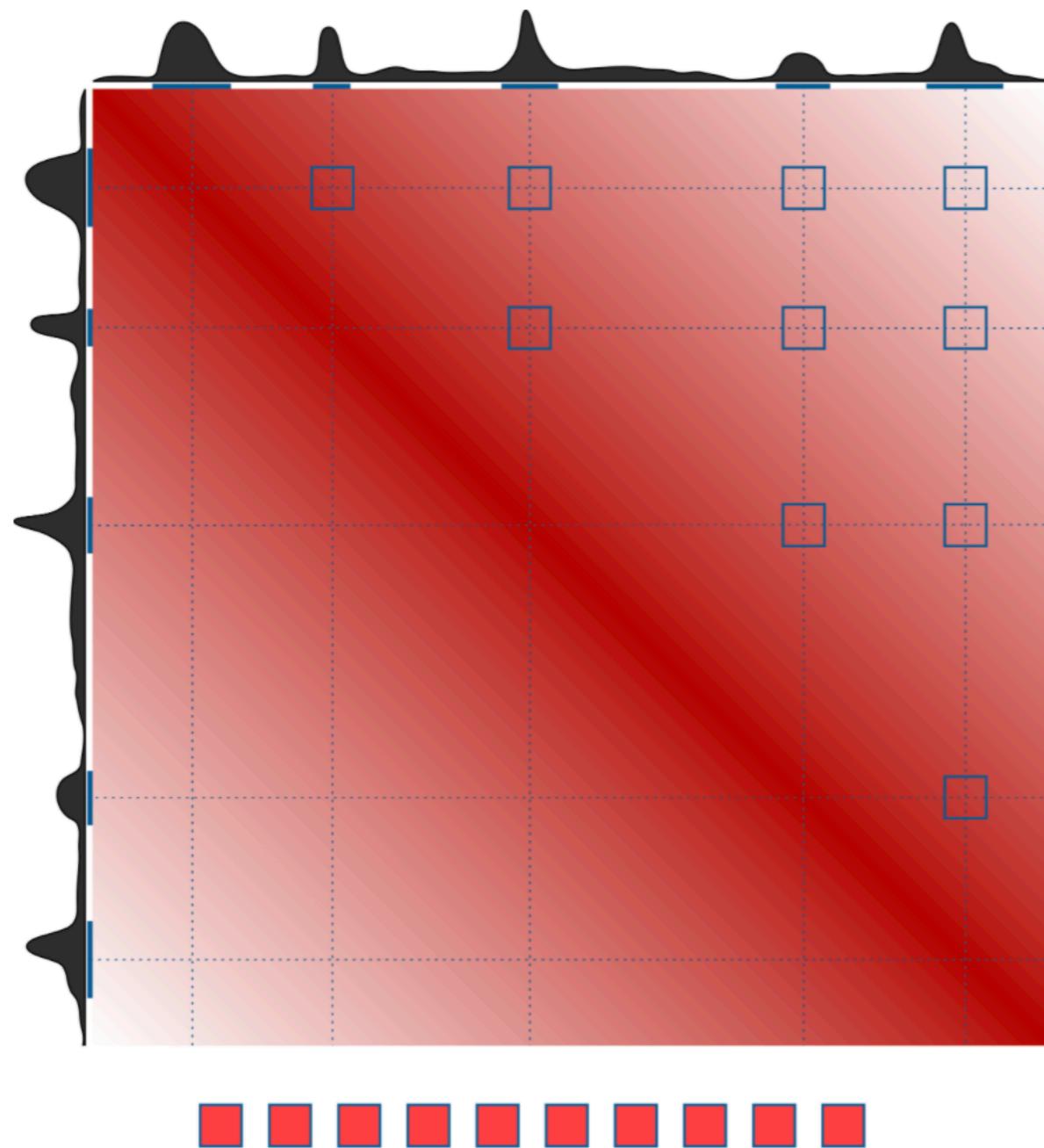


# Поиск обогащенных взаимодействий

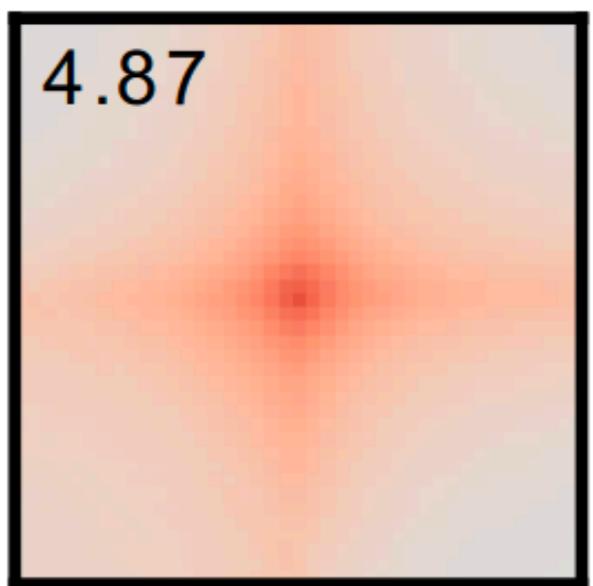
- С некоторыми отличиями могут называться также:  
peaks/dots/loops calling, detection of enriched contacts.
- Сравнение разных алгоритмов:



# Построение усредненных карт



CTCF  
loops



# Variety of Hi-C processing tools

Hi-C tasks	Galaxy HiCExplorer	HiC-bench	HiFive *	Hi-Cpipe	HiCNorm	hiclib	HiTC	HOMER	Hi-Corrector	HiC-Pro	TADbit	HiCUP	.....
Web server	x		x										
Alignment	x	x	x	x		x				x	x	x	
Filtering	x	x	x	x		x		x		x	x	x	
Genome browser tracks	x	x											
Quality assessment plots	x	x	x	x			x	x		x		x	
Contact matrices	x	x	x	x		x		x		x			
Matrix correction	x	x	x		x	x	x	x	x	x	x		
Matrix comparison	x	x											
Boundary scores	x	x											
Domains	x	x									x		
Boundary comparison	x	x											
Specific interactions	x	x		x	x					x		x	
Distance vs. counts	x						x						
Correlation of samples	x												
A/B compartments	x						x						
Annotations		x					x	x					
Allele-specific interactions										x		x	
Visualization	x	x	x	x			x	x					
Integration with ChIP-seq data	x	x					x						
Parallelization	x	x	x	x			x	x	x	x	x		
Integration of alternative tools	x	x											
Parameter exploration		x											
Reproducibility	x	x	x										
Import export different file formats	x												

# Пример анализа данных Ні-С

# Think twice before mapping raw Hi-C data...

- Готовые и простые в использовании базы данных и браузеры Hi-C.
- Огромное количество подходов и утилит для обработки Hi-C, нет золотого стандарта:

	Language	Year
<b>HOMER</b>	Perl, R	2010
<b>hiclib</b>	Python	2012
<b>Fit-Hi-C</b>	Python	2014
<b>HiCCUPS / Juicer</b>	Java	2014, 2016
<b>GOTHIC</b>	R	2015
<b>HIPPIE</b>	Python, Perl, R	2015
<b>diffHic</b>	R, Python	2015
<b>Juicer</b>	Java	2016
<b>TADbit</b>	Python	2017
<b>distiller</b>	nextflow	not published, but actively used

- Зачастую одна статья - один новый метод для обработки Hi-C
- Иногда проще написать собственное картирование и обработку Hi-C, чем воспроизвести анализ из статьи...

# Пример пайплайна обработки Hi-C:

## 1. Картирование

- Tools : **bwa, samtools**
- Output: **bam** file

## 2. Фильтрация, сортировка, создание списка контактов

- Tools: **pairtools**
- Output: **pairs** file

## 3. Бинирование и нормализация

- Tools: **cooler**
- Output: **cool** file

## 4. Агрегация

- Tools: **cooler**
- Output: multires.cool (**mcool**) file

- Все это сейчас делает **distiller-nf**
- Для более полной информации, см. немного устаревший  
[Hi-C data analysis bootcamp @ Harvard Medical School 2018](#)

# Distiller

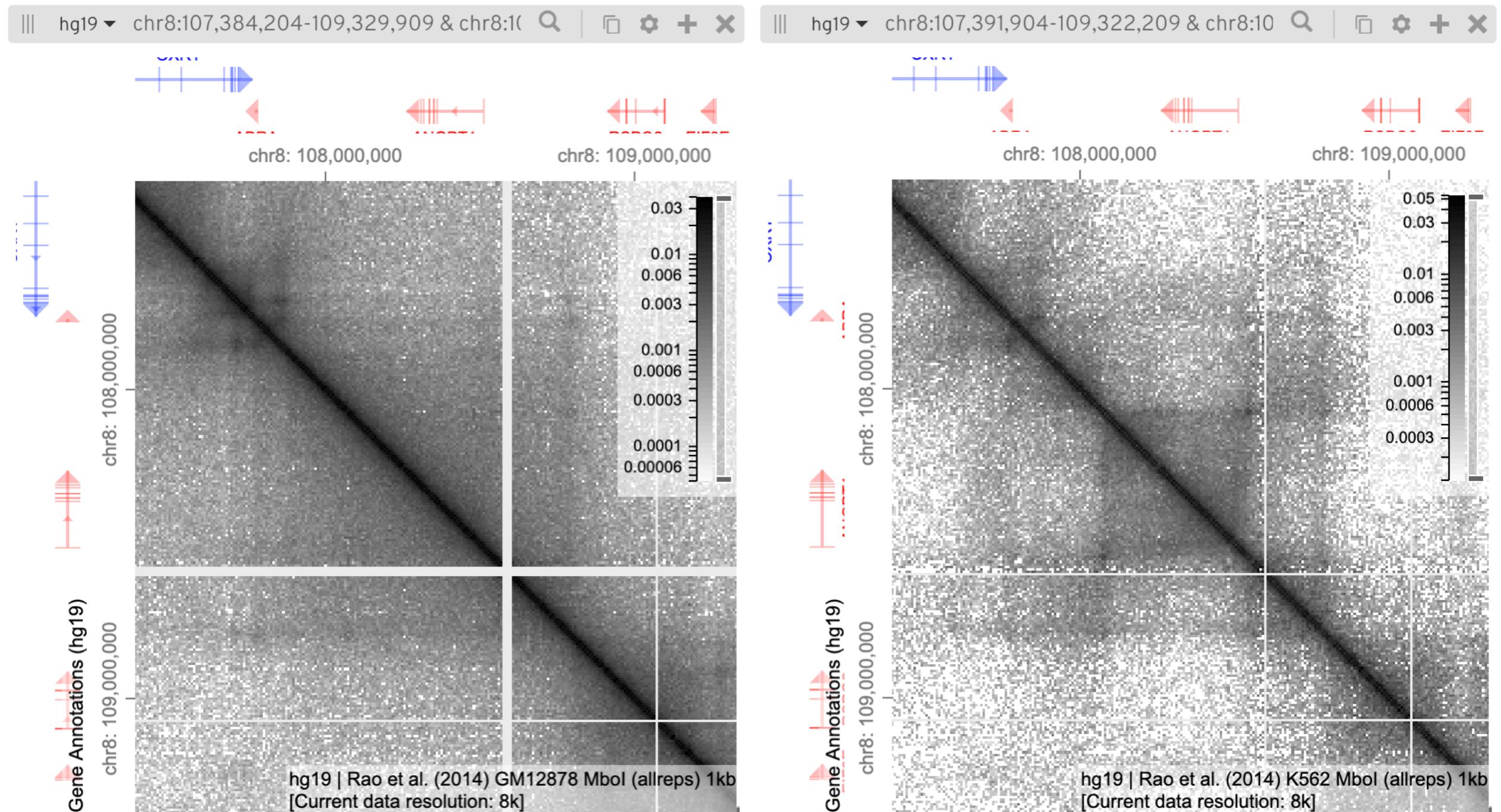
- <https://github.com/mirnylab/distiller-nf>



- Основанный на nextflow пайплайн обработки Hi-C данных,
- Позволяет получить *cool*-файлы с картами взаимодействий и *mcool*-файлы для визуализации в HiGlass.

# HiGlass: браузер для просмотра карт и многое другого

- <http://higlass.io/>
- Пример вида: <http://higlass.io/l/?d=VKtF8avOQD6mpe6Trzcj-A>



# Другие браузеры карт:

- Вместо локального **HiGlass** можно использовать [Galaxy-версию HiGlass](#)
- **Juicebox** from Aiden lab: <http://www.aidenlab.org/juicebox/>
- Карты Hi-C теперь поддерживаются **UCSC**: <https://genome.ucsc.edu/>
- Браузеры одной лаборатории:
  - Dekker Lab browser: <http://hic.umassmed.edu/heatmap/heatmap.php>
  - YUE lab browser: <http://rgromoter.bx.psu.edu/hi-c/>
- И многие другие...

# Специальные случаи

1. Применение Hi-C для сборки геномов
2. Single-cell Hi-C
3. РНК-ДНК взаимодействия

# 1. Ні-С и биоразнообразие

- Первые эксперименты Ні-С (2009-2012): человек, мышь и *Drosophila*
- Вторая волна Ні-С: другие модельные организмы, например *Caenorhabditis, Arabidopsis, Saccharomyces, Escherichia*
- Ні-С на экзотических видах: комар, археи, *Plasmodium falciparum* и многие другие (e.g., DNA Zoo)

# 1. Ні-С и биоразнообразие



Cheetah



Red panda



American alligator



Chinese alligator



California sea hare



Golden eagle



Peanut



Hog deer



Bryde's whale



American bison



Silkmoth



Field mustard



Dromedary



Dingo



Domestic dog, Go...



German Shepherd...



Canadian beaver



Southern cassowary



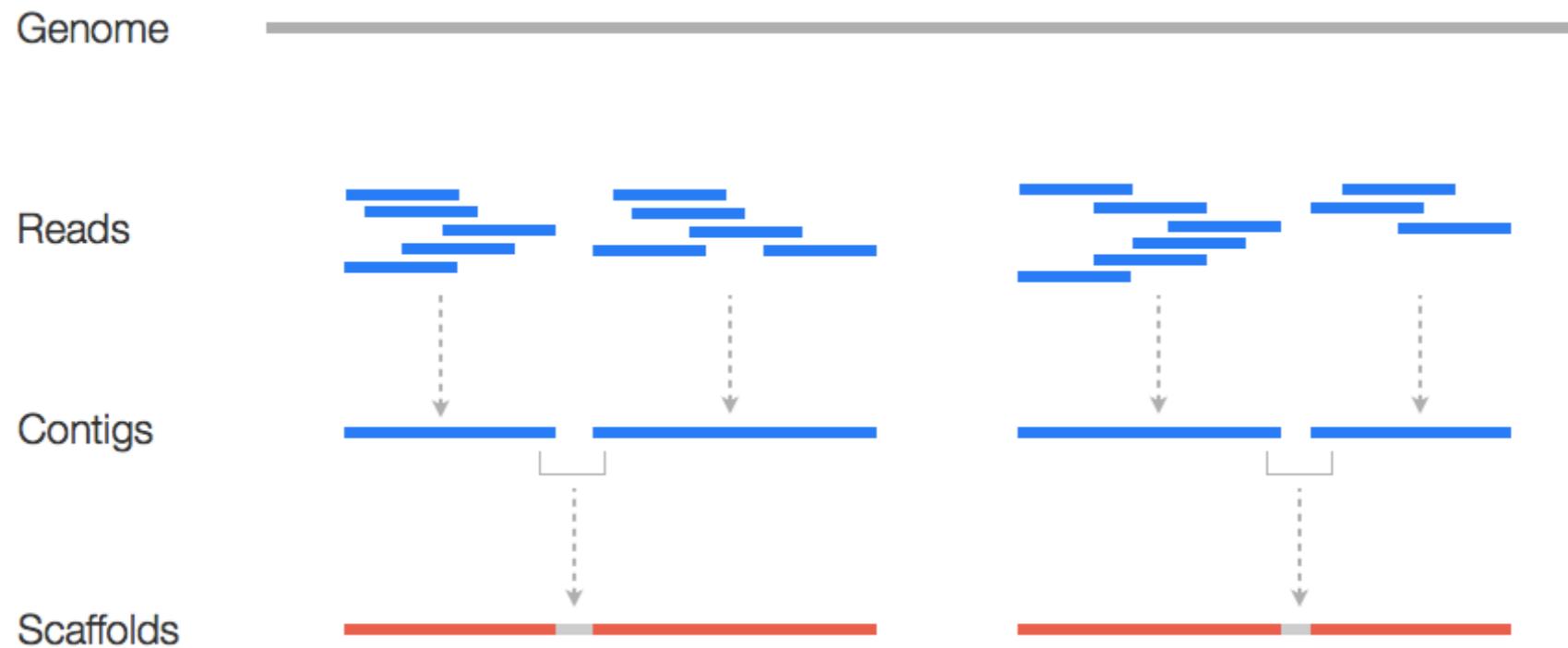
Domestic guinea pig



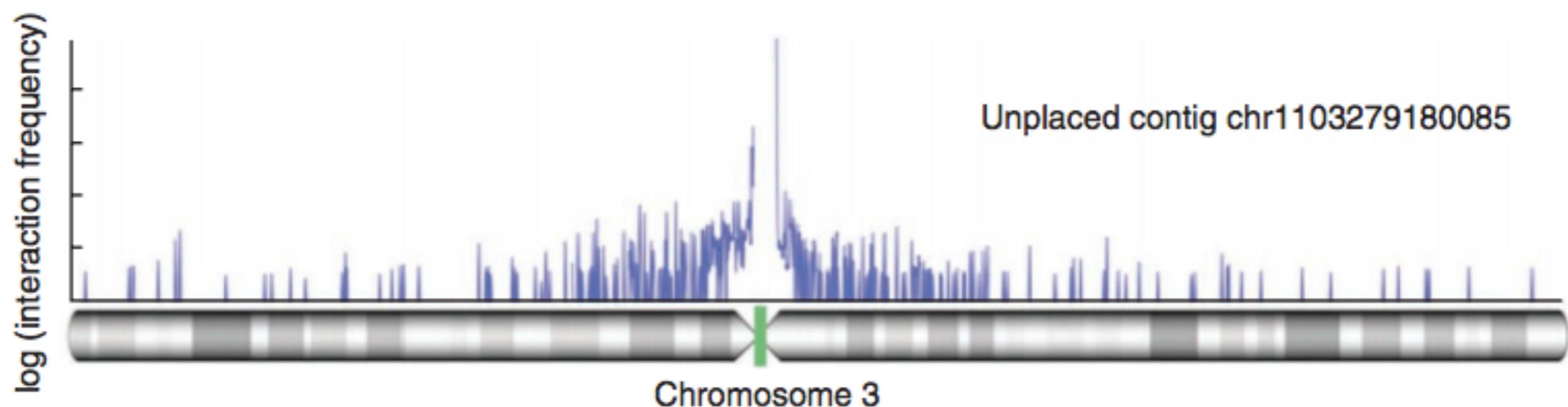
Pygmy marmoset

# 1. De novo сборка геномов

- Типичный подход к сборке геномов:

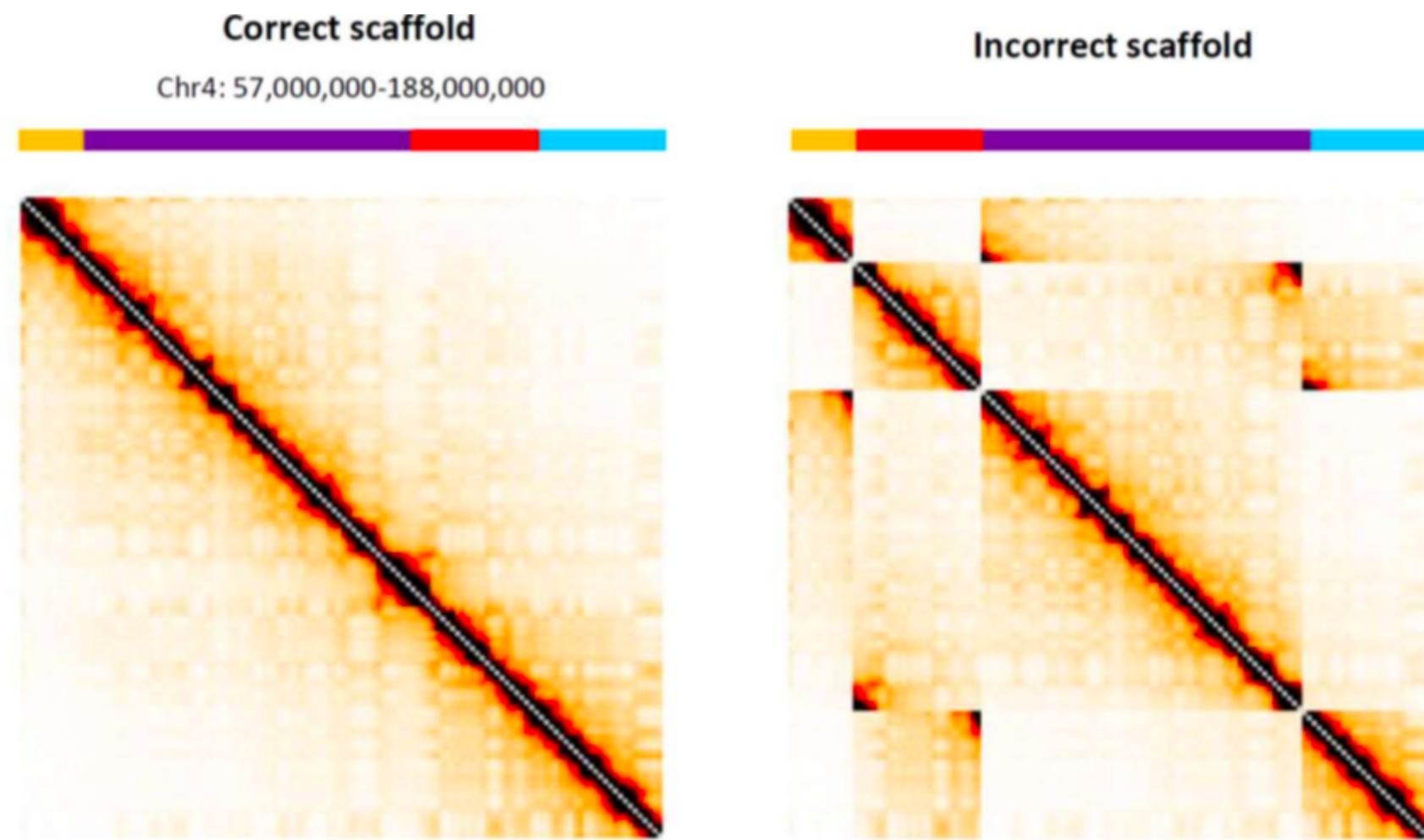


- С помощью Hi-C:



# 1. Ні-С для экзотических видов, неправильная сборка

- Ні-С содержит информацию о пространственной и, косвенно, линейной близости участков ДНК.
- Пример картирования Ні-С на две сборки генома некоторого млекопитающего:



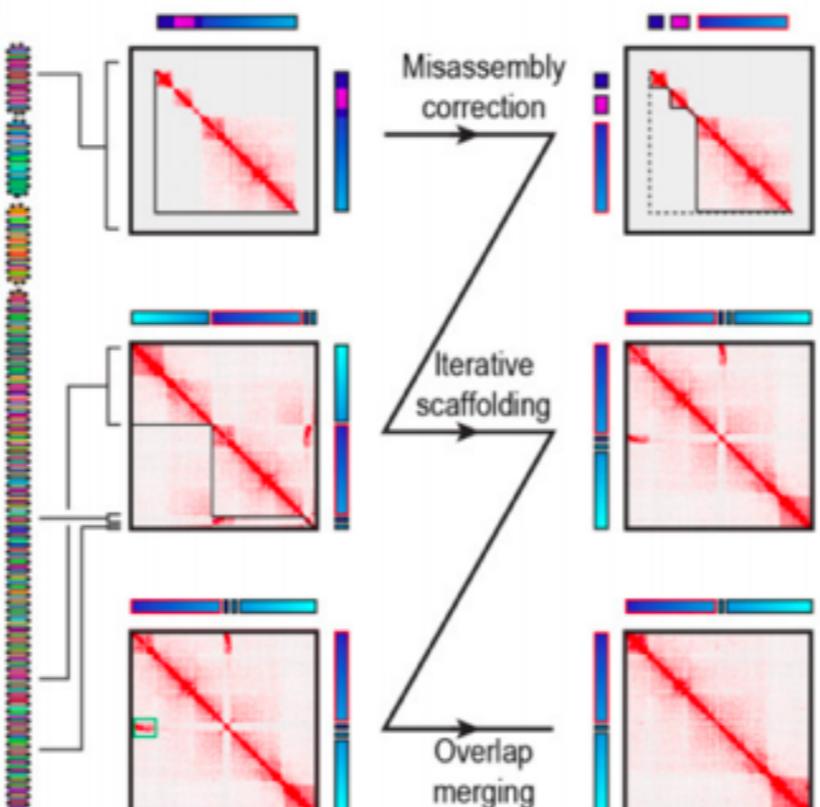
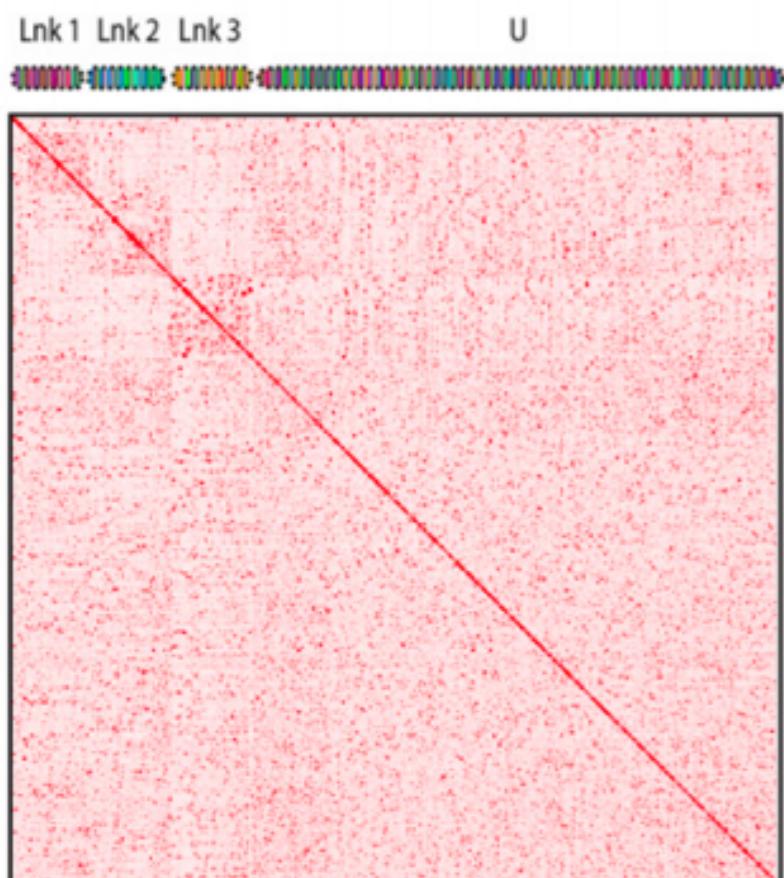
# 1. Ні-С экзотических видов

- Пример полной сборки:

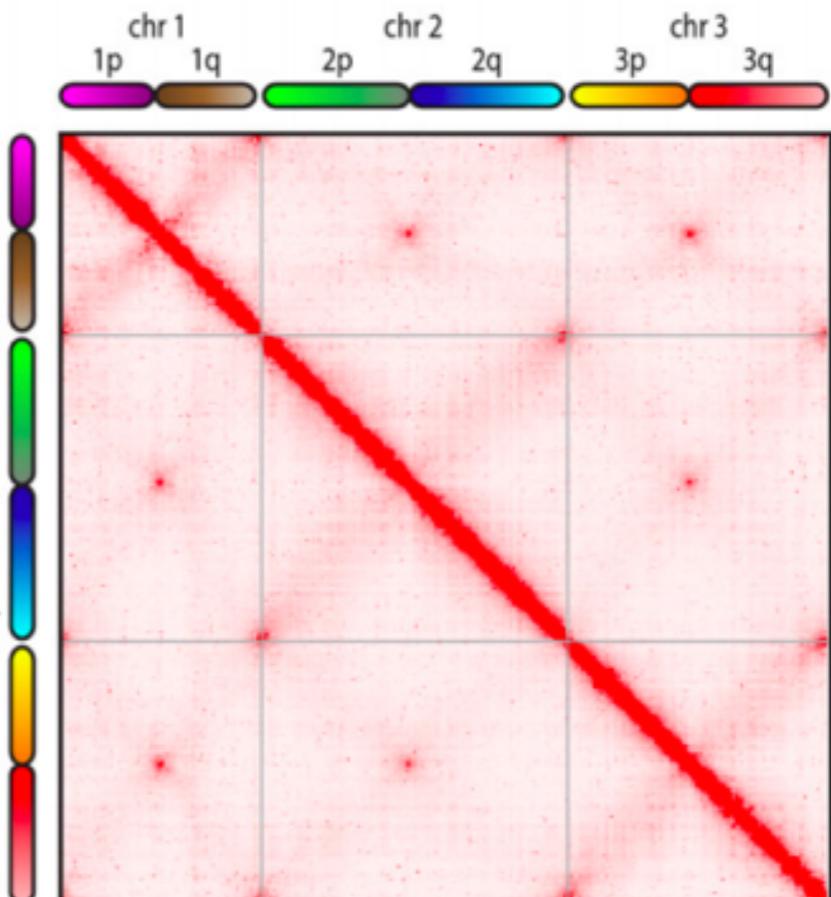
*Aedes aegypti* (the yellow fever mosquito)



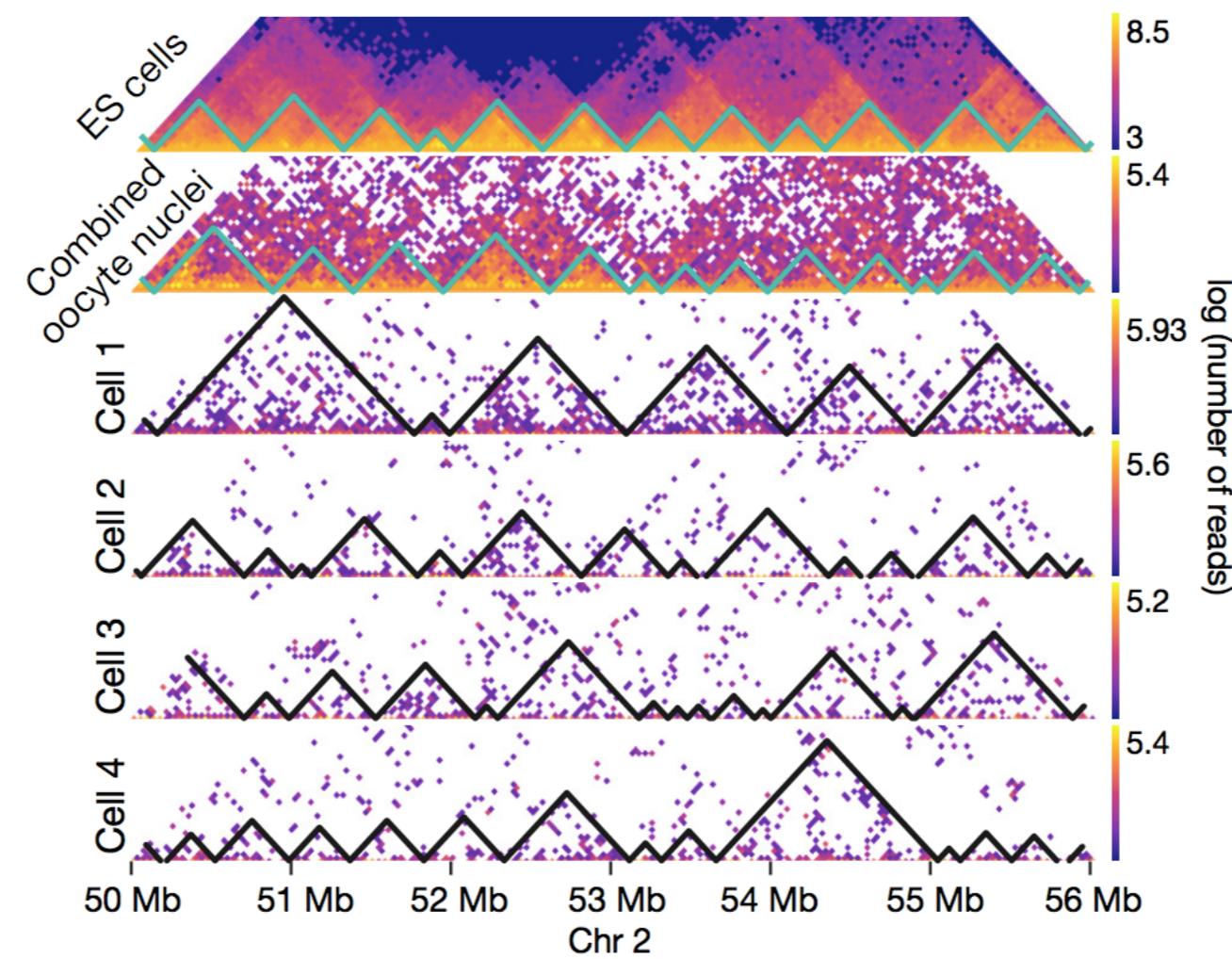
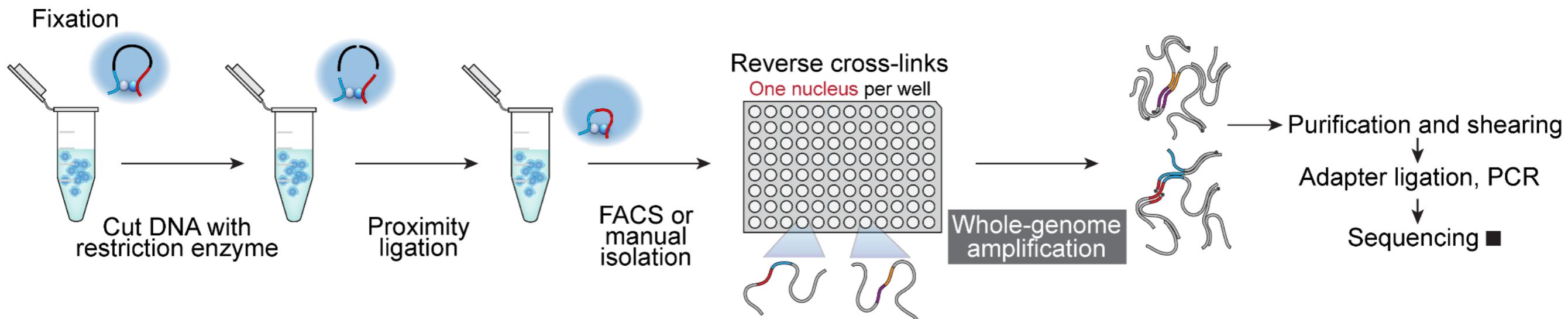
Draft assembly



End-to-end assembly

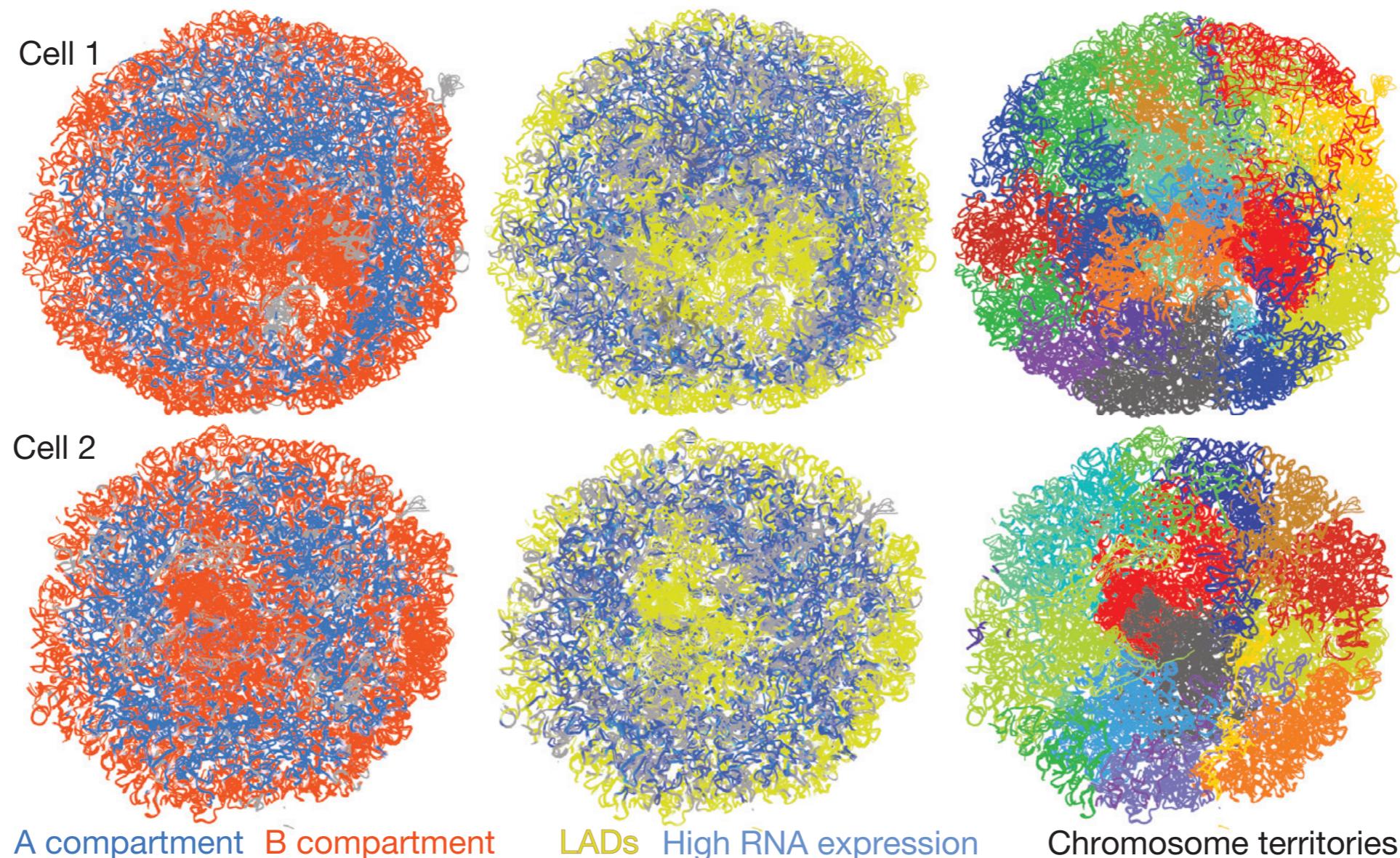


## 2. Single-cell Hi-C: Hi-C одиночных клеток



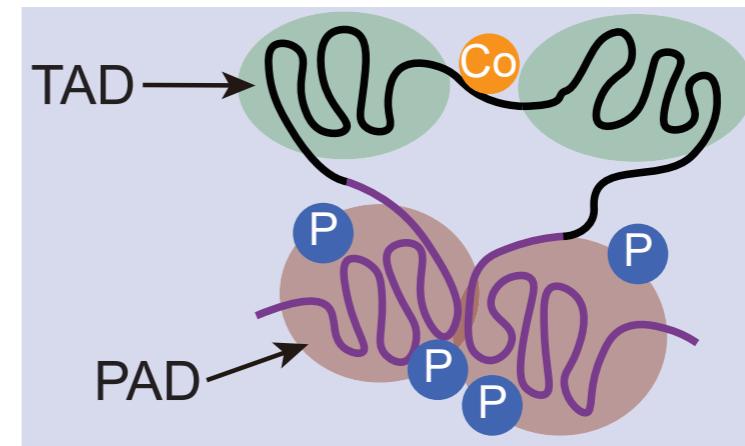
# Single-cell Hi-C: моделирование хромосом и ядра

Моделирование по single-cell Hi-C воспроизводит то, что мы наблюдали с помощью микроскопии:



# Основные механизмы формирования структуры хроматина:

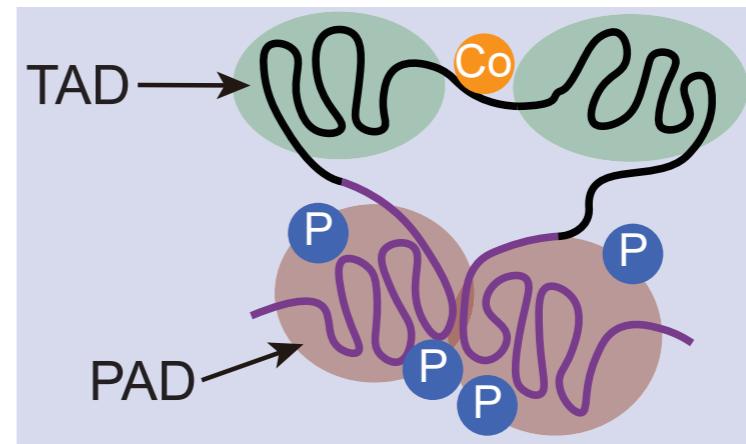
- **Polycomb-петлеобразование:**  
ДНК, связанная с белками комплекса Polycomb, образует агрегаты



# Основные механизмы формирования структуры хроматина:

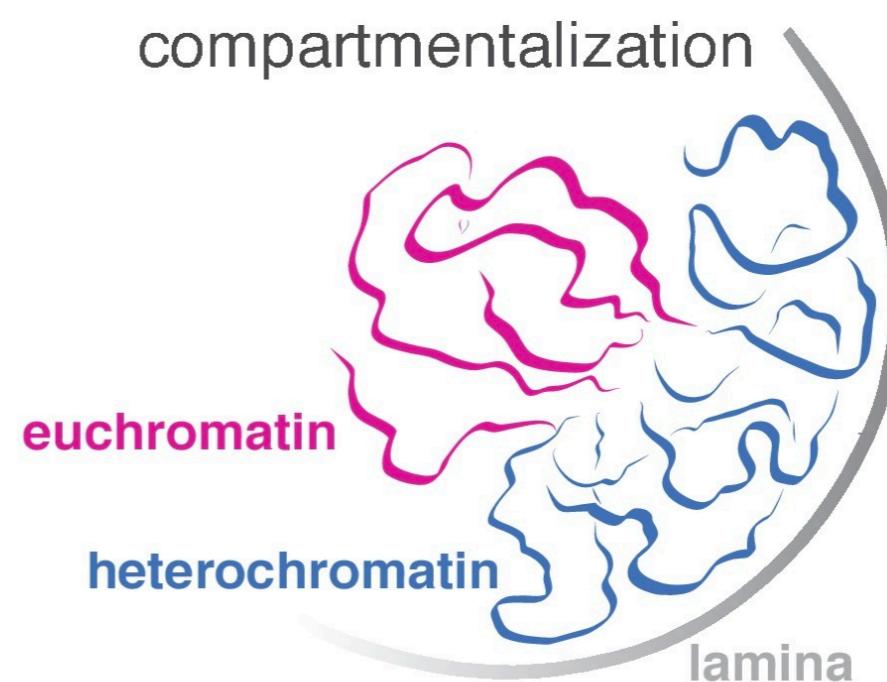
- **Polycomb-петлеобразование:**

ДНК, связанная с комплексом Polycomb, образует агрегаты



- **Комpartmentализация:**

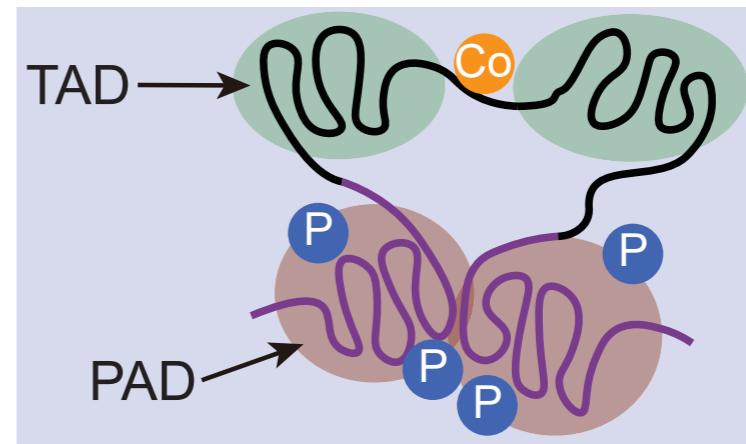
Транскрипционно активные регионы генома оказываются сближены в пространстве



# Основные механизмы формирования структуры хроматина:

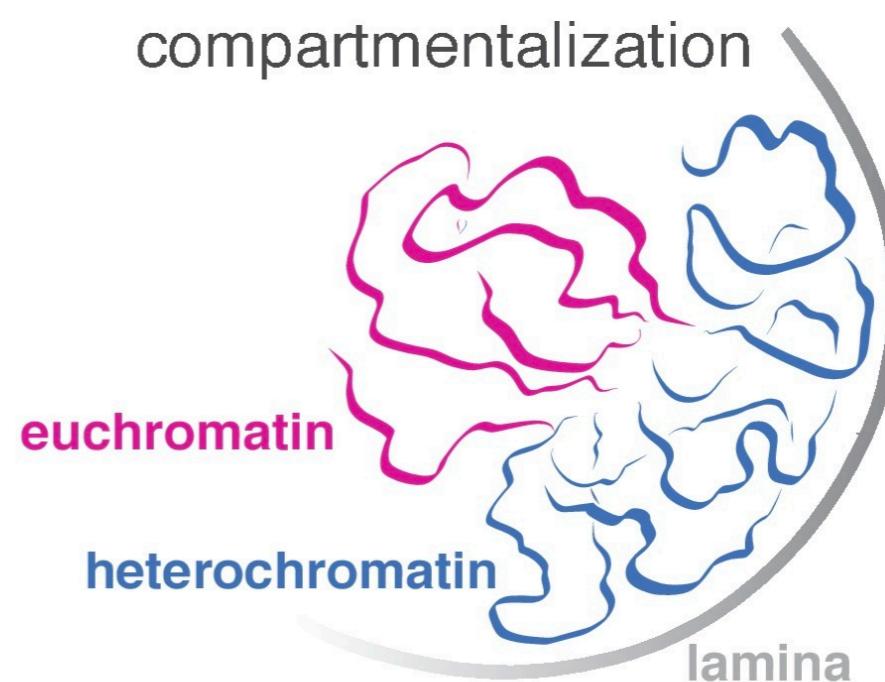
- **Polycomb-петлеобразование:**

ДНК, связанная с комплексом Polycomb, образует агрегаты



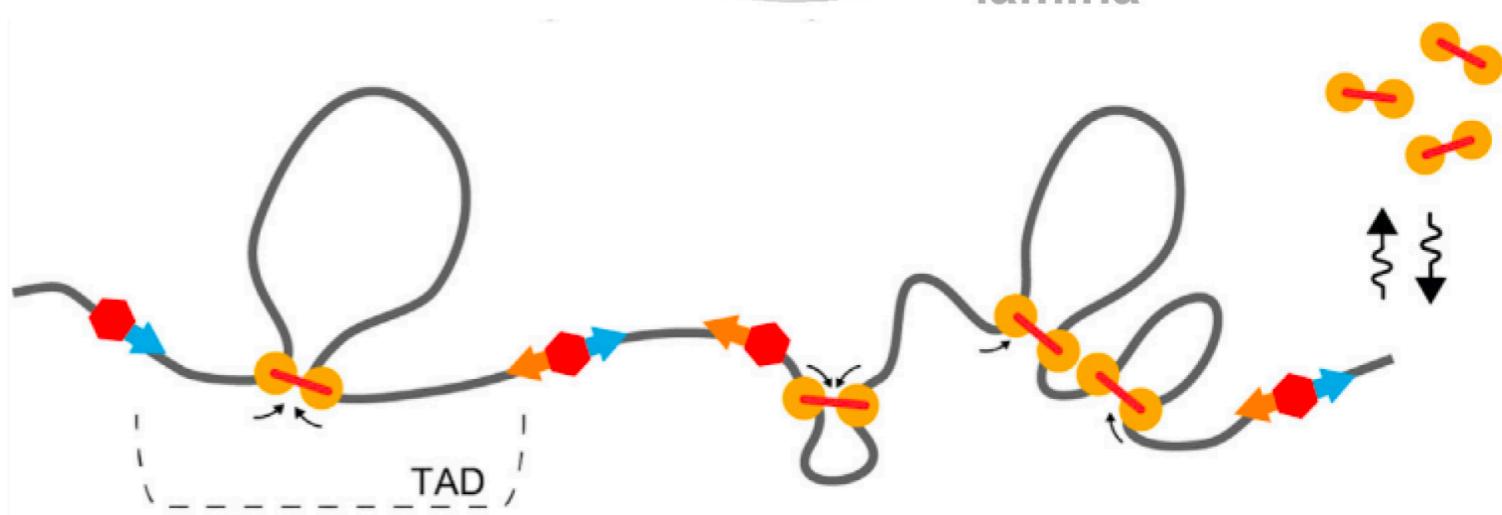
- **Комpartmentализация:**

Транскрипционно активные регионы генома оказываются сближены в пространстве



- **Экструзия петель:**

Локальный белок-экструдер (когезин) выпячивает ДНК

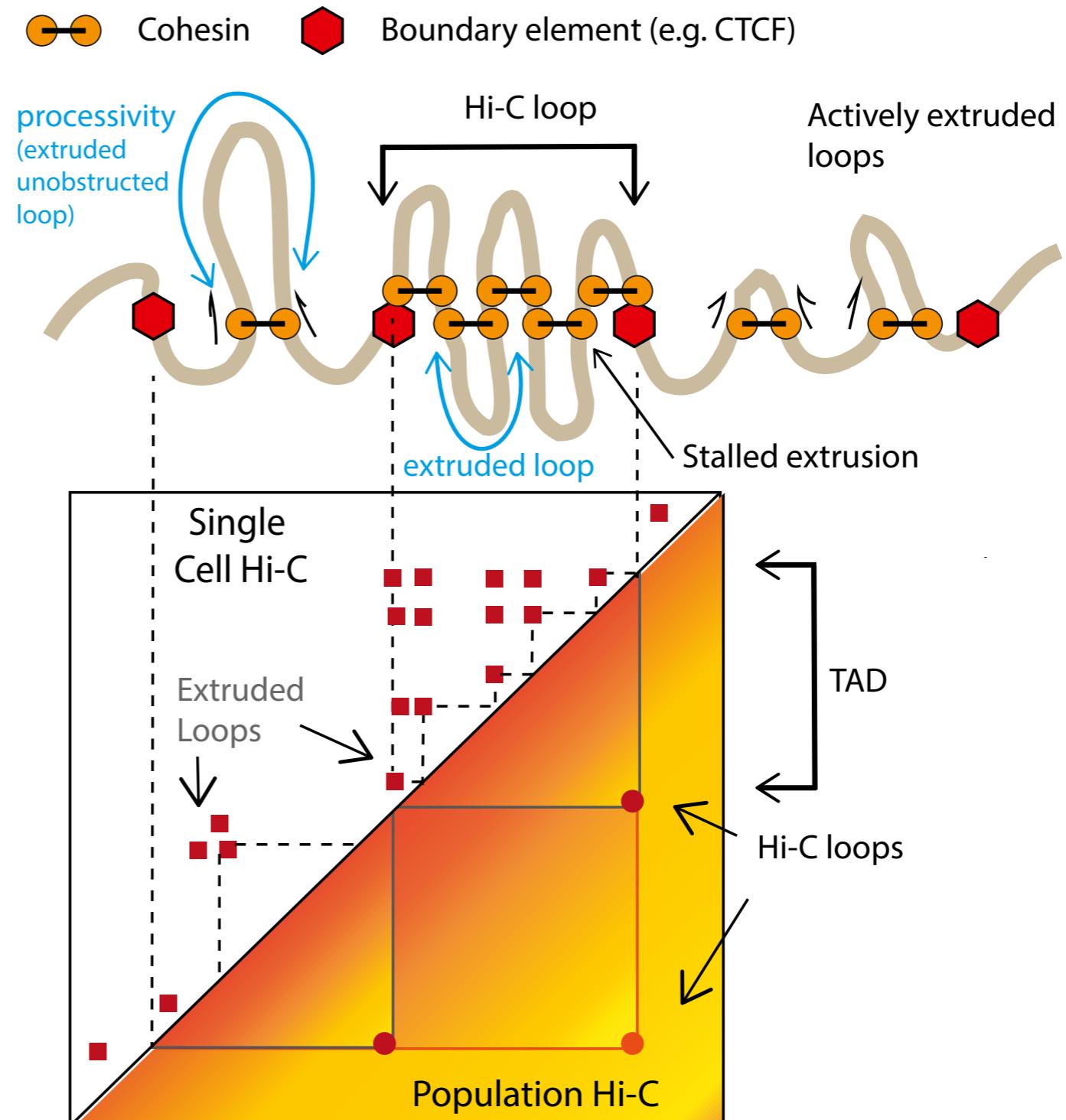


# Как формируются контакты Hi-C?

- Контакты наблюдаются как на уровне одиночных клеток, так и на уровне популяции клеток:

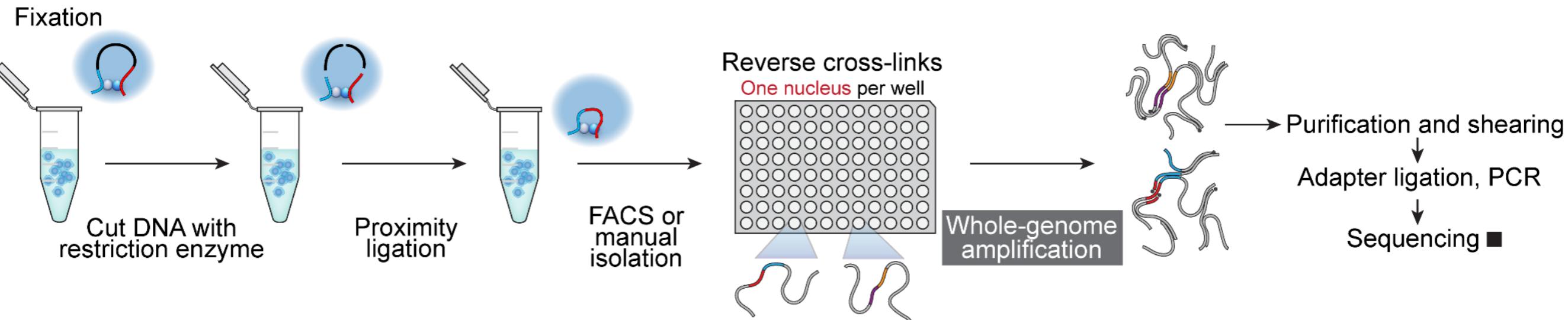
Экструзия петель как основной механизм петлеобразования:

Два типа Hi-C:

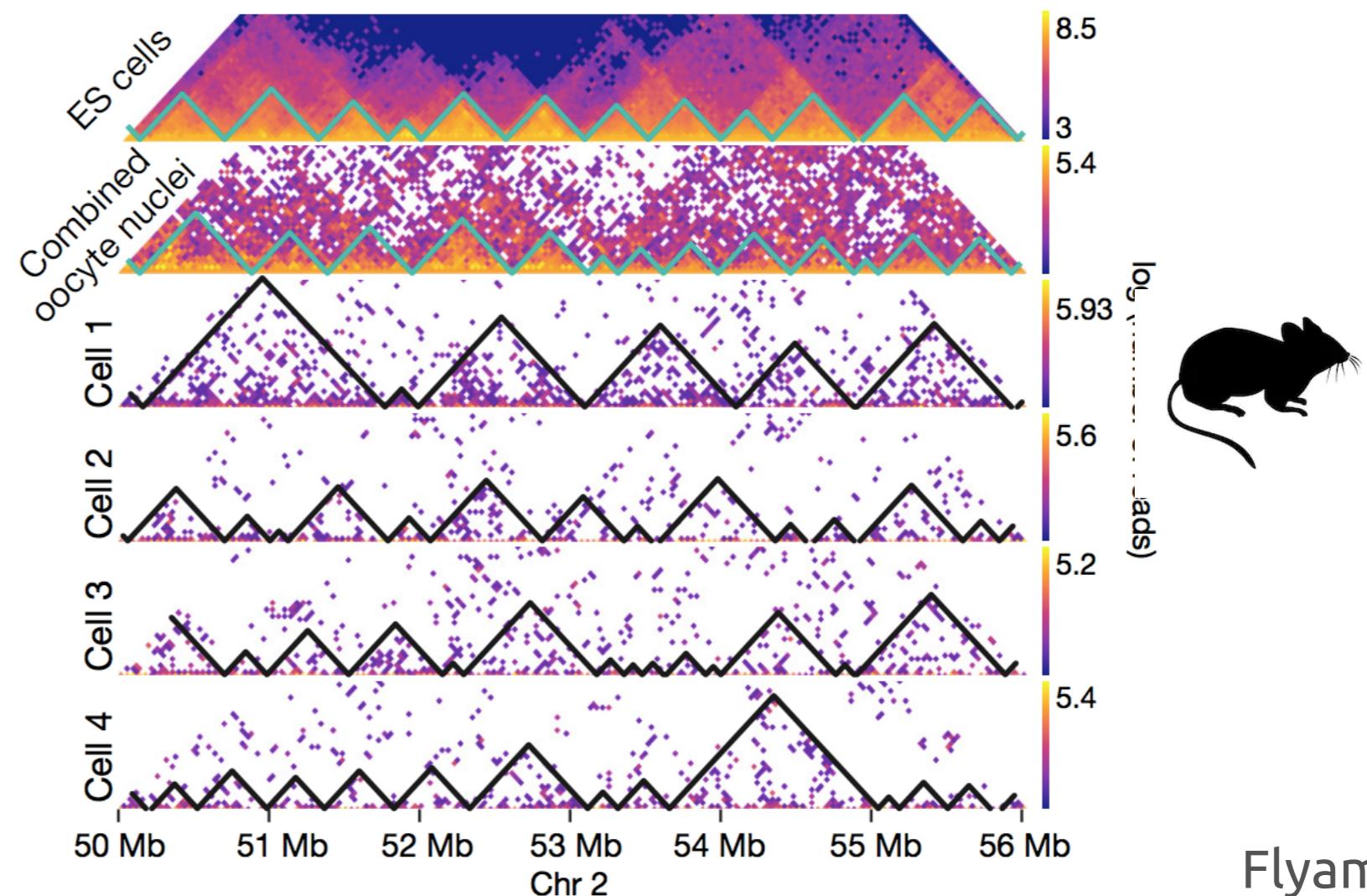


# Single-cell Hi-C: Hi-C одиночных клеток

- Single-cell Hi-C протокол (scHi-C, snHi-C):



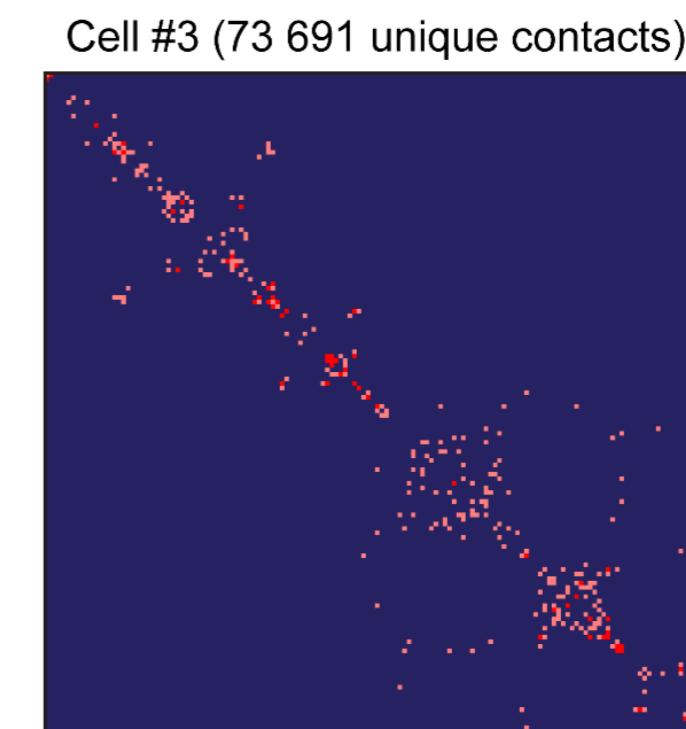
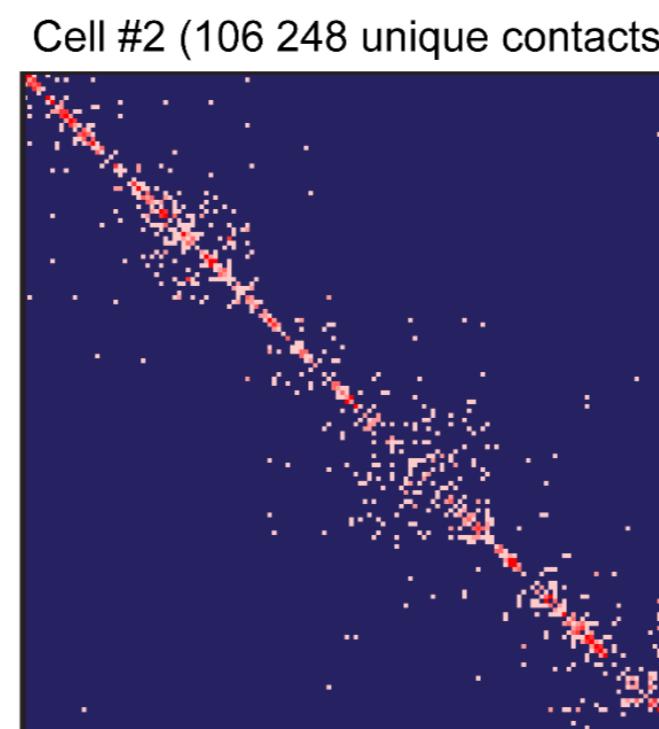
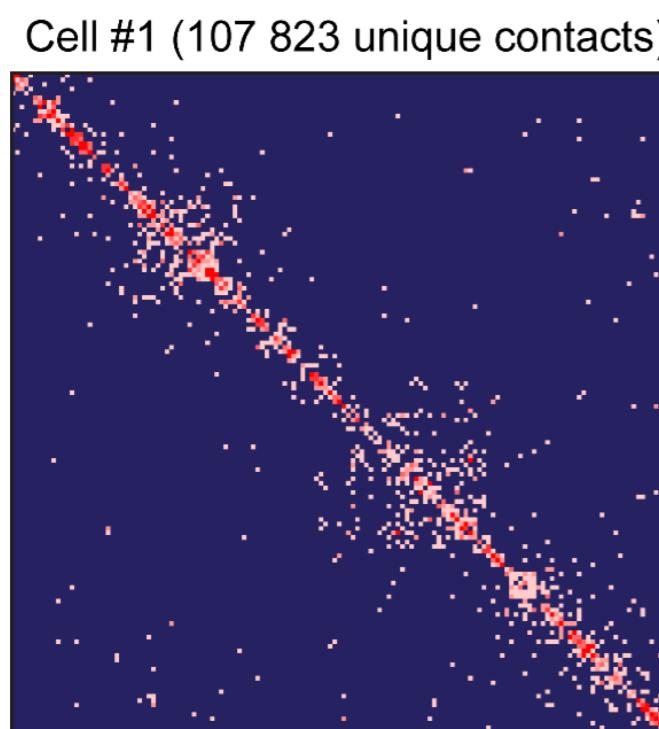
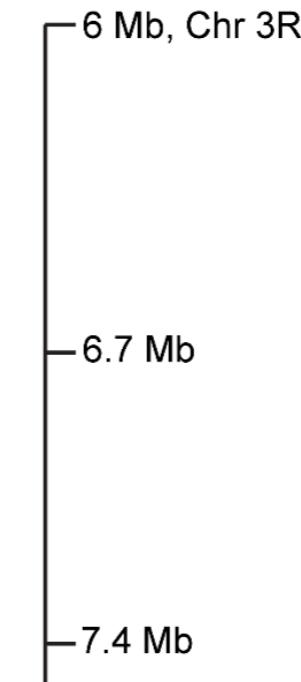
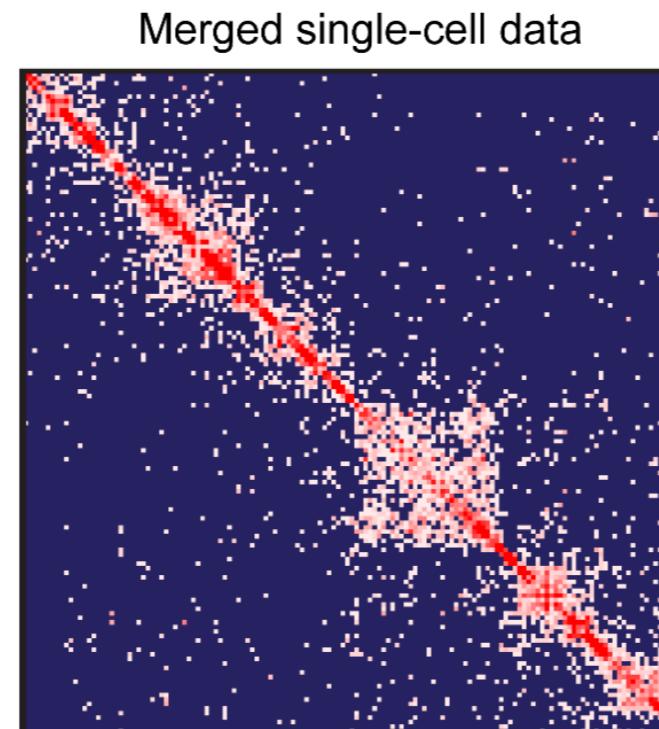
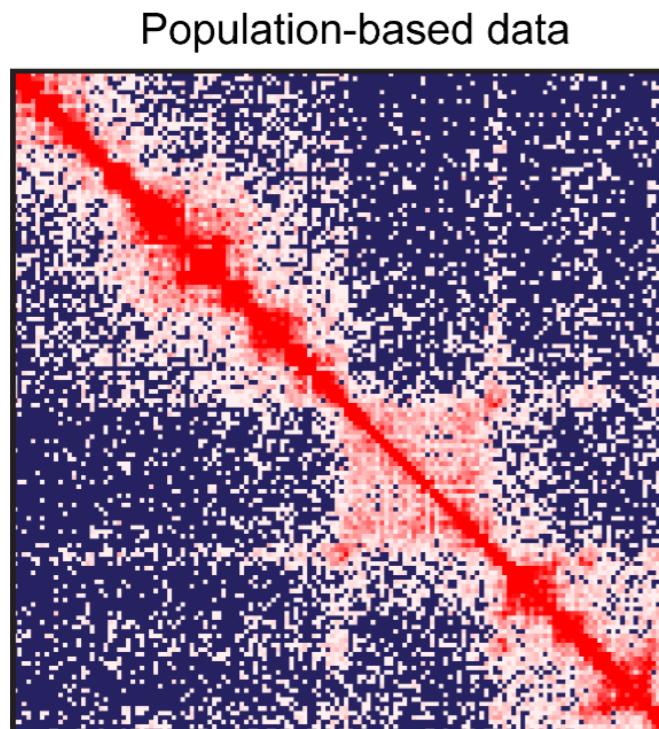
- Данные для мыши:



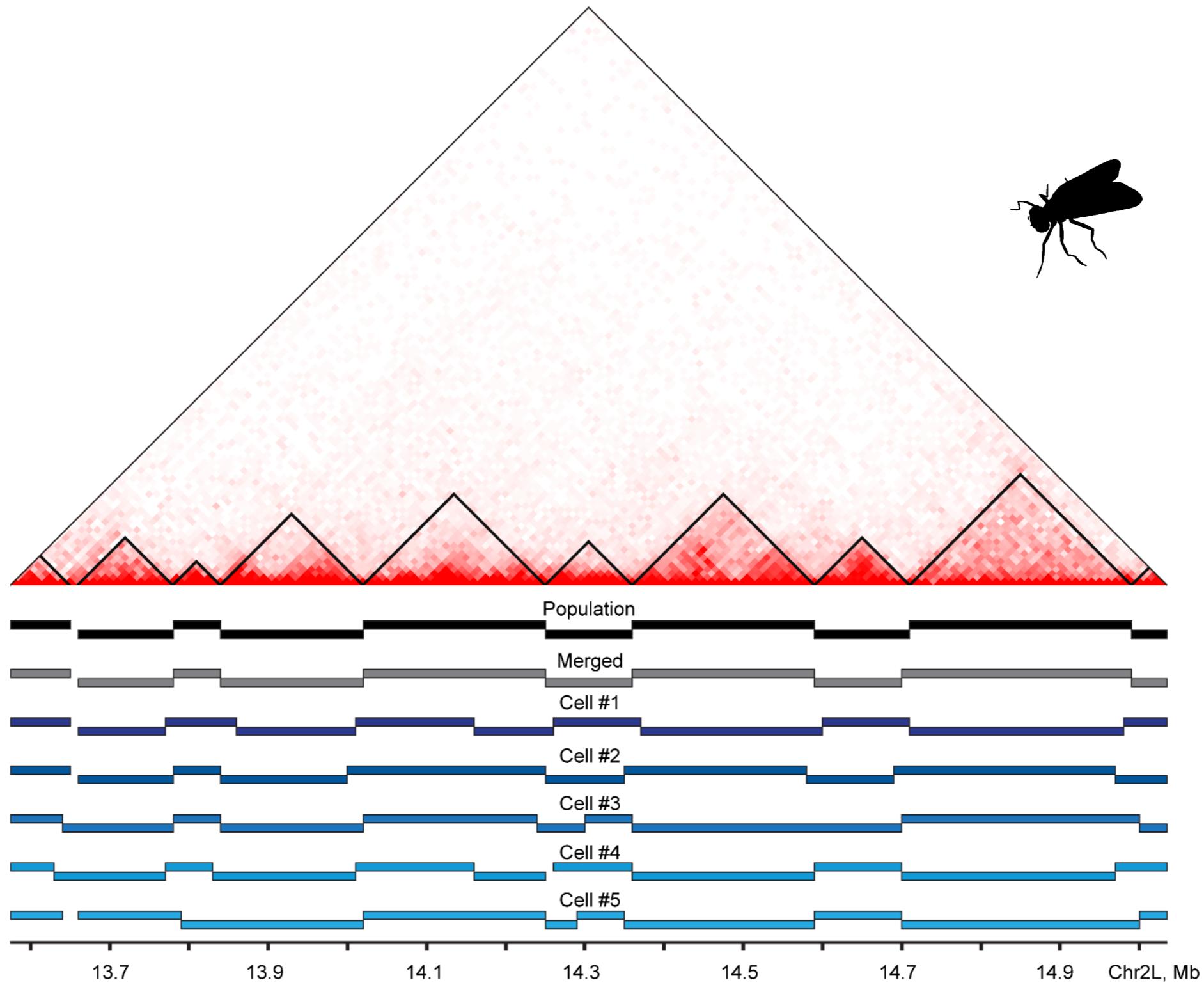
# Single-cell Hi-C для *Drosophila*

Work in progress

- ТАДы намного более стабильны в *Drosophila*:



- Границы ТАДов сохраняются в разных клетках *Drosophila*:



# Single-cell Hi-C: *Drosophila* и мышь

Work in progress

- В среднем, ~40% границ ТАДов совпадают между клетками муhi
- ~20-30% границ ТАДов совпадают между клетками мыши
- Возможно, в *Drosophila* действуют другие механизмы образования структуры, или смещен баланс между известными процессами:

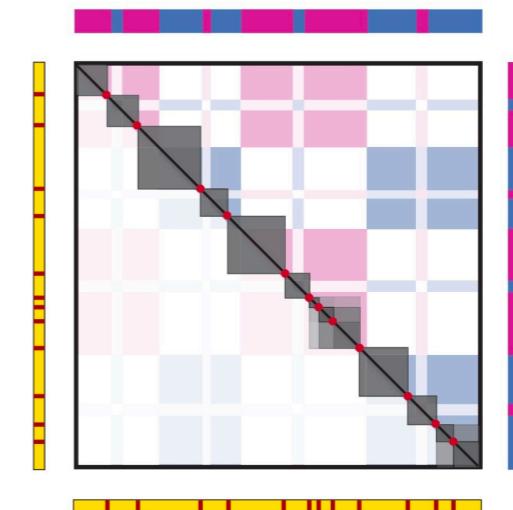
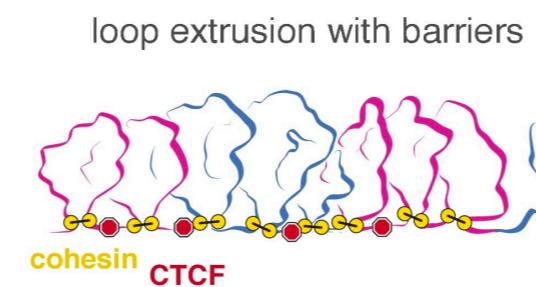
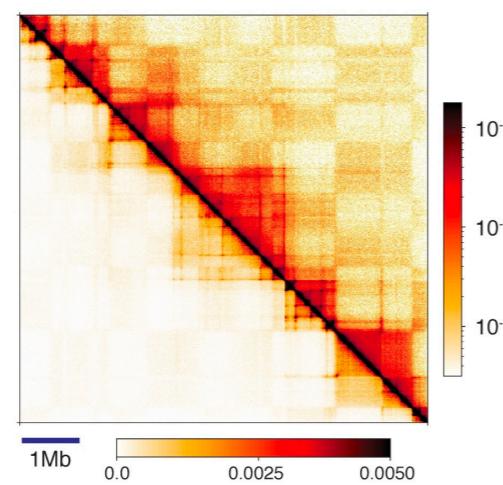
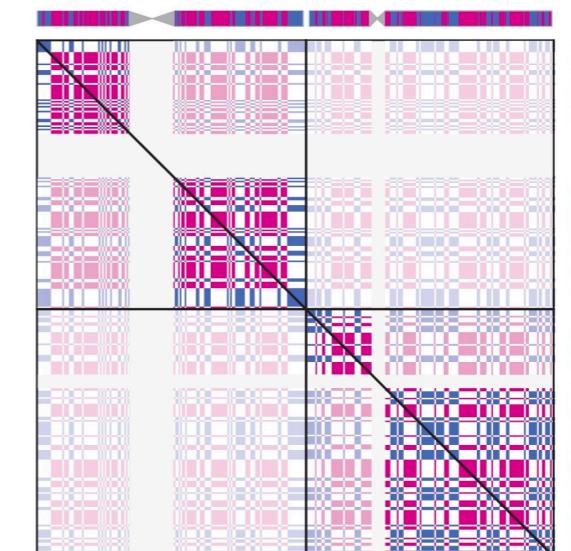
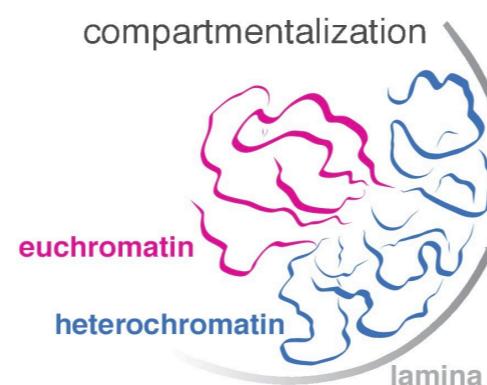
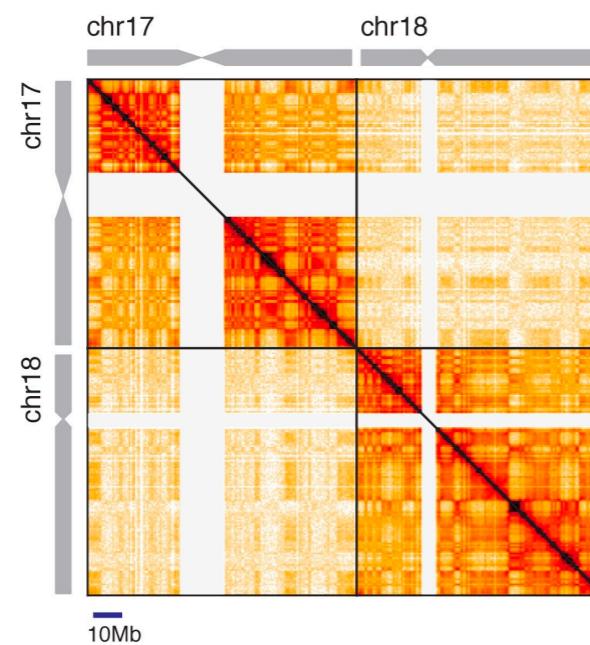
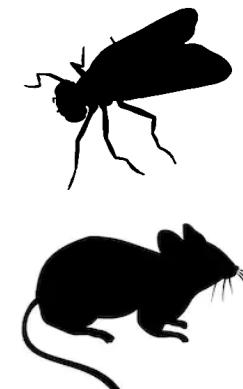
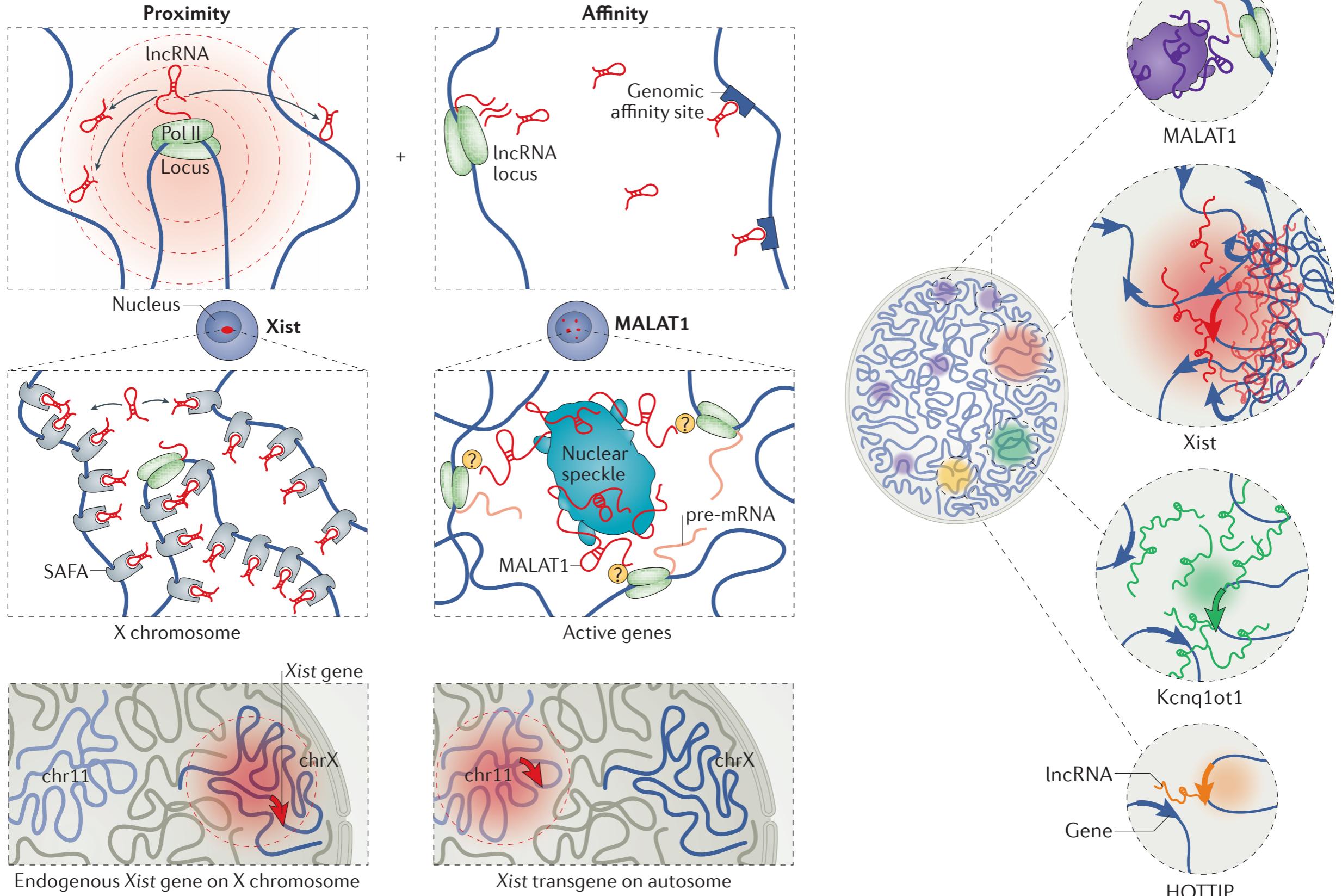


иллюстрация из Migny 2019

### 3. Картирование РНК-ДНК контактов



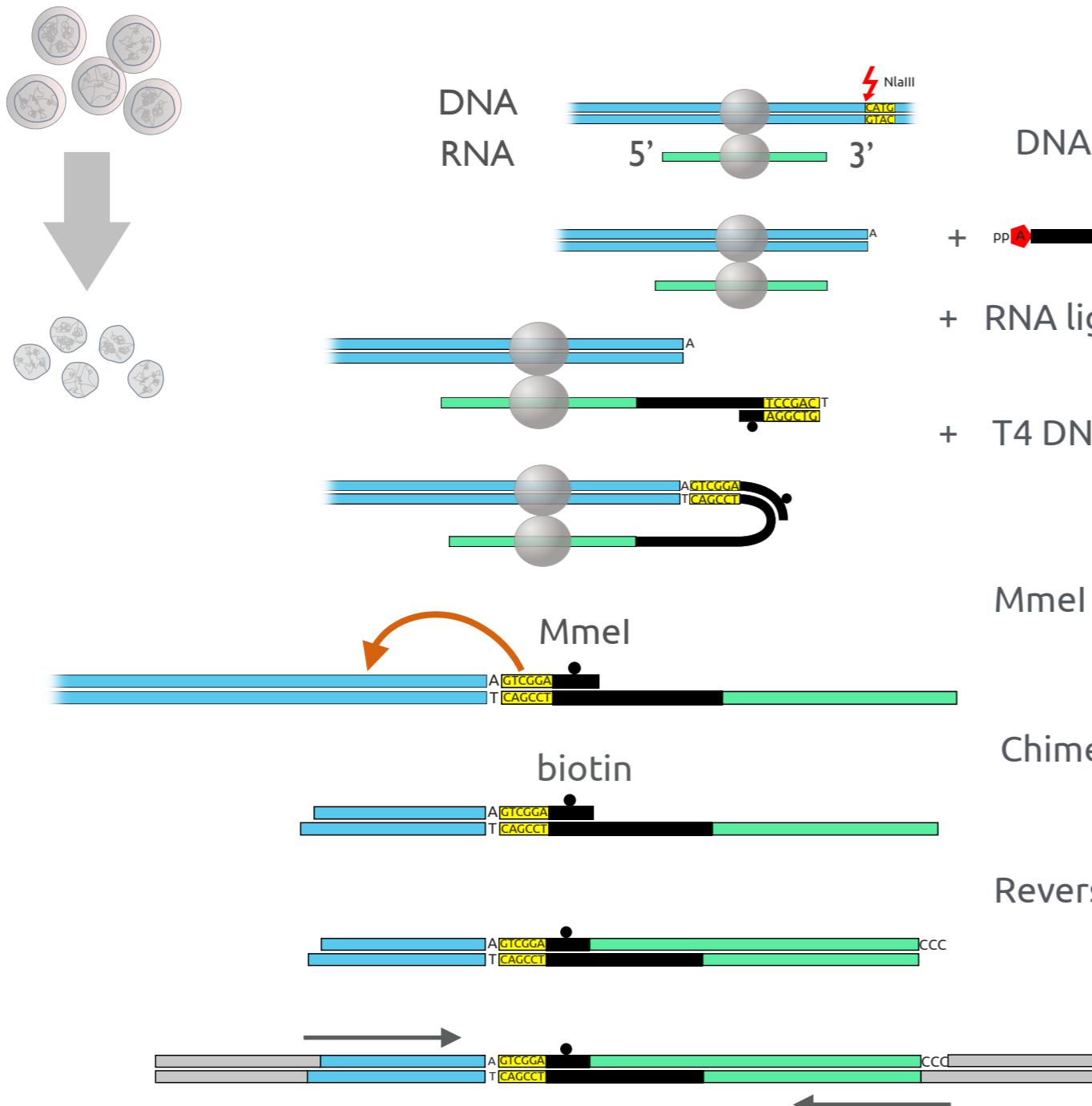
### 3. Разнообразие методов картирования РНК-ДНК контактов

- Chromatin-associated RNA detection:
  - **Nuclear fractionation** - Werner & Ruthenburg 2015
- One-vs-all targeting:
  - **ChIRP-seq** (chromatin isolation by RNA purification) - Chu et al. 2015
  - **CHART** (capture hybridization analysis of RNA targets) - Simon et al. 2011
  - **RAP** (RNA Antisense Purification) - Engreitz et al. 2014
- Many-vs-many targeting:
  - **GRID-Seq** (global RNA interactions with DNA by deep sequencing) - Li et al. 2017
  - **MARGI** (mapping RNA-genome interactions) - Sridhar et al. 2017
  - **ChAR-Seq** (chromatin-associated RNA sequencing) - Bell et al. 2018
  - other

# 3. РНК-ДНК контакты: Red-C

Work in progress

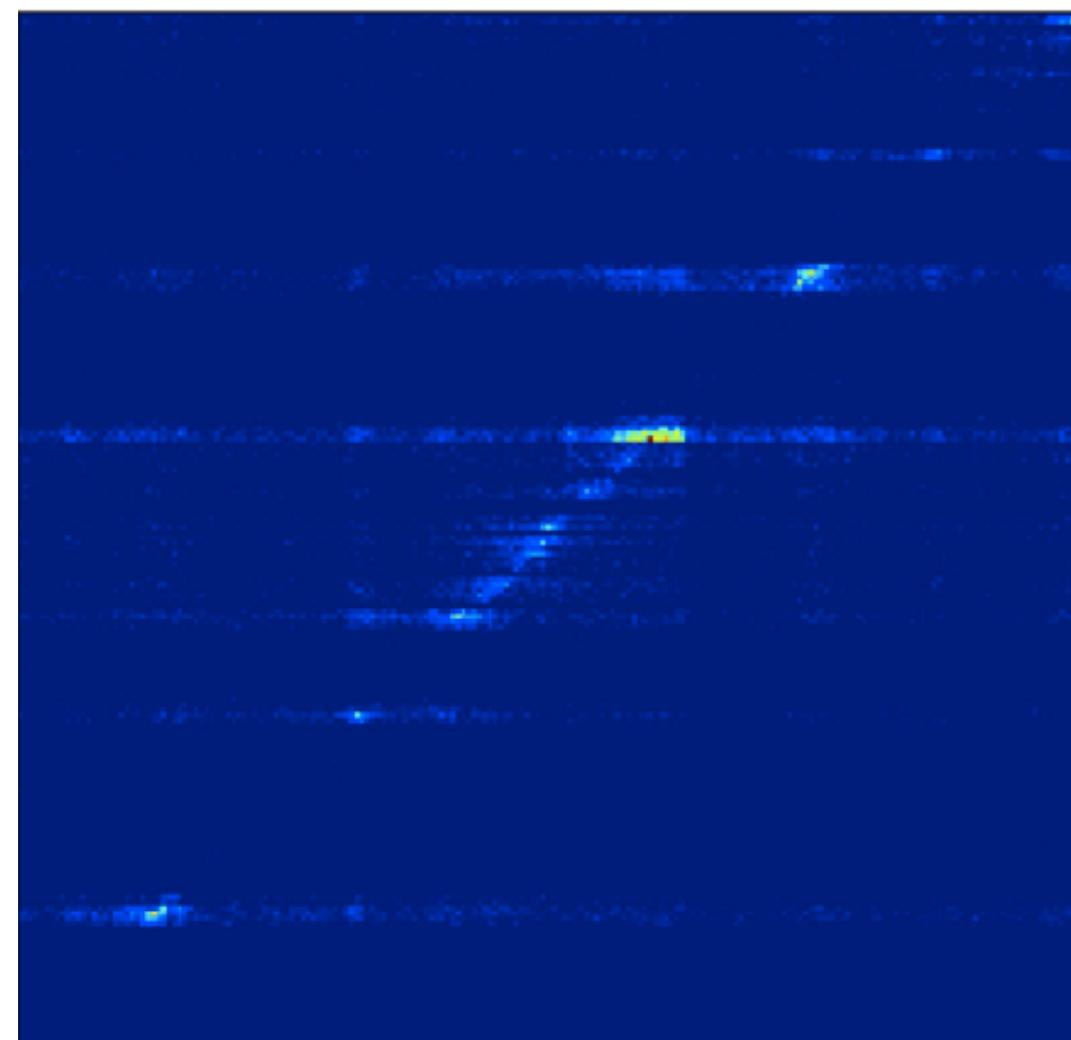
Formaldehyde crosslinking, nuclei isolation



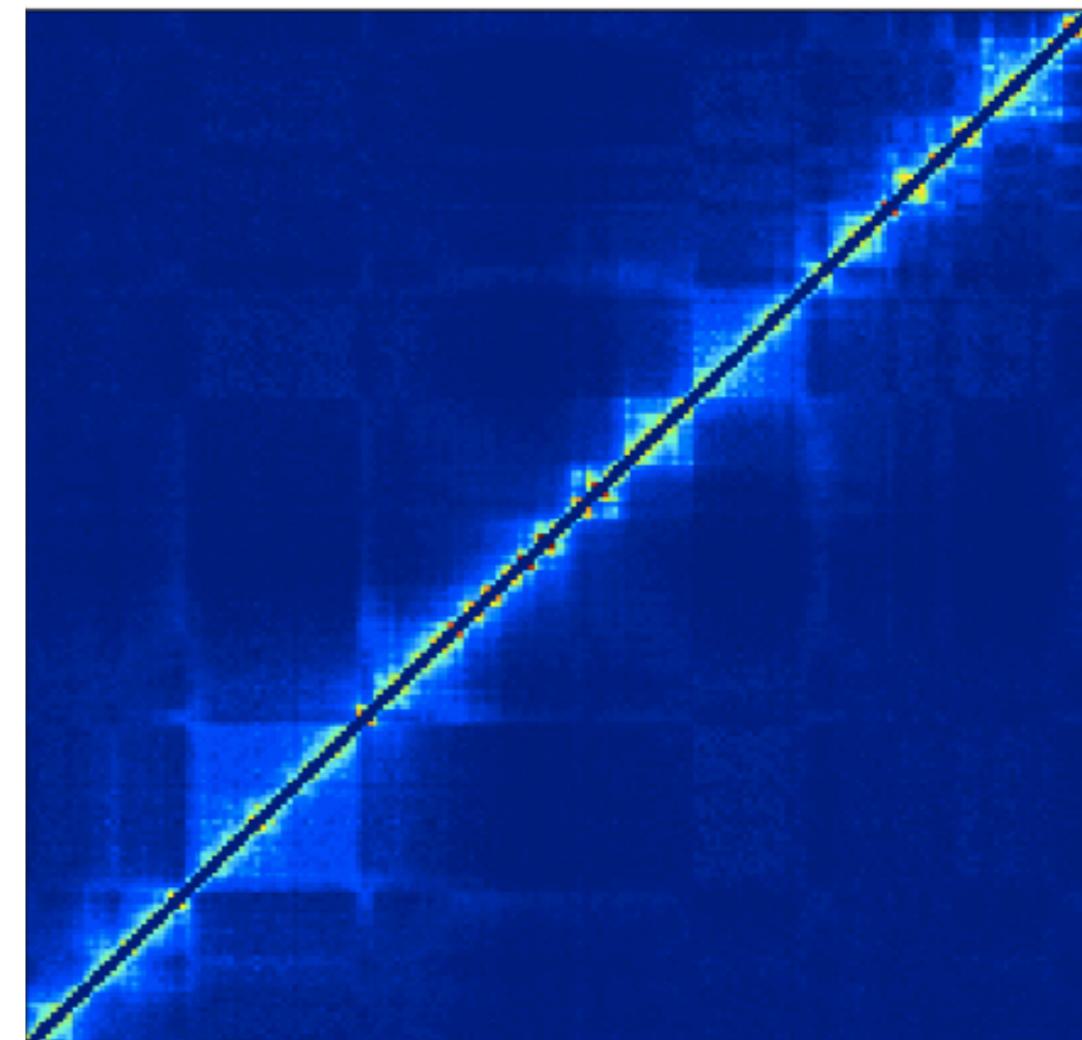
### 3. РНК-ДНК контакты: Red-C

Work in progress

Obtained RNA-DNA interactome



Hi-C from Rao et al. 2014



2.5 Mb

DNA

# Для Ваших отзывов и комментариев:



agalitzina@gmail.com



@agalicina