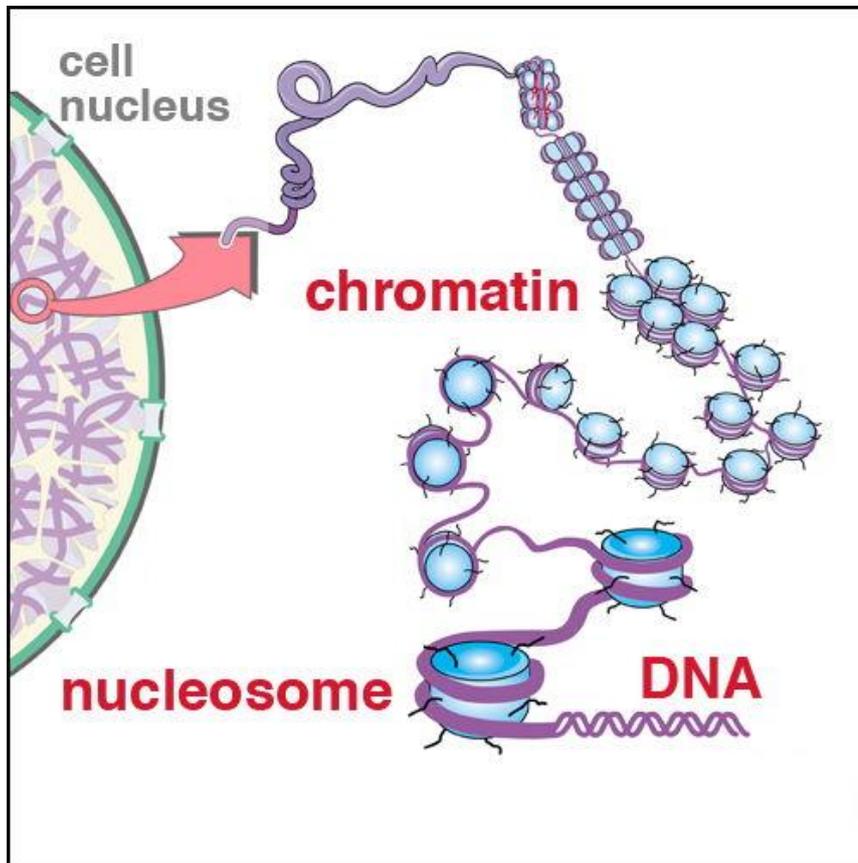


3D-структура хроматина и технологии ее определения



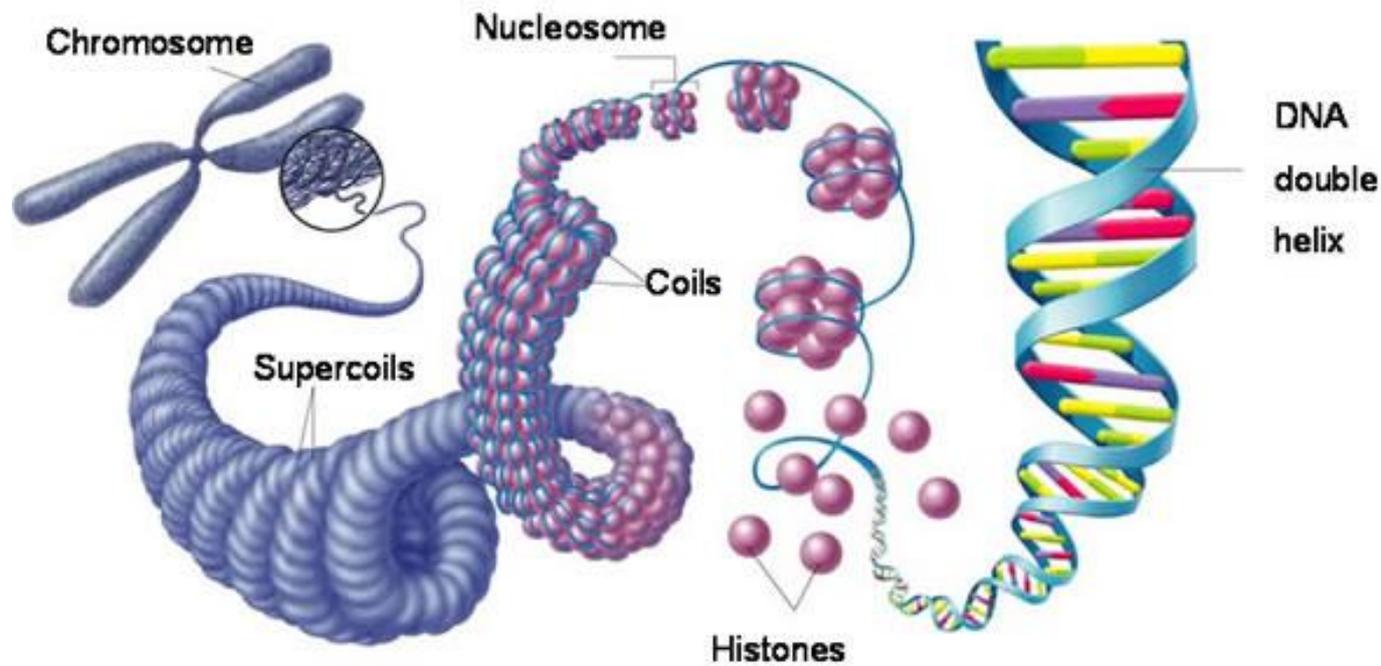
Екатерина Храмеева

*к.б.н., научный сотрудник,
Сколковский институт науки и технологий*

План лекции

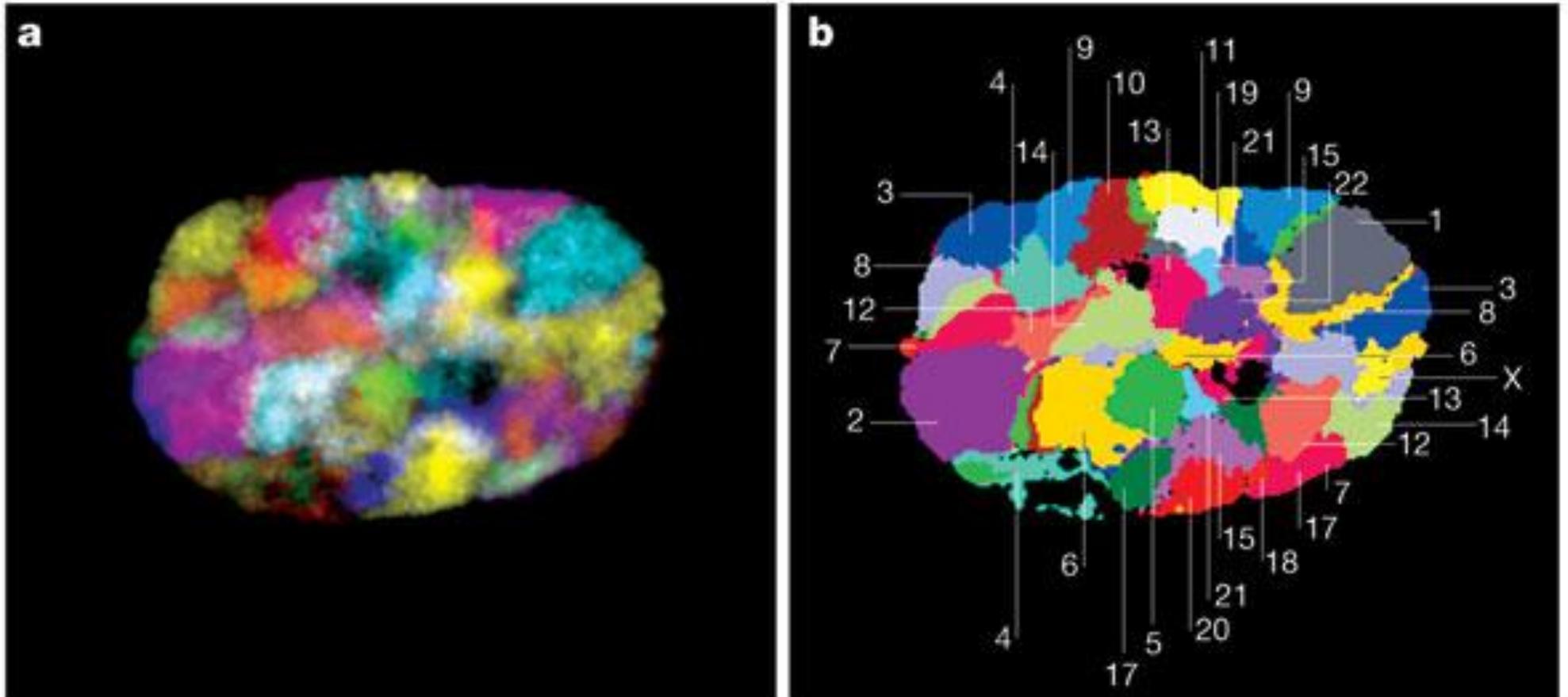
- Принципы организации хроматина
- Экспериментальные методы
 - Обзор 3С-методов
 - Секвенирование
 - Метод Hi-C
- Биоинформатический анализ
 - Анализ HiC данных
 - Построение теплокарт
 - Предсказание топологических доменов
- Какую науку делают с этими данными?

Структура хромосом



- ◆ Хромосомы состоят из ДНК и белков
- ◆ ДНК намотана на специальные белки – *гистоны*, из которых формируются *нуклеосомы*
- ◆ Нитки нуклеосом скручиваются и суперскручиваются, образуя *хромосомы*

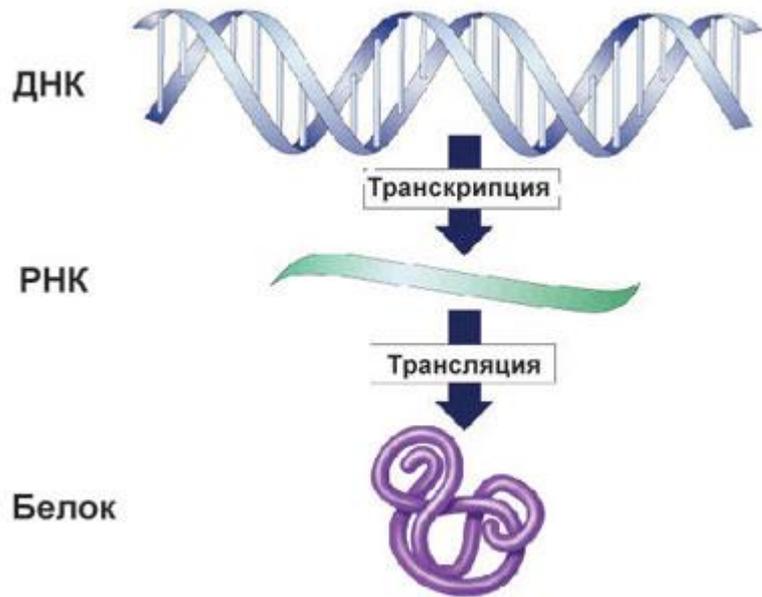
Микрофотография всех хромосом в ядре клетки человека



Copyright © 2005 Nature Publishing Group
Nature Reviews | **Genetics**

Michael R. Speicher & Nigel P. Carter, Nature, 2005

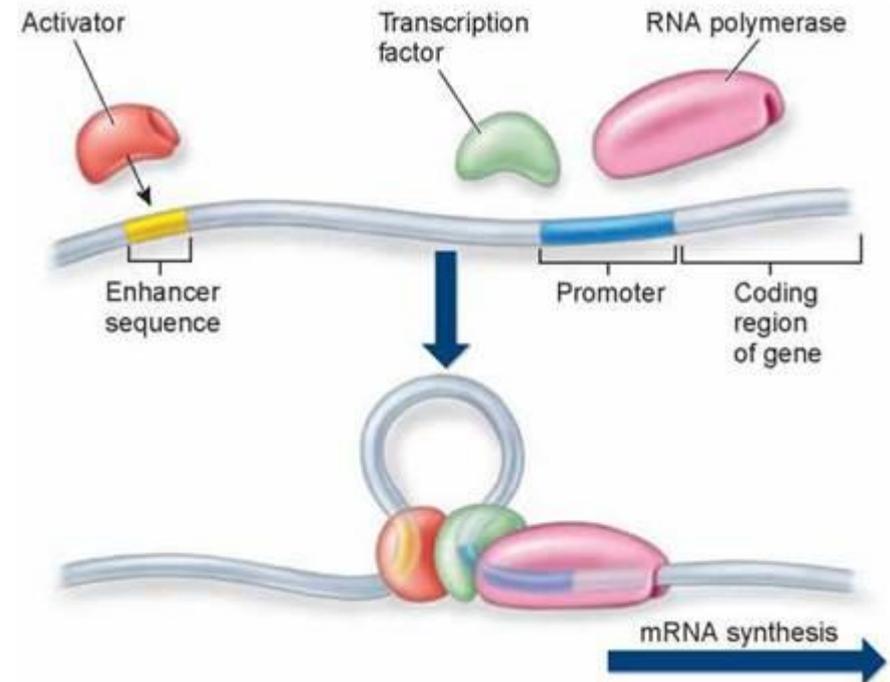
Не только упаковка



В нашем геноме:

- ~ 20 000 генов
- ~ 100 000 регуляторных последовательностей (энхансеров)

- РНК-полимераза — **осуществляет** транскрипцию
- Транскрипционный фактор — **регулирует** транскрипцию



Регуляция работы генов в тканях

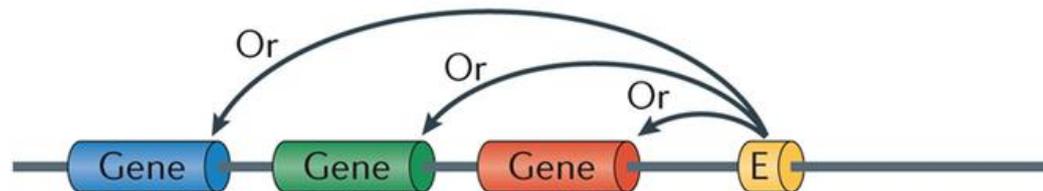
a One gene, multiple enhancers, one tissue



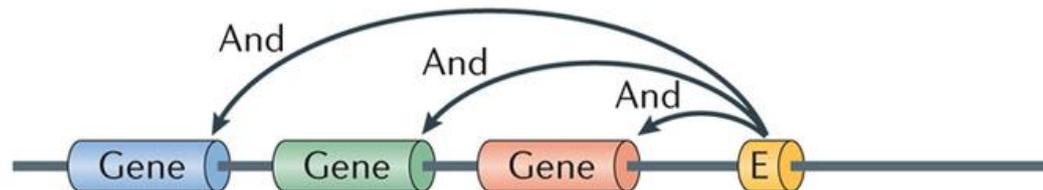
b One gene, multiple enhancers, more tissues



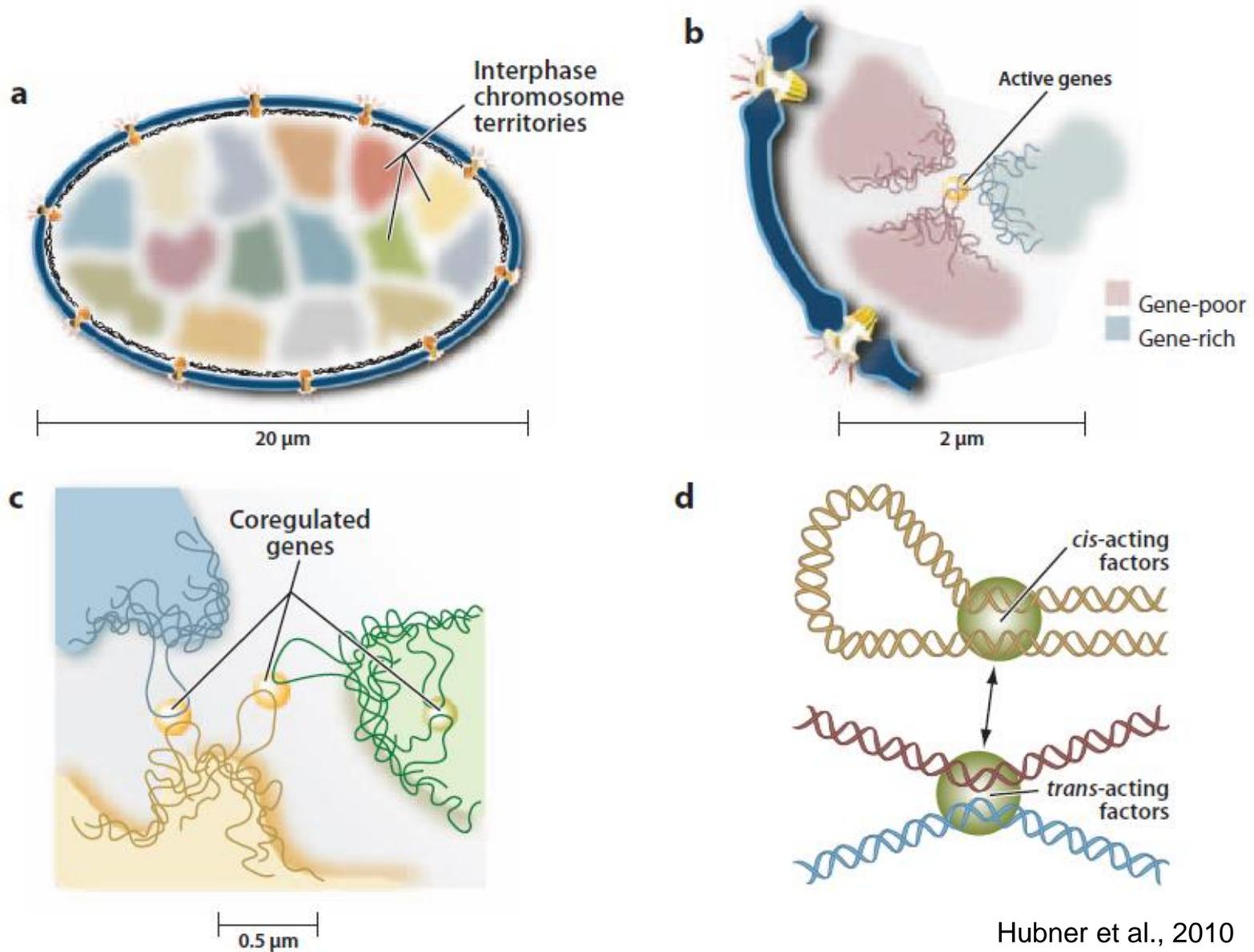
c Gene competition for a shared enhancer: winner takes all



d Gene competition for a shared enhancer: we are all winners



Принципы организации хроматина



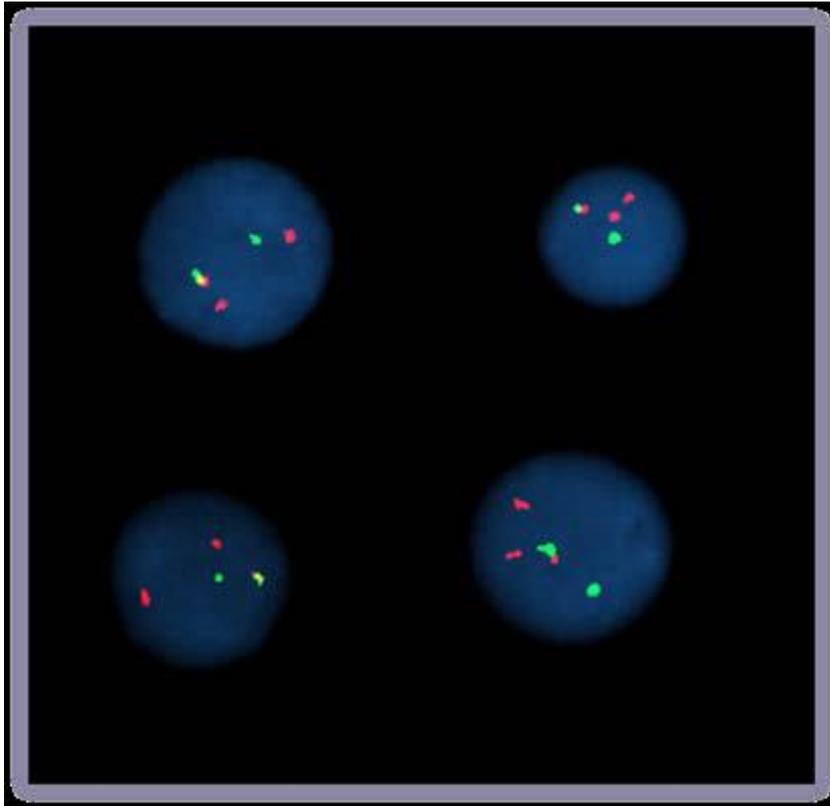
Как изучают 3D структуру генома

- Микроскопия
- FISH (DNA fluorescence in situ hybridization)
- DamID – позволяет определить, какие участки хроматина находятся на периферии ядра
- 3C-методы:
 - 3C (chromosome conformation capture)
 - 4C (chromosome conformation capture-on-chip)
 - 5C (chromosome conformation capture carbon copy)
 - **HiC** и 3C-seq
 - ChIA-PET (chromatin interaction analysis by paired-end tag sequencing)
 - TCC (tethered conformation capture)

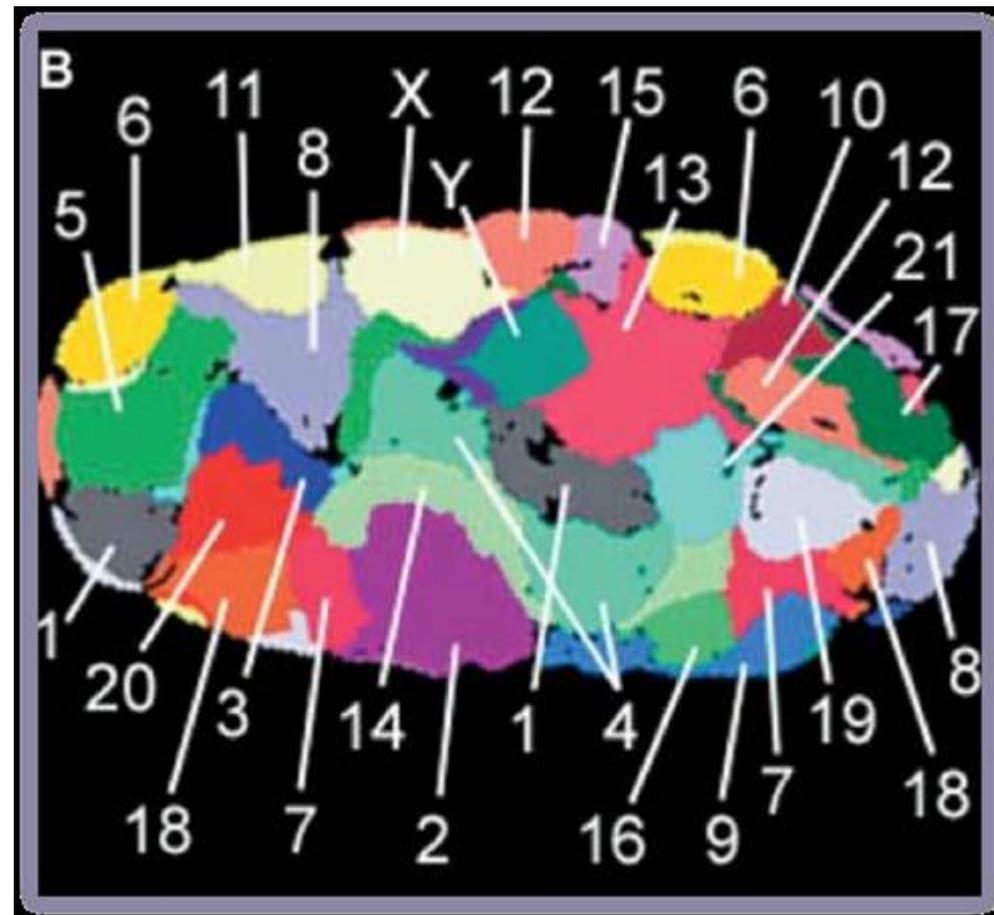
Экспериментальные методы

FISH

(fluorescence in situ hybridization)



Chromosome painting



3C-методы

a 3C: converting chromatin interactions into ligation products



b Ligation product detection methods

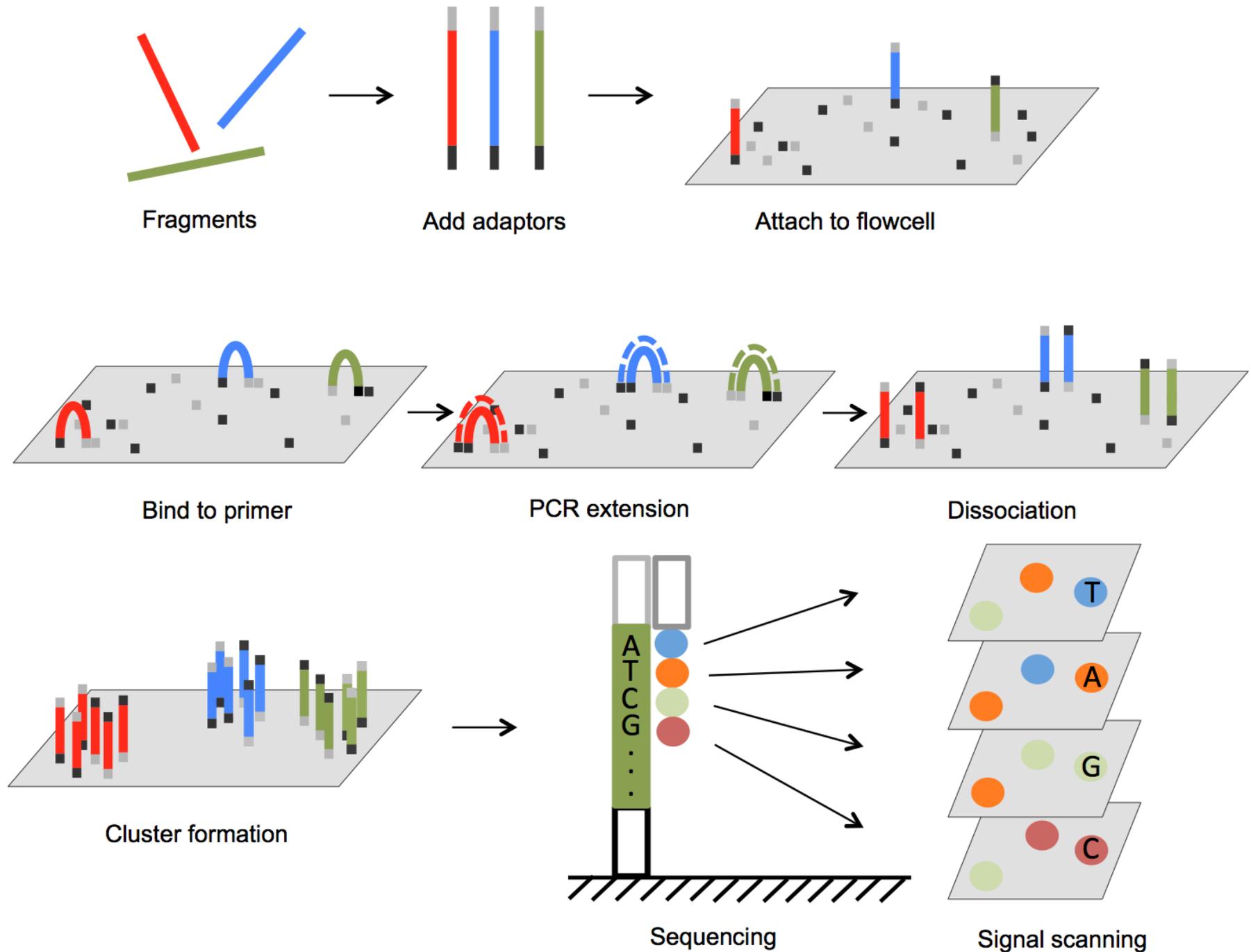
3C	4C	5C	ChIA-PET	Hi-C
One-by-one All-by-all	One-by-all	Many-by-many	Many-by-many	All-by-all
			<ul style="list-style-type: none"> • DNA shearing • Immunoprecipitation 	<ul style="list-style-type: none"> • Biotin labelling of ends • DNA shearing
PCR or sequencing	Inverse PCR sequencing	Multiplexed LMA sequencing	Sequencing	Sequencing

Что такое секвенирование?

- Секвенировать = установить последовательность нуклеотидов в молекуле ДНК.
- Сначала ДНК надо выделить.
- Нельзя прочесть молекулу ДНК целиком, поэтому ДНК режут на кусочки в случайных местах и «размножают».



Как делают секвенирование? (Illumina)



Зачем секвенировать?

- **Секвенирование генома и транскриптома**
основа любых исследований, поиск крупных геномных перестроек
- **Ресеквенирование**
поиск мутаций, связанных с болезнями, картирование генов
- **Анализ экспрессии генов**
сравнение уровней экспрессии, поиск новых изоформ генов
- **ДНК-белковые взаимодействия**
поиск сайтов связывания транскрипционных факторов
- **ДНК-ДНКовые взаимодействия**
изучение пространственной организации хроматина

Секвенирование генома и транскриптома *de novo*

Голый землекоп



- Колониальный грызун
 - Холоднокровный
 - Выносит кислородное голодание
 - Уровень заболеваемости раком крайне низок
 - Чем он отличается от остальных млекопитающих, которые болеют раком?
-
- Ген устойчивости к раку? Как узнать? Секвенировать геном!
 - Как узнать, где в геноме гены? Секвенировать транскриптом!
 - Как найти геномные перестройки в раках? Секвенировать геном!

Ресеквенирование

Ресеквенировать = детектировать мутации, ассоциированные с определенным заболеванием

←
Тестирование на известные мутации

→
Поиск неизвестных мутаций

Мутации:

- Однонуклеотидные полиморфизмы
- Вставки и делеции небольшого размера

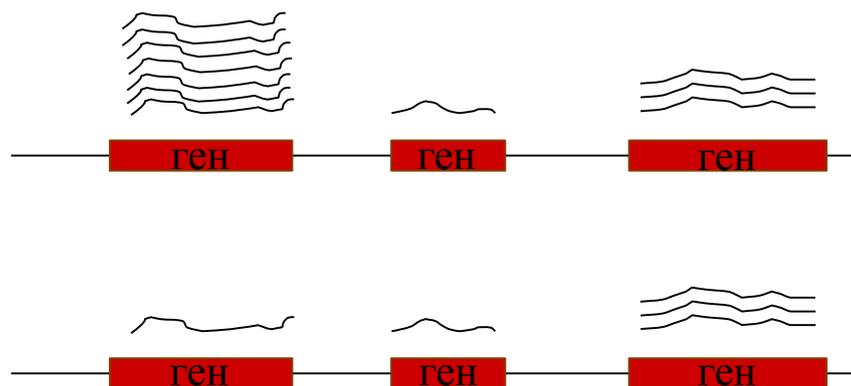
- Большая глубина секвенирования (т.е. количество ридов) не нужна => дешево
- Нужен референсный геном
- Риды надо картировать на геном. Отличия = мутации.

Анализ экспрессии генов

Голый землекоп



- Гена устойчивости к раку не нашли.
- Но почему он другой? Гены те же, но работают по-другому?
- Как узнать? Сравнить уровни экспрессии генов у землекопа и мыши.

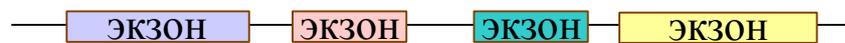


Анализ изоформ генов

ДНК



РНК



сплайсинг



РНК

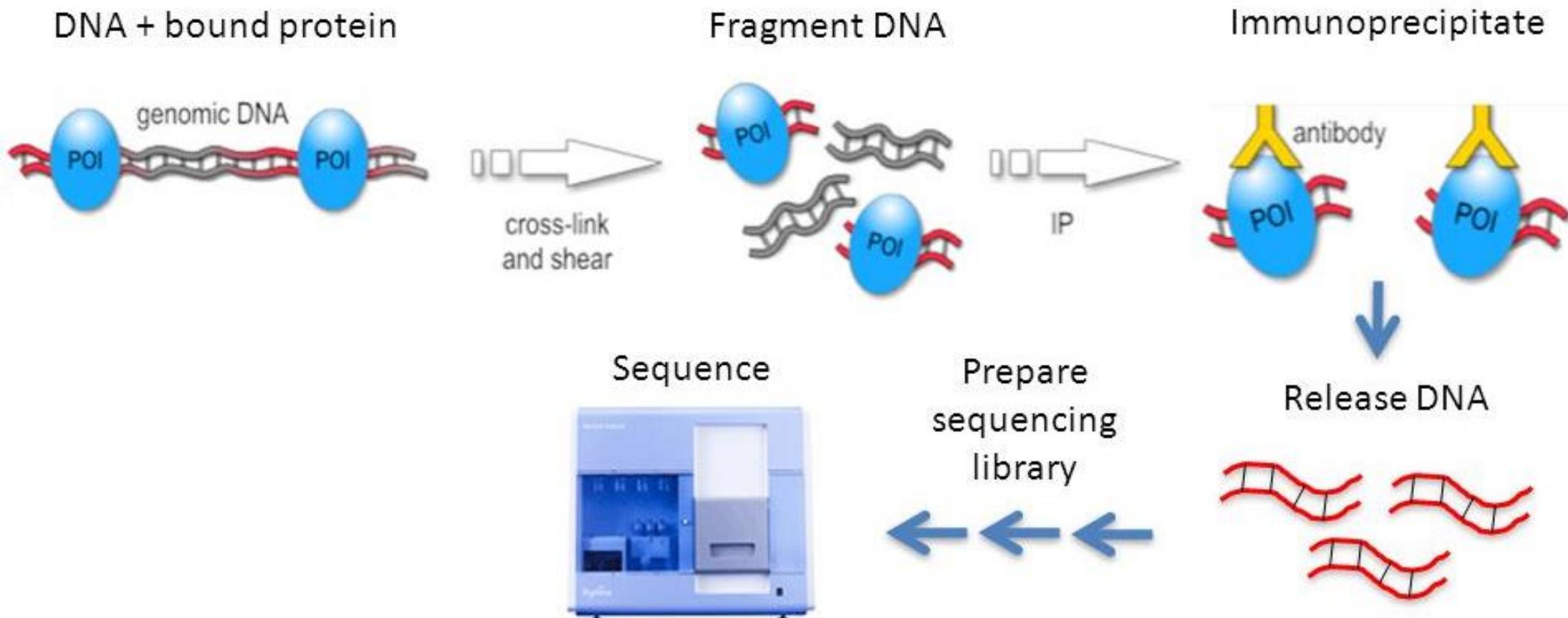


РНК



ДНК-белковые взаимодействия

- Нашли разницу в экспрессии генов. Чем она вызвана?
- Регуляторы экспрессии (= транскрипционные факторы)
- ChIP-seq (чипсек) – метод иммунопреципитации хроматина



ДНК-ДНКовые взаимодействия

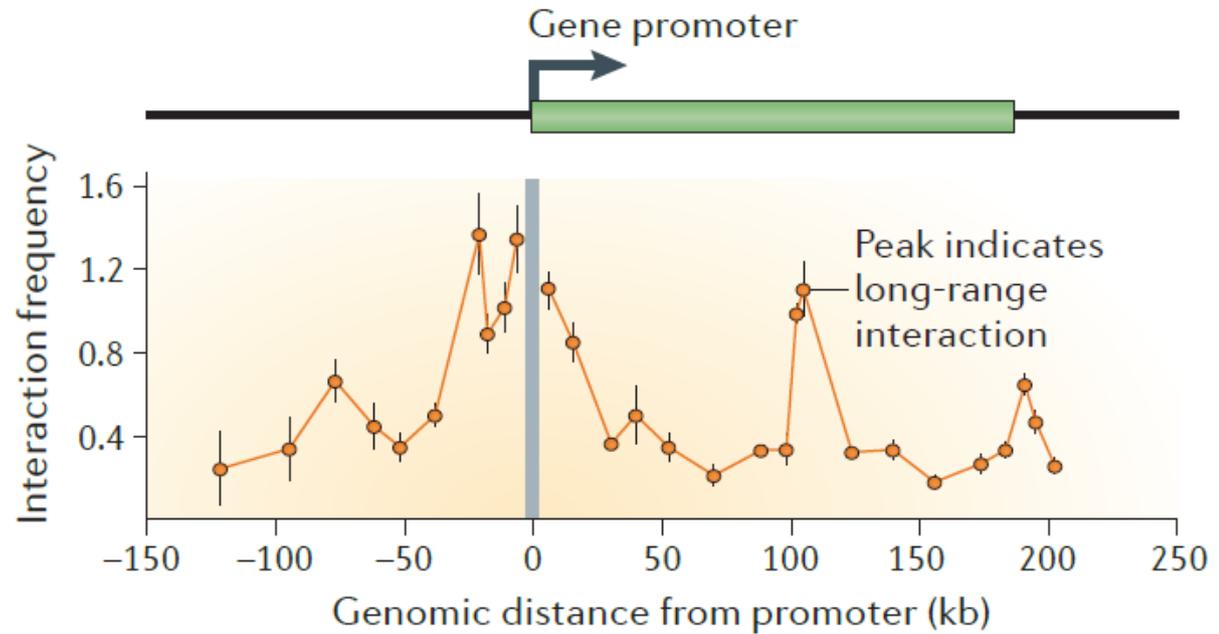
- Изучение пространственной структуры хроматина

Метод 3C



Пример 3C данных:

3C
One-by-one All-by-all
PCR or sequencing

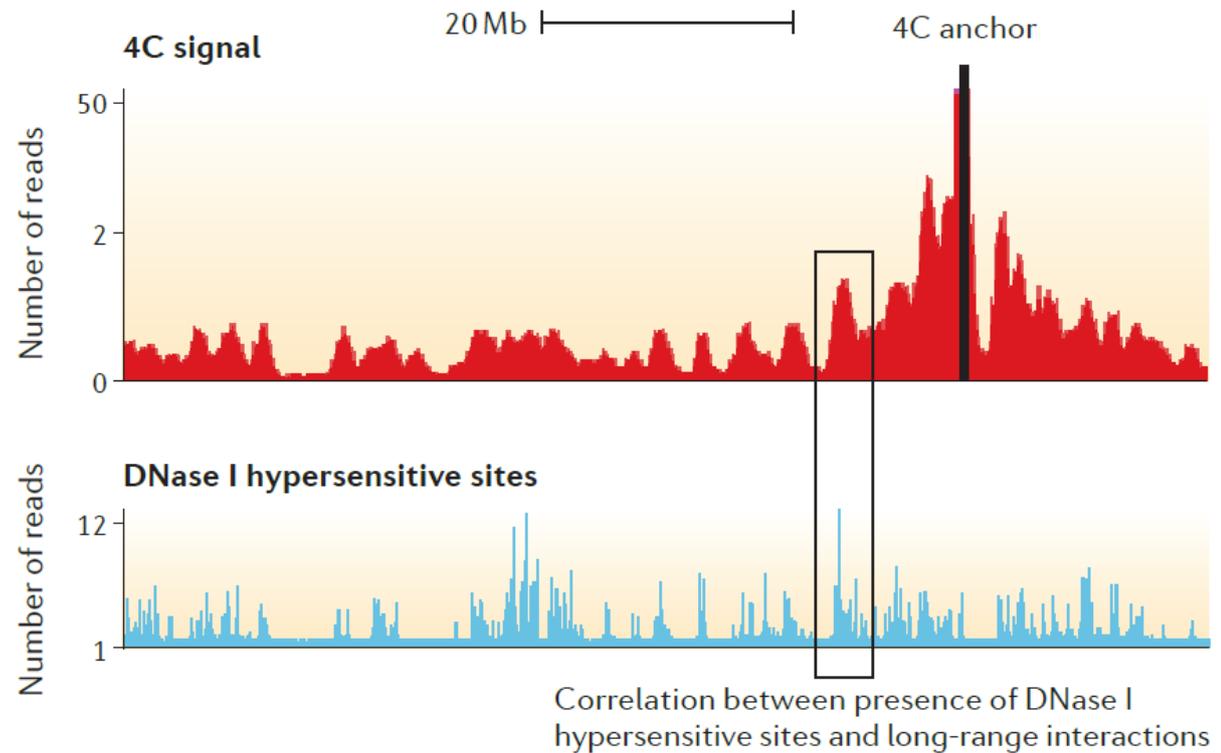


Метод 4C



Пример 4C данных:

3C	4C
One-by-one All-by-all	One-by-all
PCR or sequencing	Inverse PCR sequencing

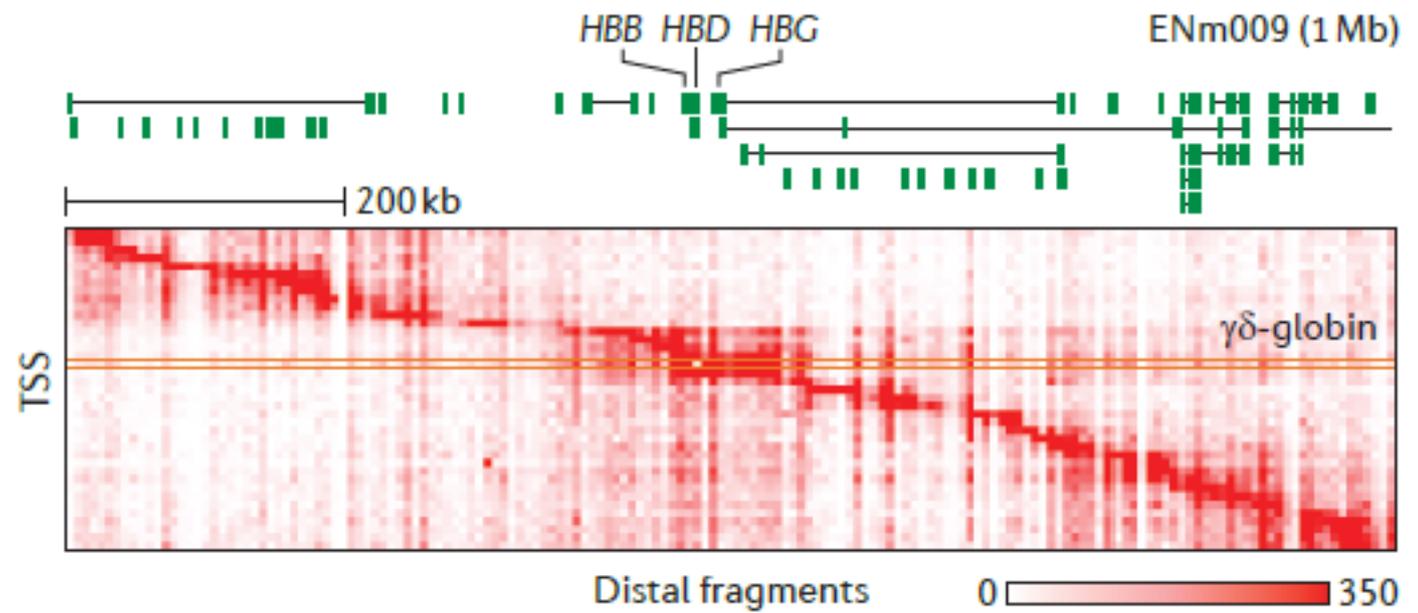


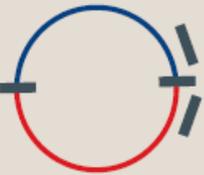
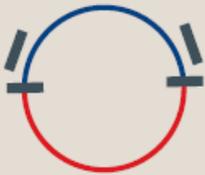
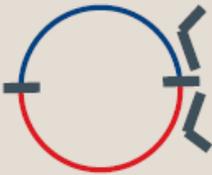
Метод 5C



3C	4C	5C
One-by-one All-by-all	One-by-all	Many-by-many
PCR or sequencing	Inverse PCR sequencing	Multiplexed LMA sequencing

Пример 5C данных:



3C	4C	5C
One-by-one All-by-all	One-by-all	Many-by-many
		
PCR or sequencing	Inverse PCR sequencing	Multiplexed LMA sequencing

Метод ChIA-PET



3C	4C	5C	ChIA-PET
One-by-one All-by-all	One-by-all	Many-by-many	Many-by-many
			<ul style="list-style-type: none"> • DNA shearing • Immunoprecipitation
PCR or sequencing	Inverse PCR sequencing	Multiplexed LMA sequencing	Sequencing

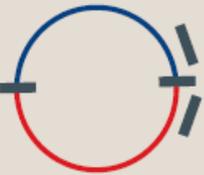
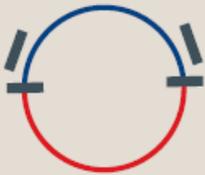
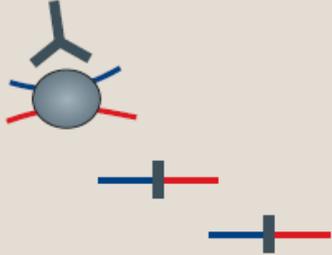
Метод Hi-C



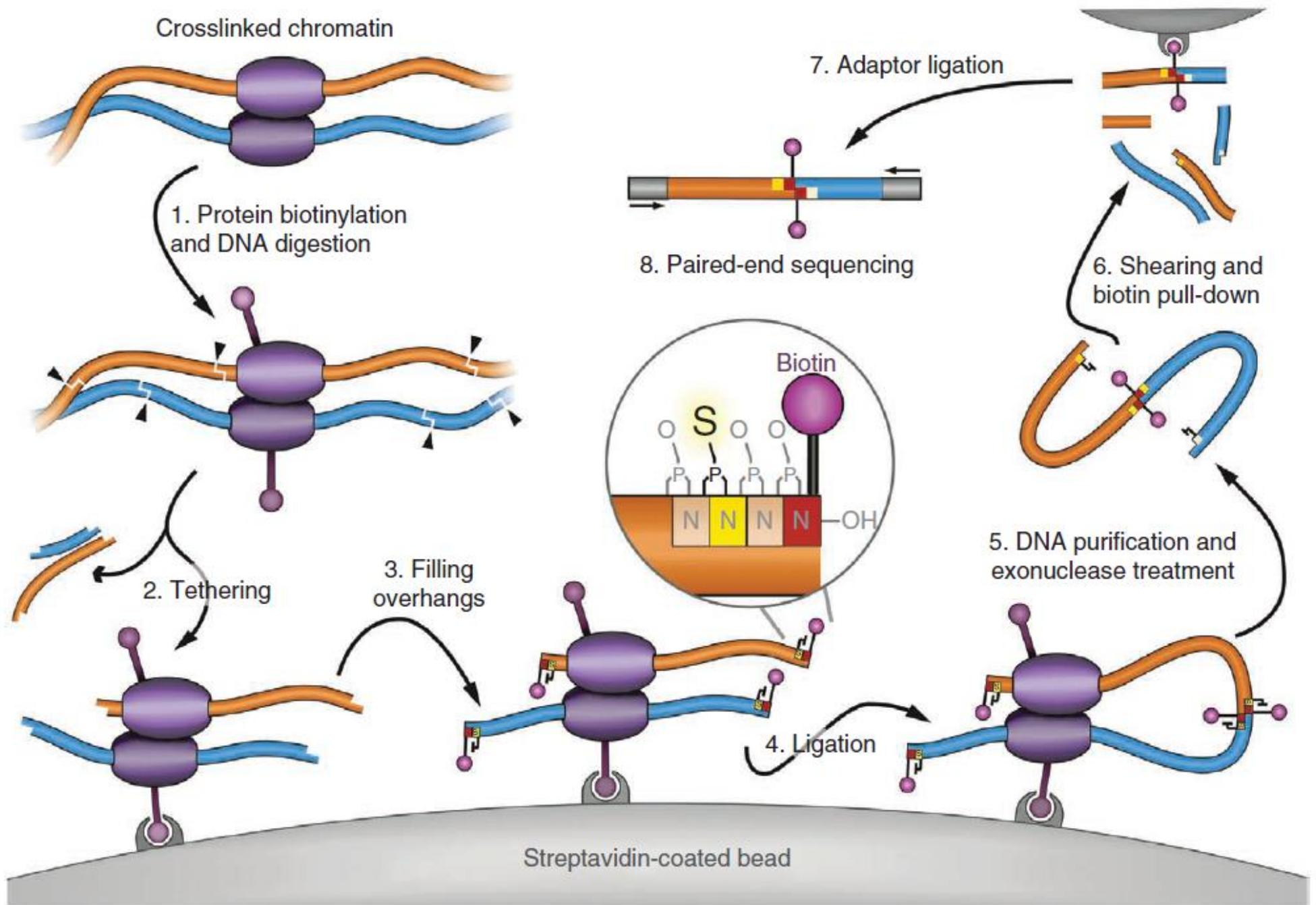
3C	4C	5C	ChIA-PET	Hi-C
One-by-one All-by-all	One-by-all	Many-by-many	Many-by-many	All-by-all
			<ul style="list-style-type: none"> • DNA shearing • Immunoprecipitation 	<ul style="list-style-type: none"> • Biotin labelling of ends • DNA shearing
PCR or sequencing	Inverse PCR sequencing	Multiplexed LMA sequencing	Sequencing	Sequencing

Метод 3C-seq

- То же самое, что Hi-C, но без биотинилирования.
- На сиксенс идут не только химерные кусочки.
- Приходится больше секвенировать!

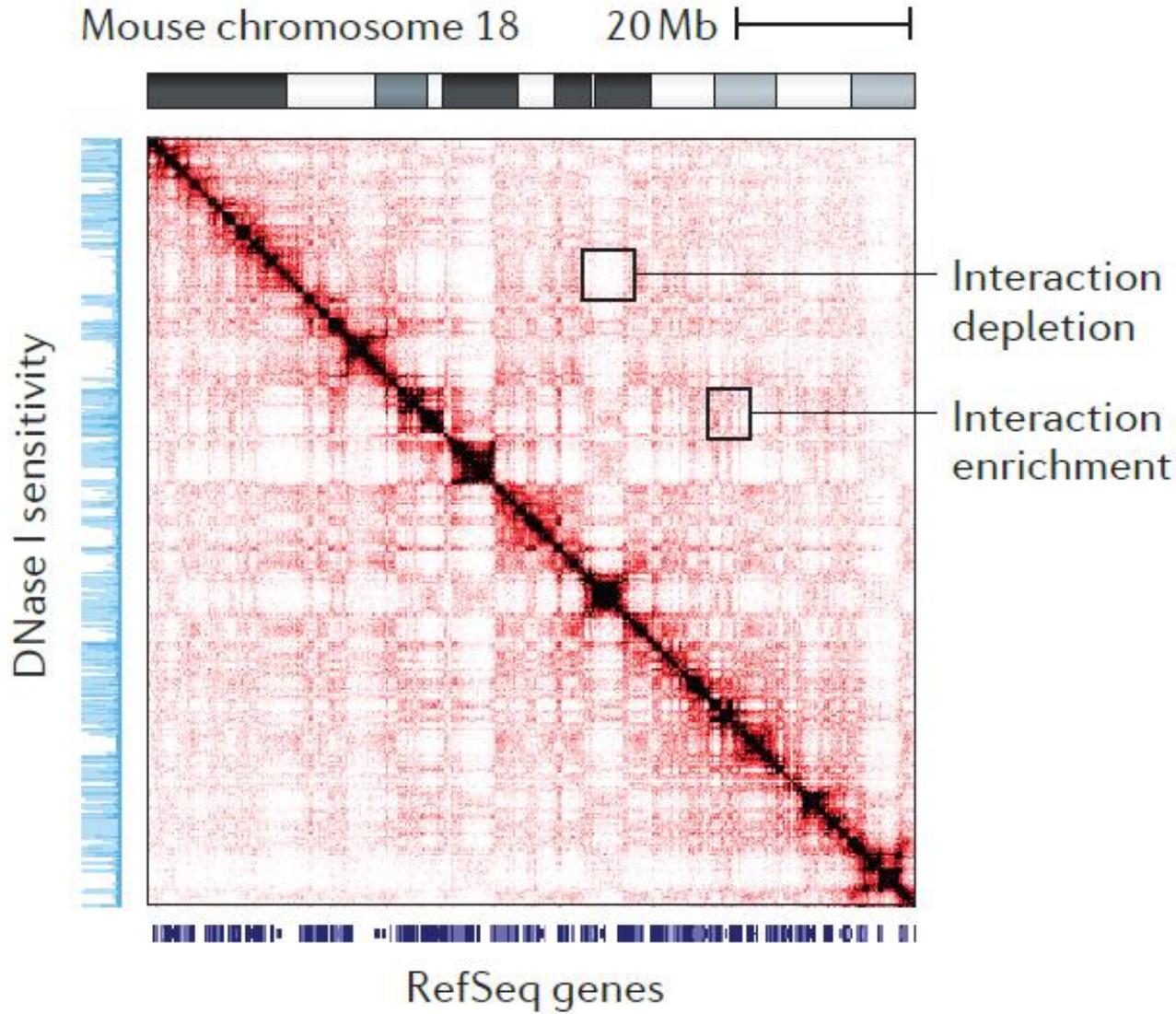
3C	4C	5C	ChIA-PET	Hi-C
One-by-one All-by-all	One-by-all	Many-by-many	Many-by-many	All-by-all
			<ul style="list-style-type: none"> • DNA shearing • Immunoprecipitation 	<ul style="list-style-type: none"> • Biotin labelling of ends • DNA shearing 
PCR or sequencing	Inverse PCR sequencing	Multiplexed LMA sequencing	Sequencing	Sequencing

Метод ТСС



Метод Hi-C

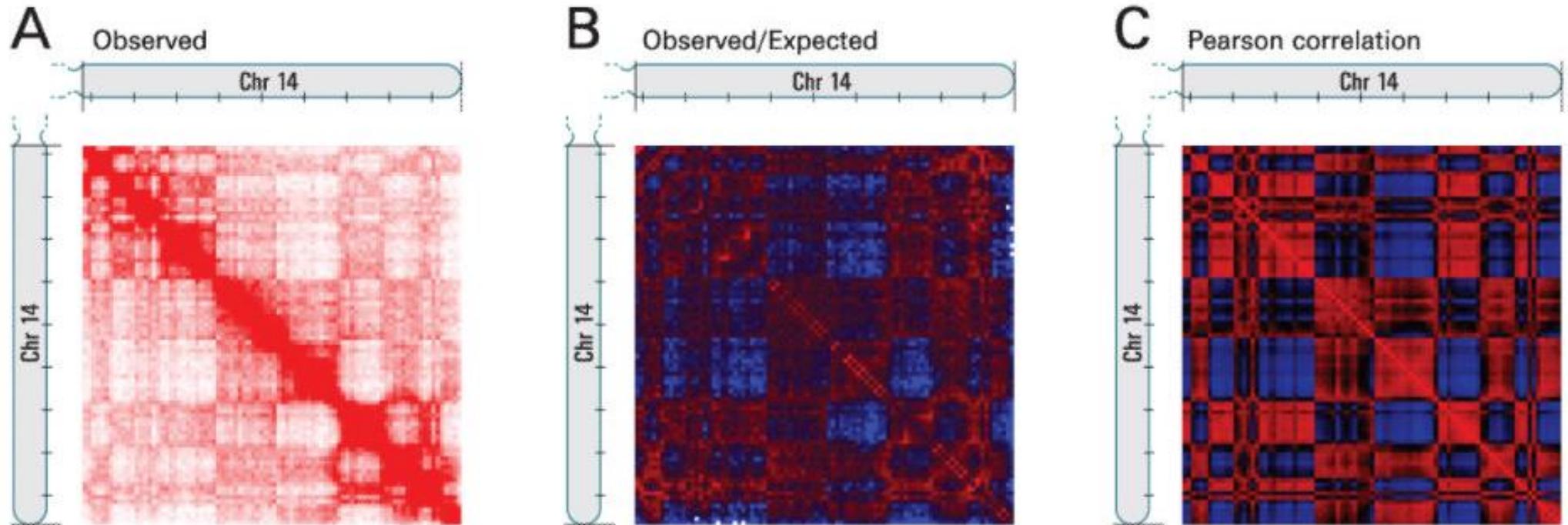
Пример Hi-C данных:

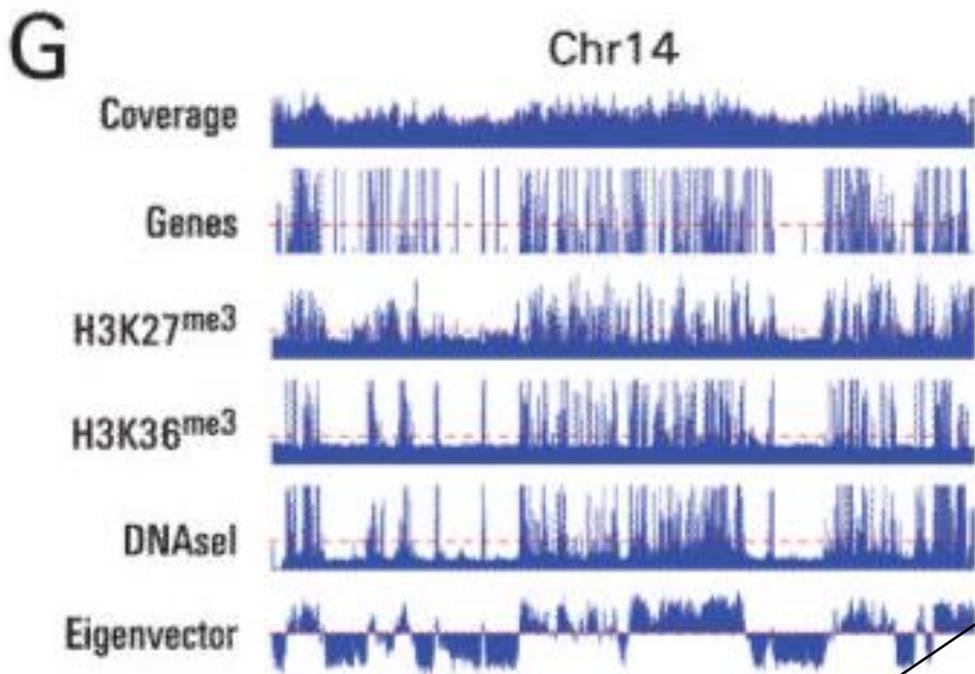


Hi-C
All-by-all
<ul style="list-style-type: none">• Biotin labelling of ends• DNA shearing

Sequencing

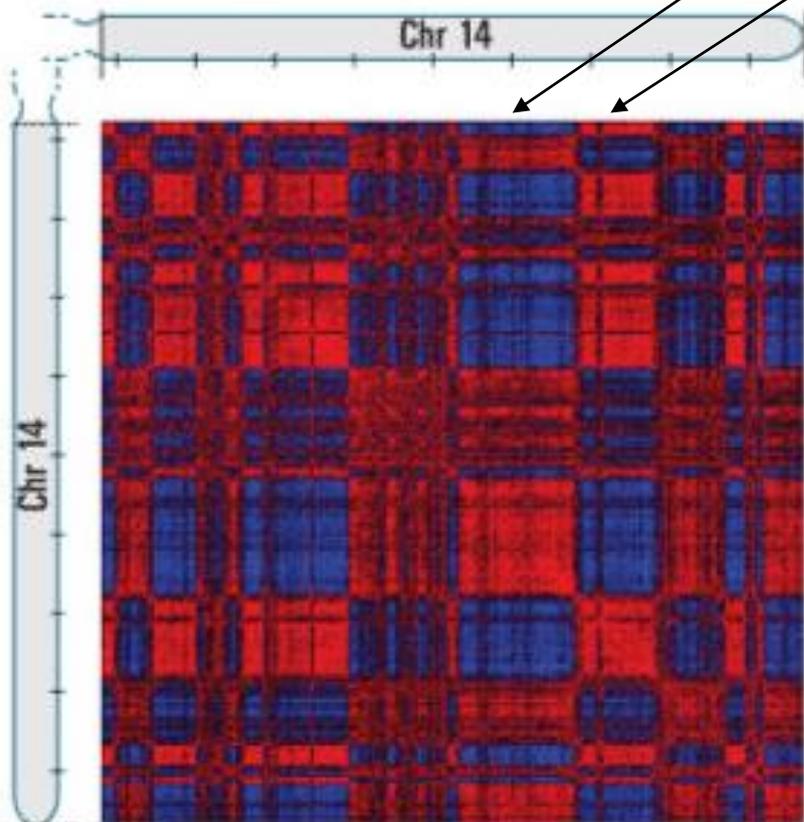
Матрица контактов Hi-C





Первая главная компонента (собственный вектор, или eigenvector) коррелирует с плотностью генов и маркерами открытого хроматина.

Компартмент А
Компартмент В



Корреляционная матрица Hi-C для хромосомы 14, разрешение 100 тыс. п.о.

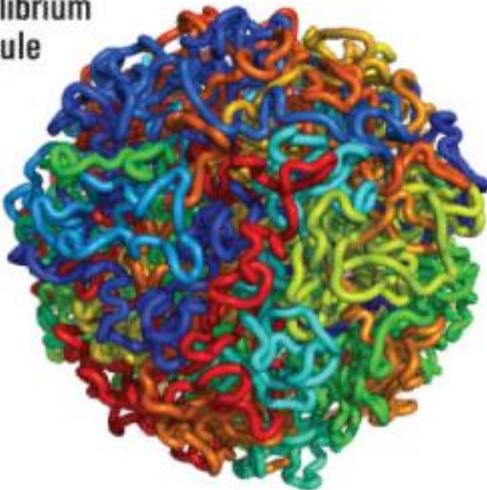
C

UNFOLDED POLYMER

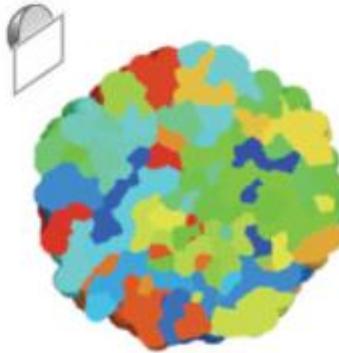


FOLDED POLYMER

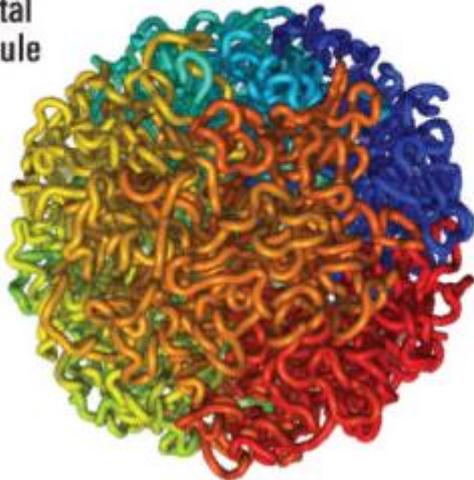
Equilibrium globule



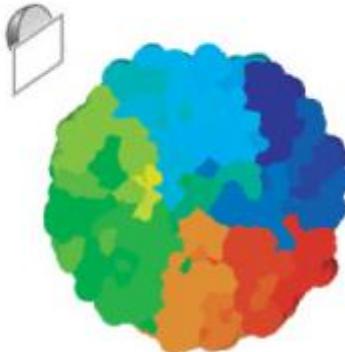
Cross-section view



Fractal globule



Cross-section view



Фрактальная глобула

Биоинформатический анализ

Hi-C workflow

1. Read mapping

Iterative mapping.

Paired-end mode is not used.

2. Fragment assignment

The mapped read is assigned according to its 5' mapped position.
Mapped read positions should fall close to a restriction site.

3. Fragment filtering

PCR duplicates, undigested restriction sites.

4. Binning

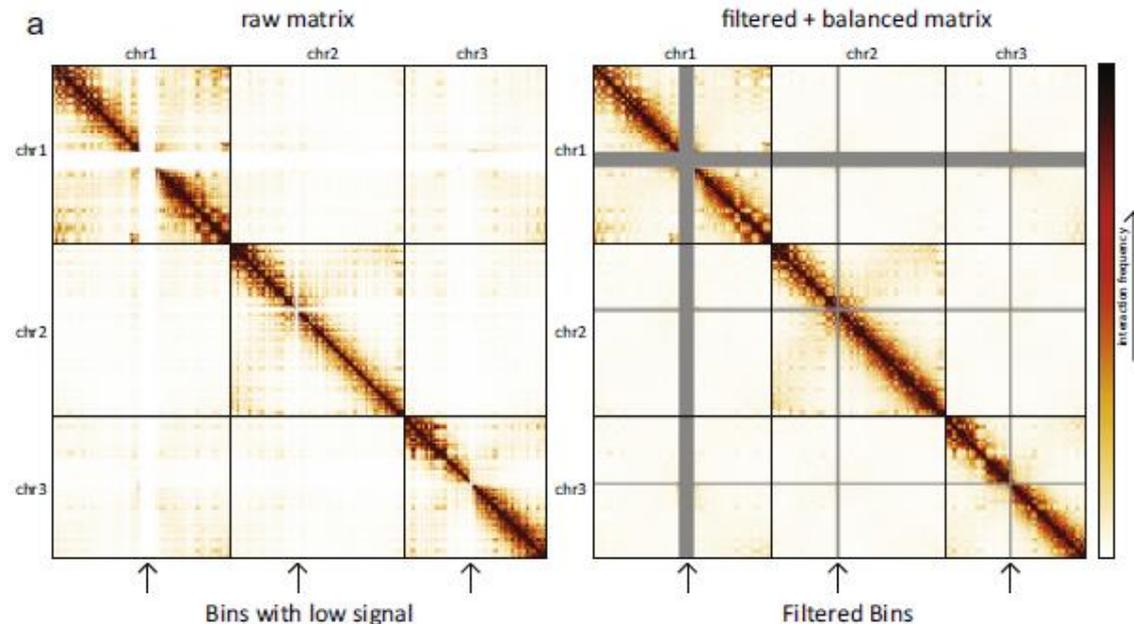
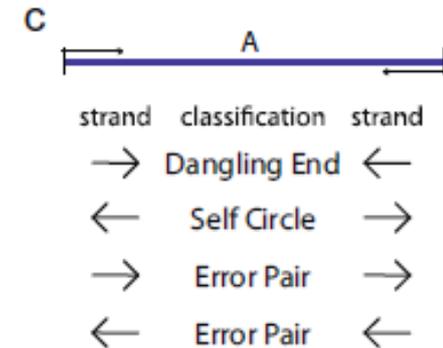
5. Bin level filtering

Remove 1% low signal rows/columns.

6. Balancing

Iterative correction.

Iterative Mapping
(paired end)

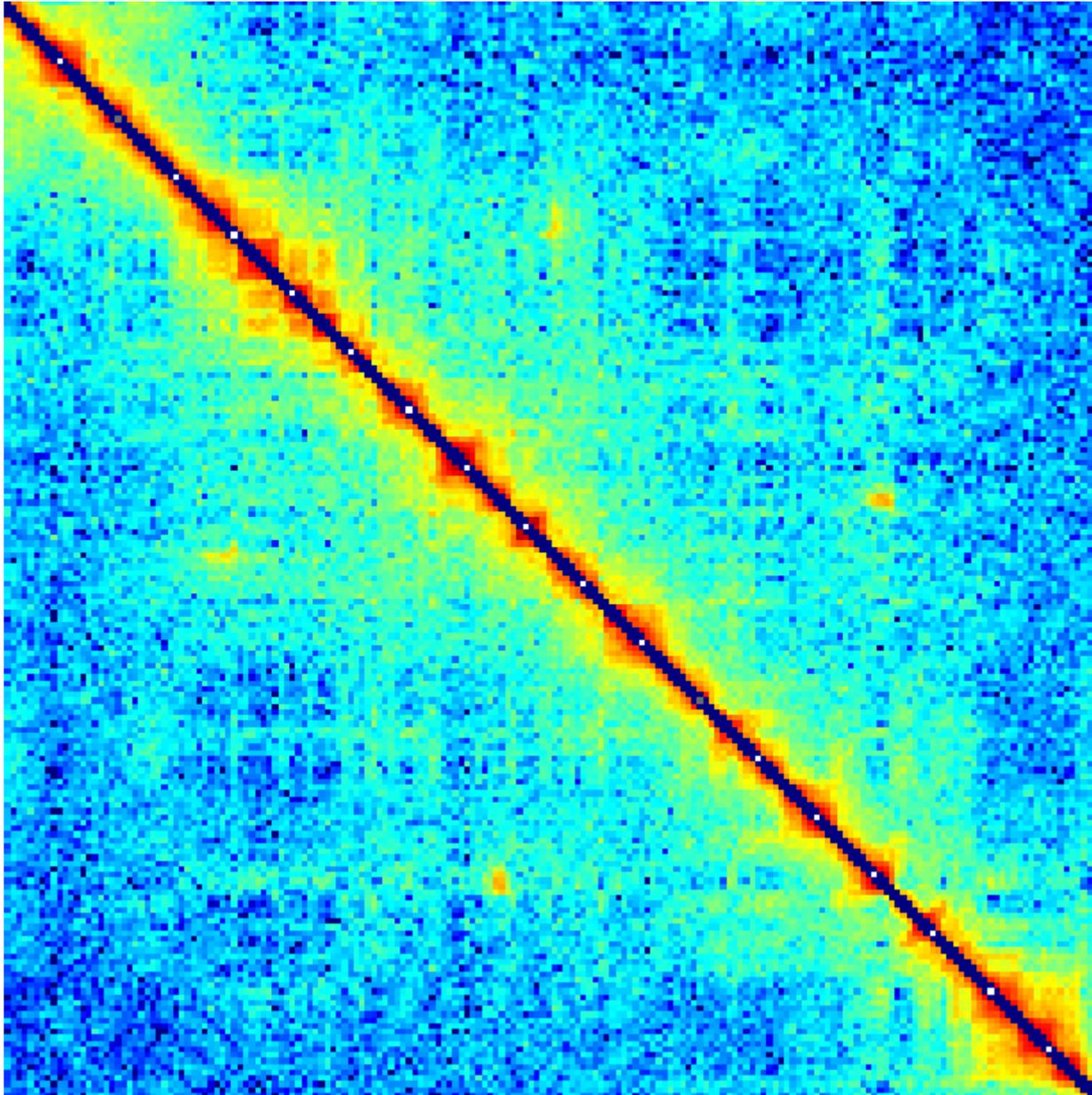


Lajoie, B. R., Dekker, J. & Kaplan, N.

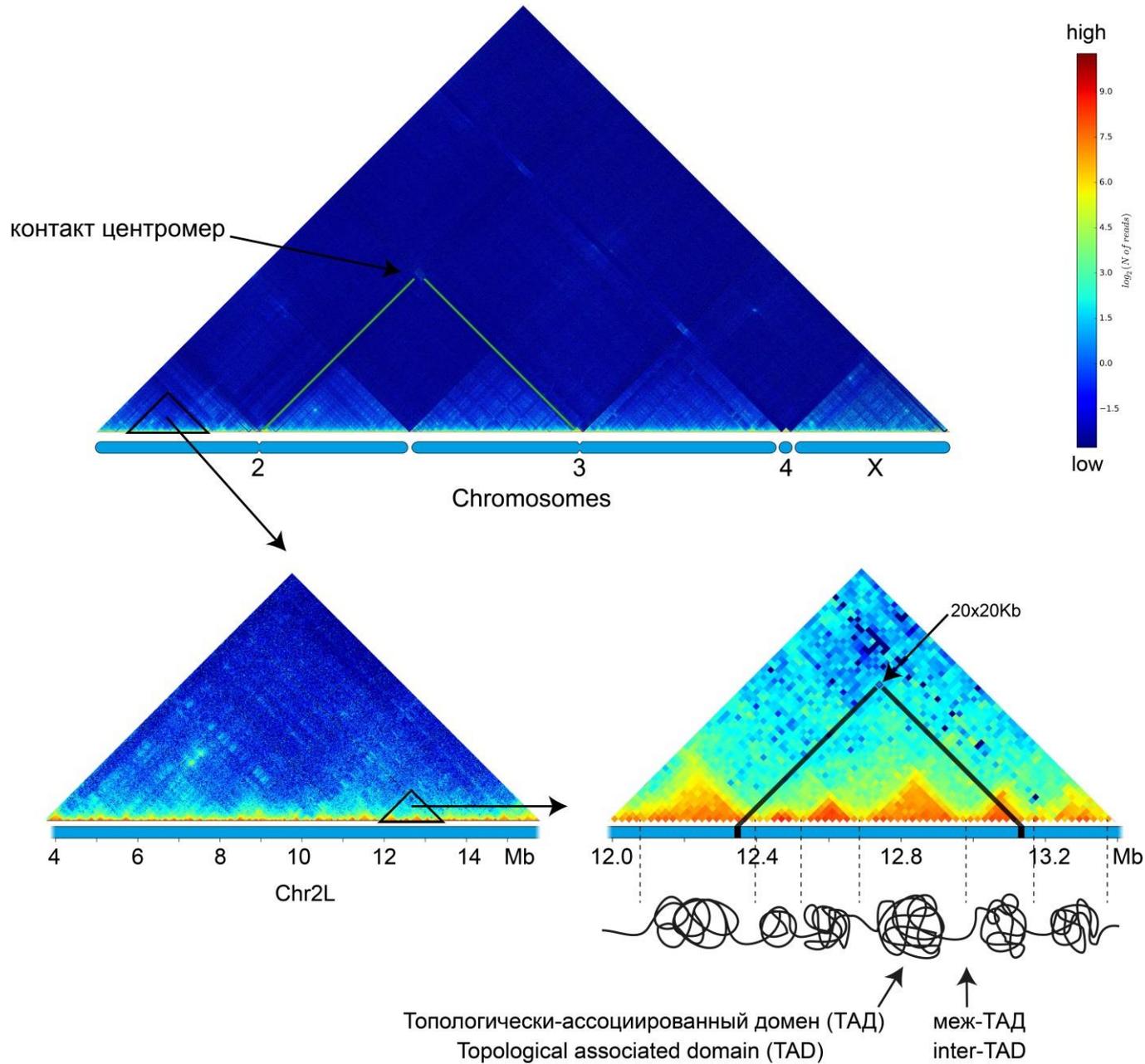
The Hitchhiker's guide to Hi-C analysis: Practical guidelines.

Methods 72, 65–75 (2015).

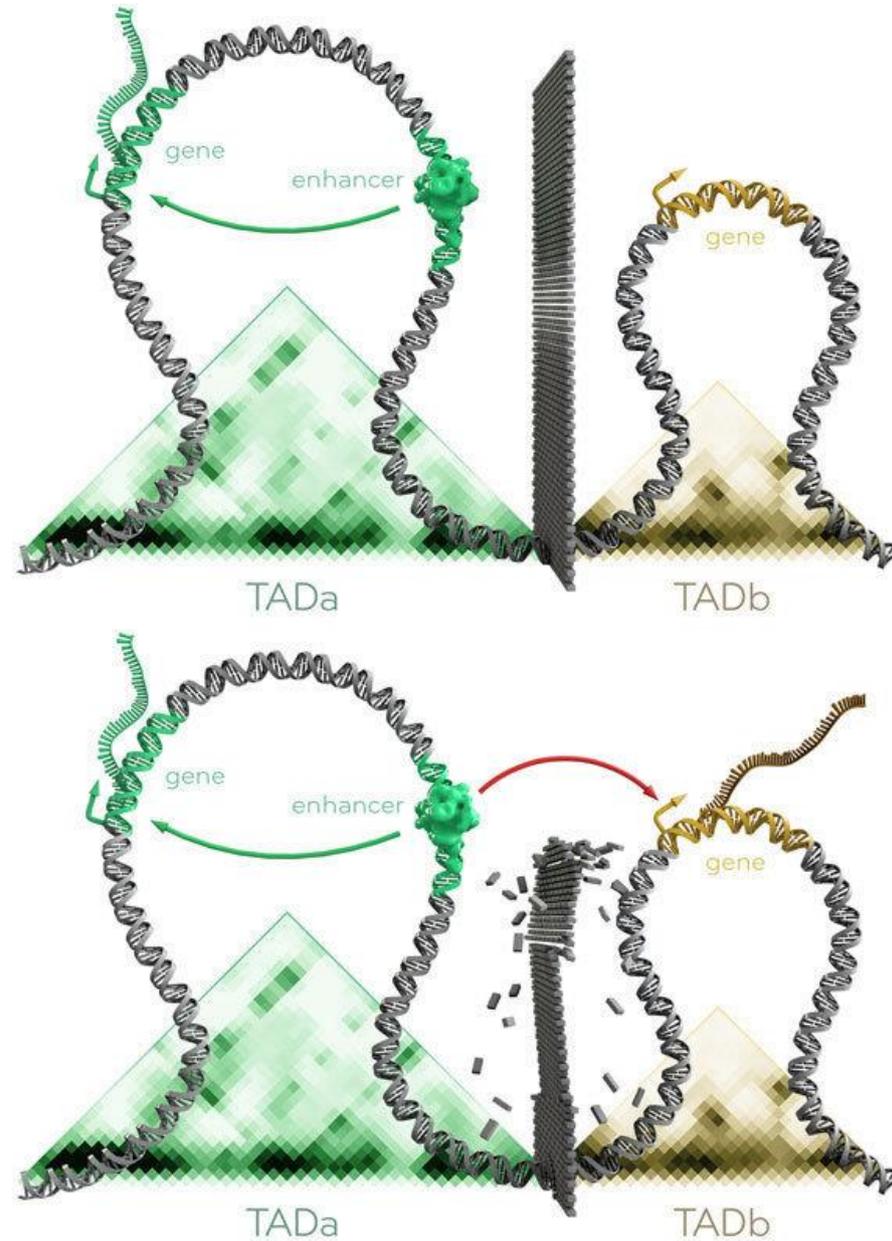
Iteratively corrected heatmap



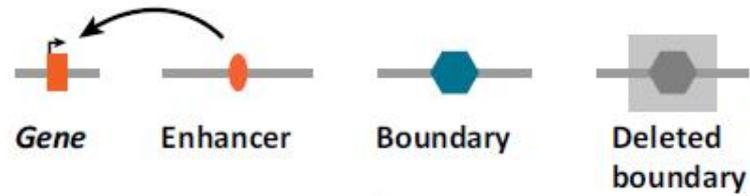
Топологические домены (ТАДы)



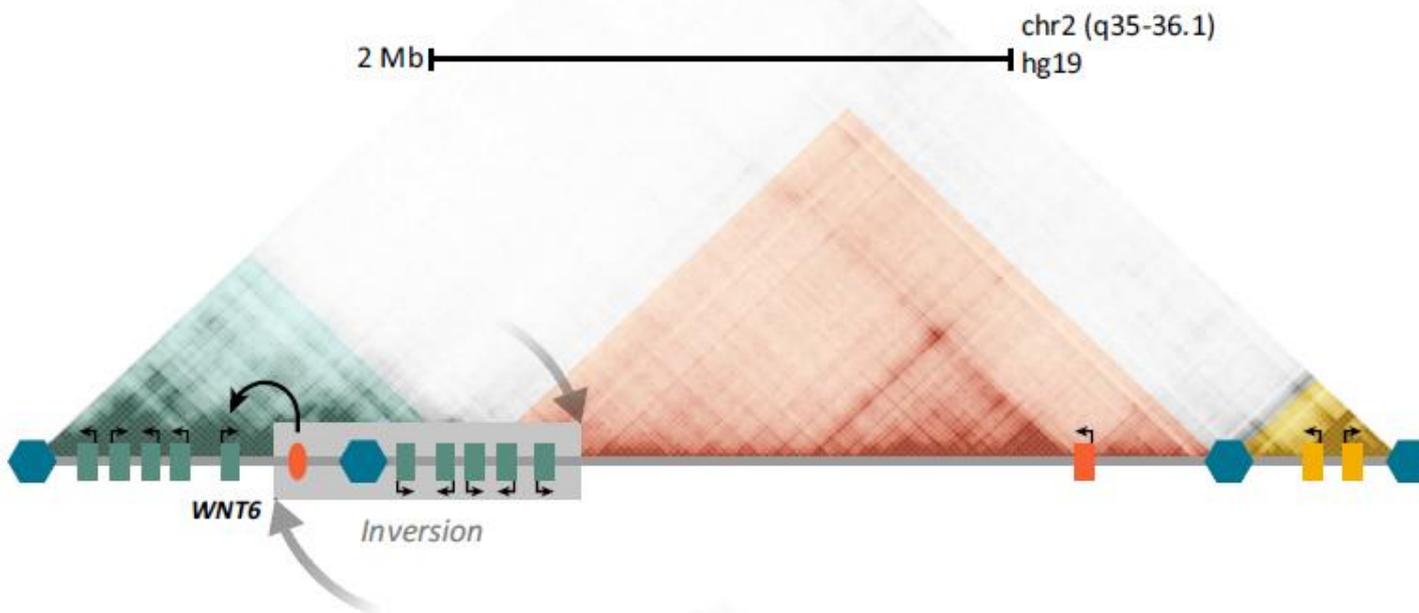
ТАДы и болезни



ТАДы и синдактилия

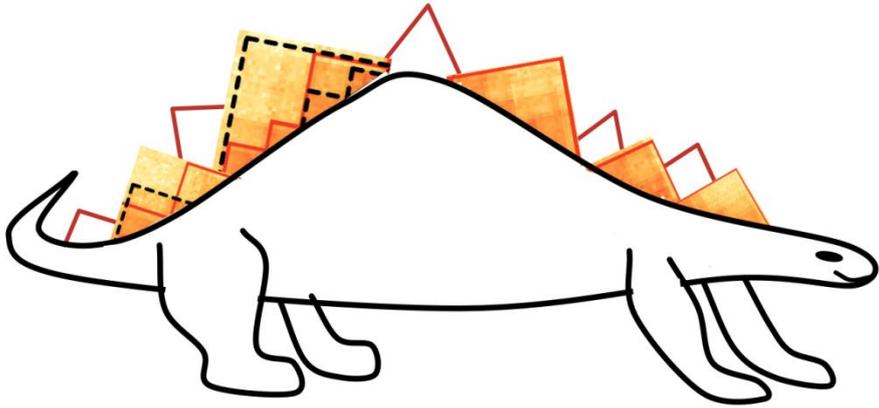


(A)



F-syndrome

Автоматизированное предсказание ТАДов

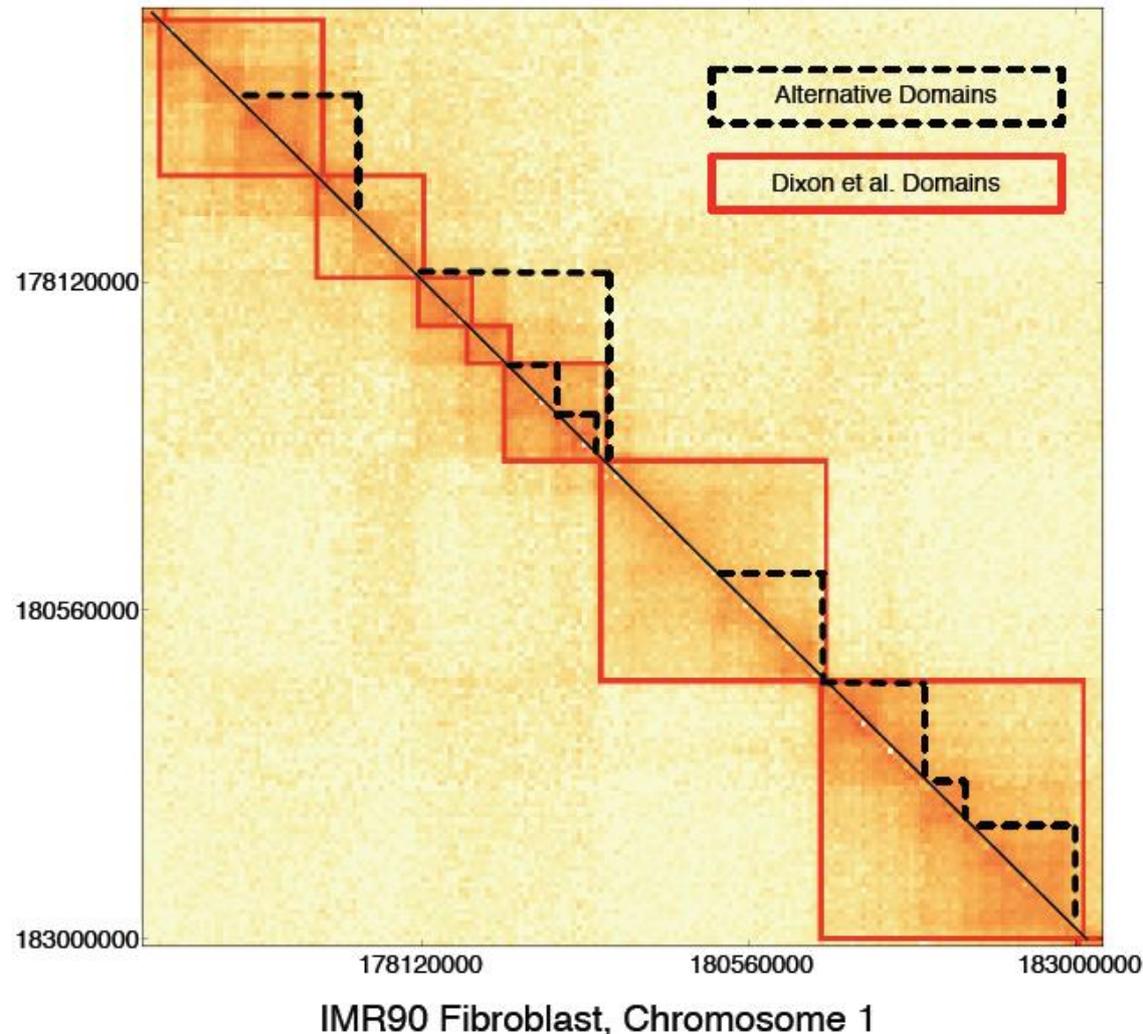


<https://github.com/kingsfordgroup/armatus>

Armatus – программа предсказания ТАДов по матрицам из C-методов.

Позволяет получить несколько разметок ТАДов с разной средней длиной ТАДа.

Иерархическая структура ТАДов: ТАДы вложены друг в друга.

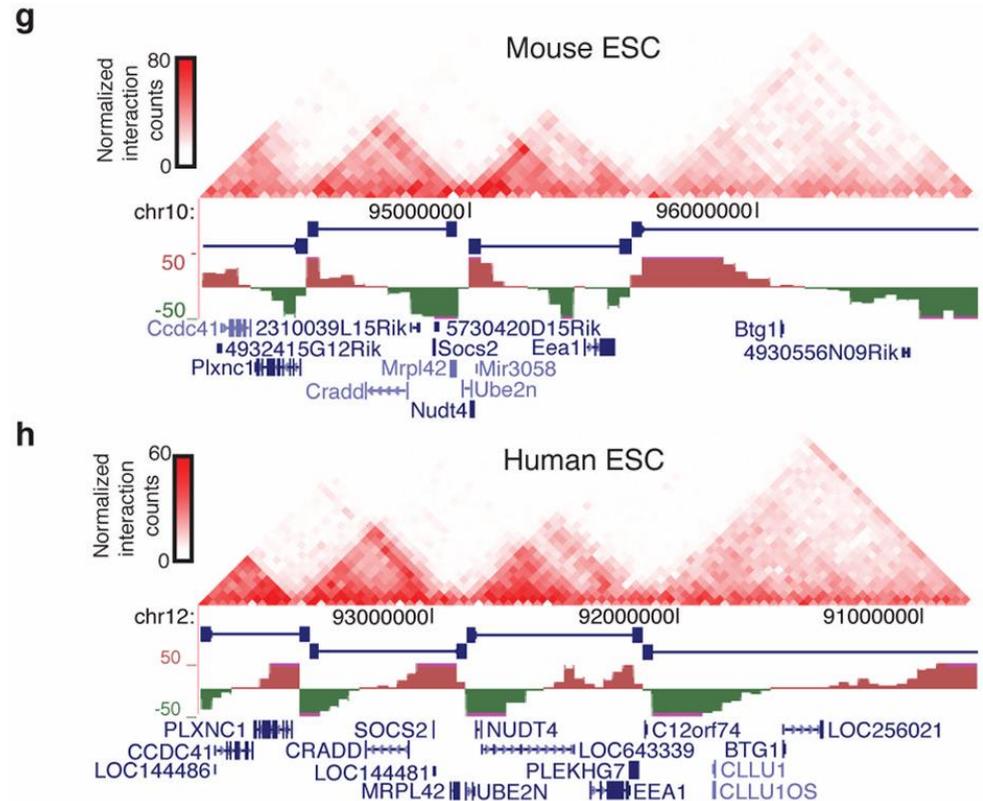
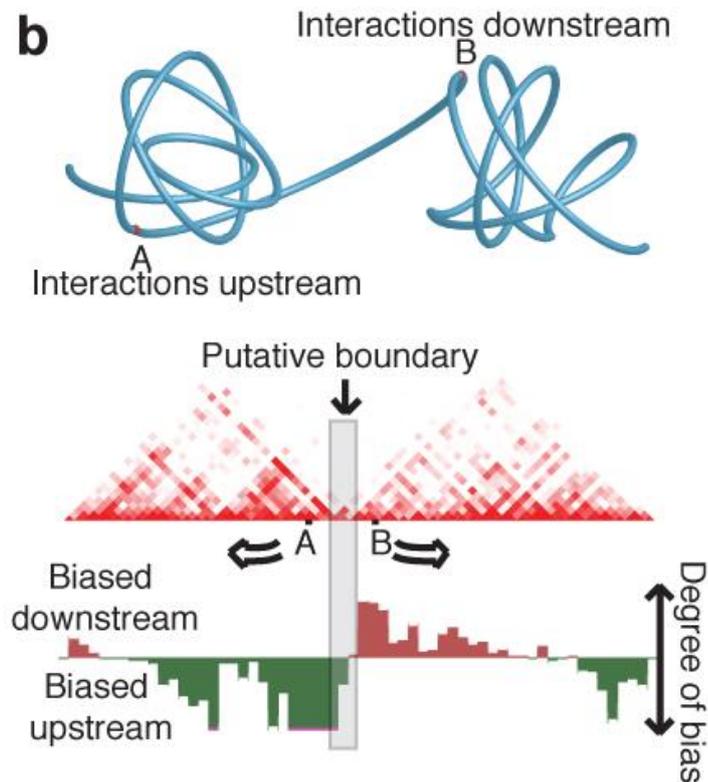


Какую науку делают с этими данными?

Мышь и человек

Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature **485**, 376–80 (2012).

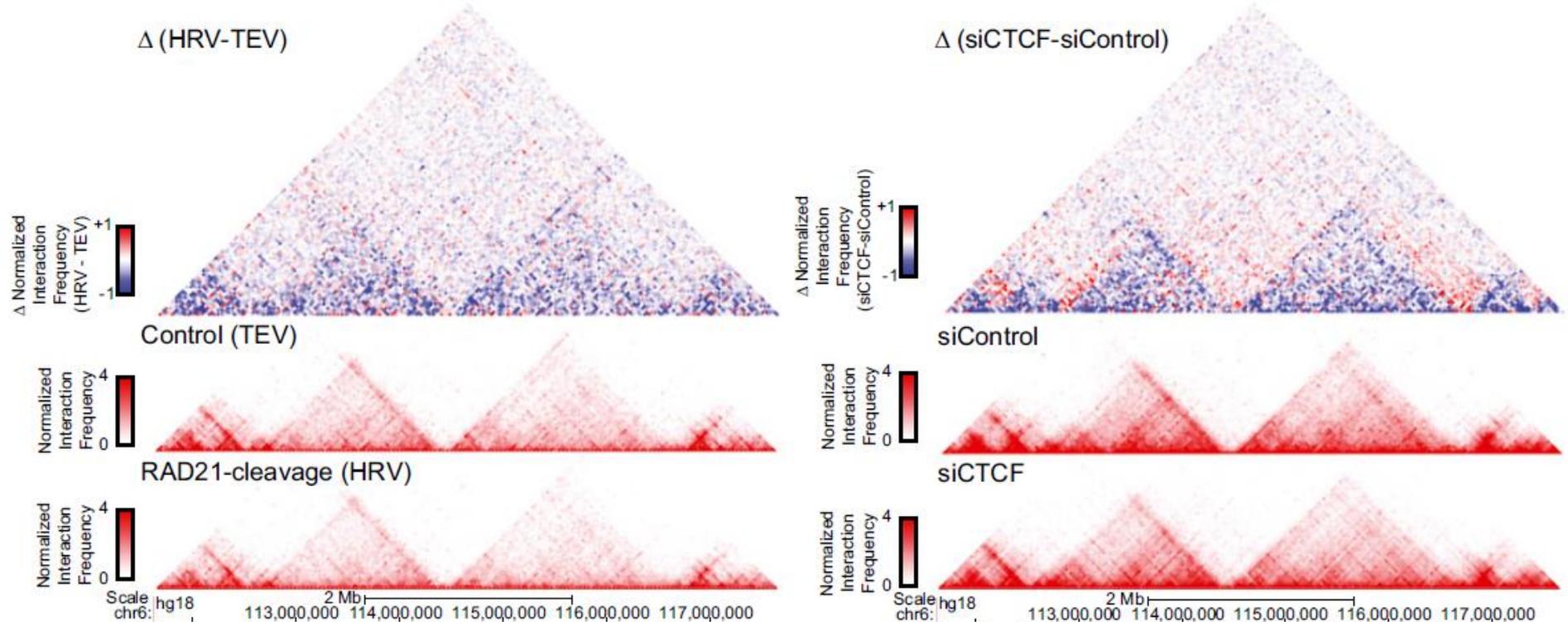
- Human and mouse Hi-C in embryonic and differentiated cell types.
- TAD. “Directionality Index” to quantify the degree of US or DS interaction bias.
- TADs are stable across different cell types and highly conserved across species.
- Boundaries are enriched for CTCF, housekeeping genes, tRNAs, and SINE.



Человек

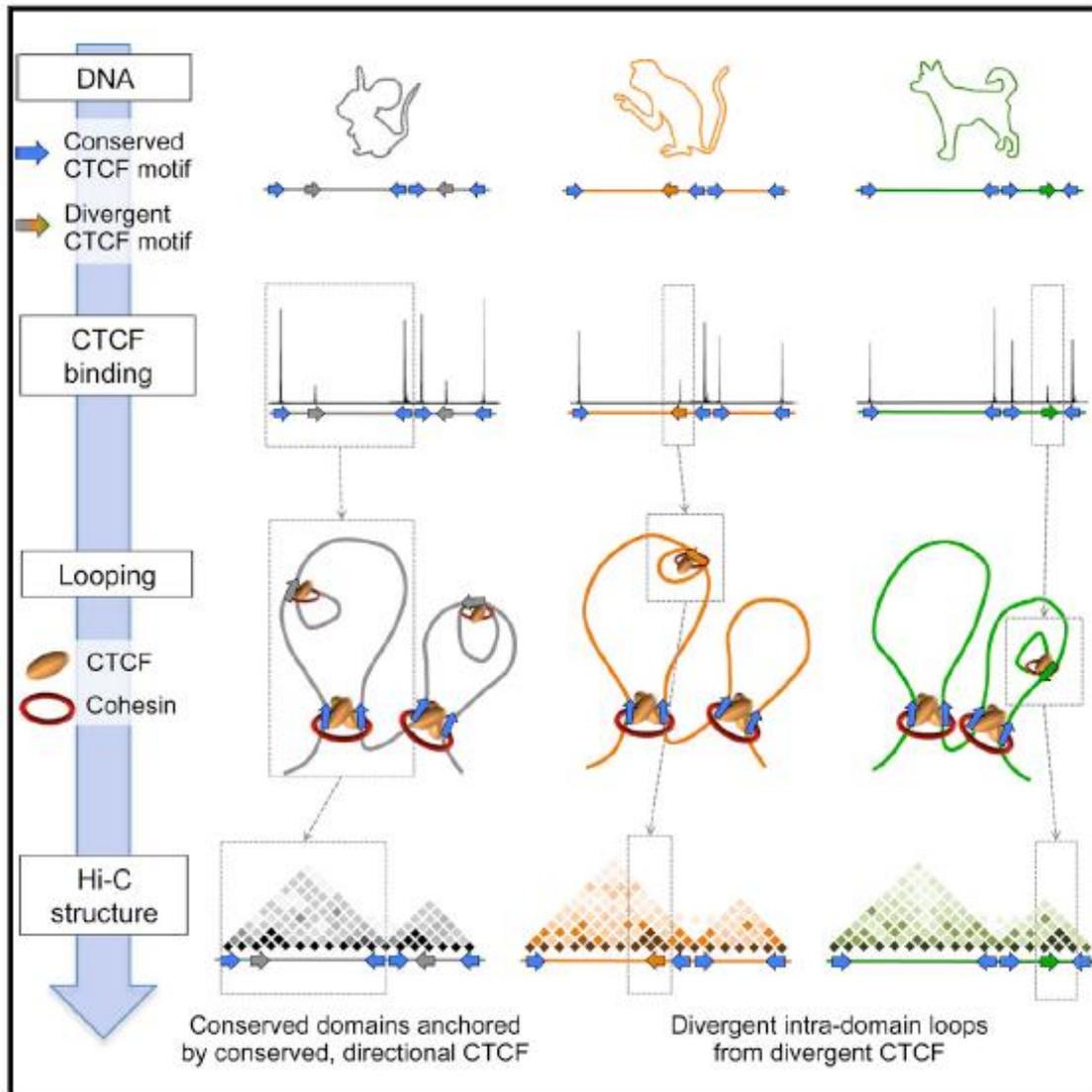
Zuin, J. et al. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc. Natl. Acad. Sci. U. S. A. 111, 996–1001 (2014).

- A general loss of local chromatin interactions upon disruption of cohesin, but the topological domains remain intact.
- Depletion of CTCF not only reduced intra-domain interactions but also increased inter-domain interactions.
- Distinct groups of genes become misregulated upon depletion of cohesin and CTCF.



Сравнение между организмами

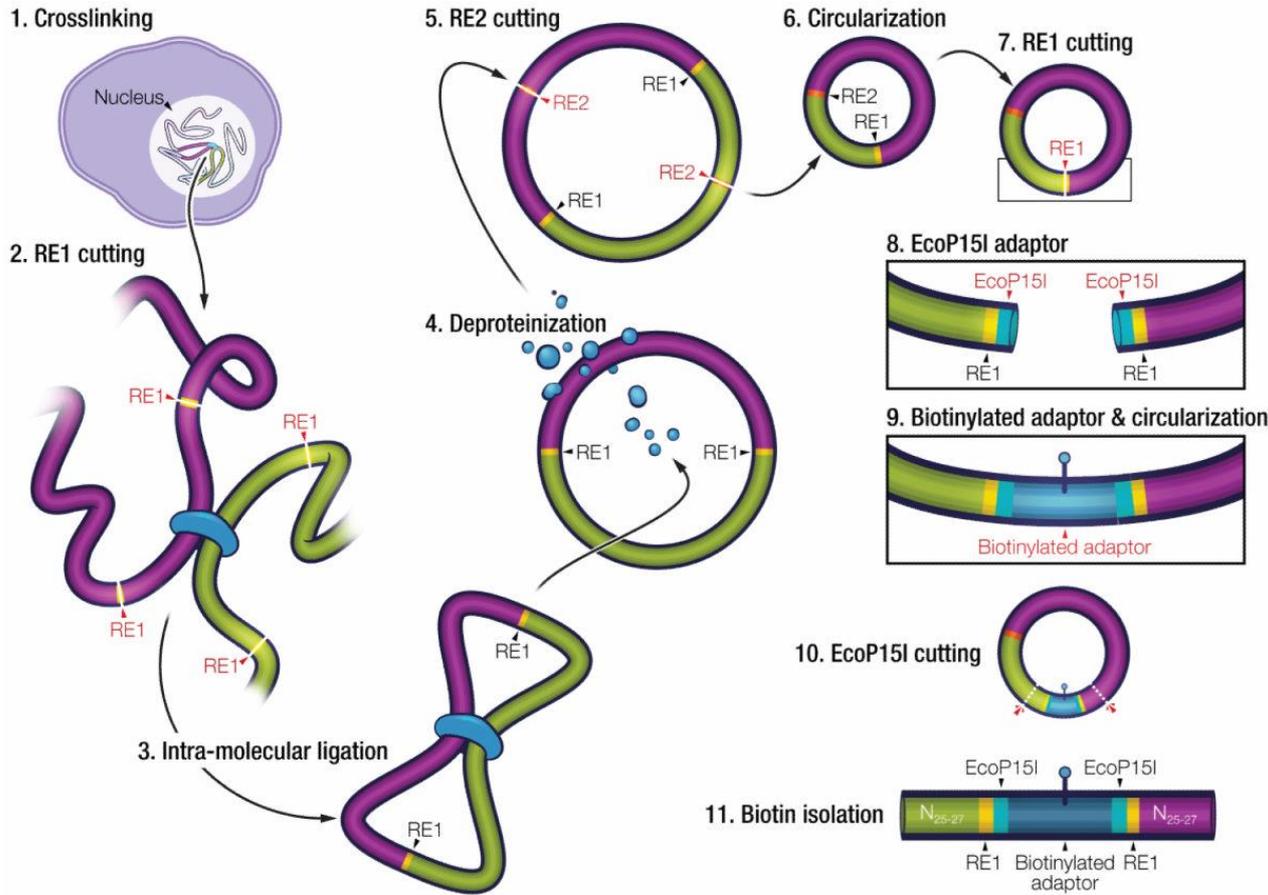
Vietri Rudan, M. et al. Comparative Hi-C Reveals that CTCF Underlies Evolution of Chromosomal Domain Architecture. Cell Rep. 10, 1297–1309 (2015).



- Multi-species Hi-C comparisons reveal robust conservation of chromosome organization.
- Conserved CTCF sites are co-localized with cohesion and enriched at strong TAD borders
- Divergent CTCF binding between species is correlated with divergence of internal domain structure.
- Large-scale domains are reorganized during genome evolution as intact modules.

Дрожжи

Duan, Z. et al. A three-dimensional model of the yeast genome. Nature **465**, 363–7 (2010).



- 4C-seq (HindIII and EcoRI).
- Rab1 configuration.

6 bp-cutter RE1 for the 3C-step digestion
4 bp-cutter RE2 for the 4C-step digestion

