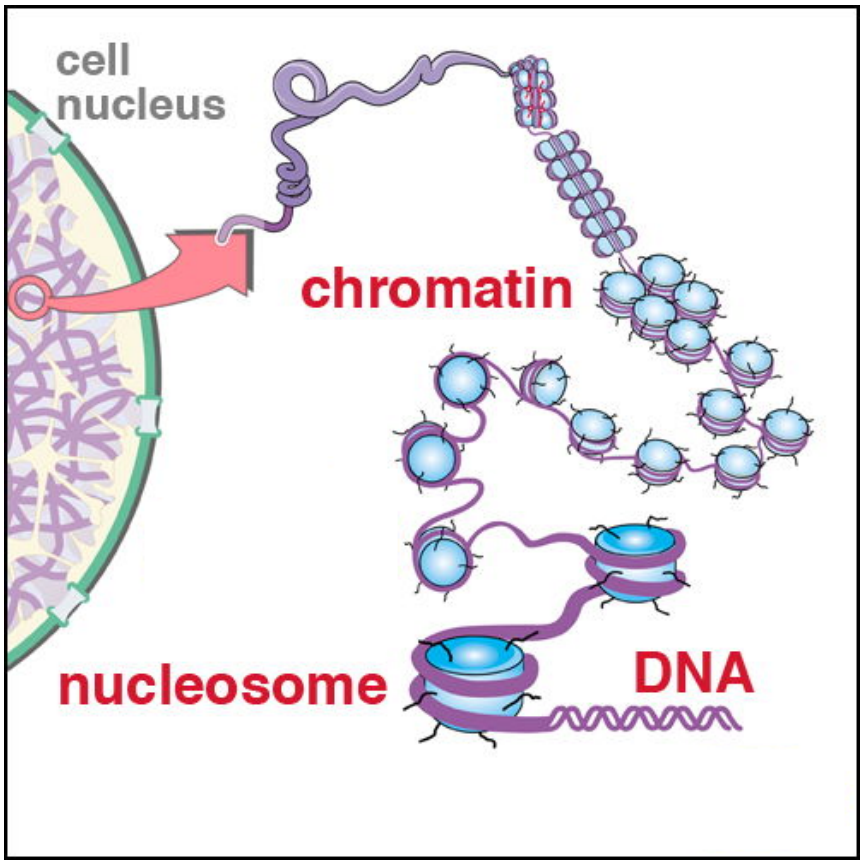
3D--структура хроматина и технологии ее определения



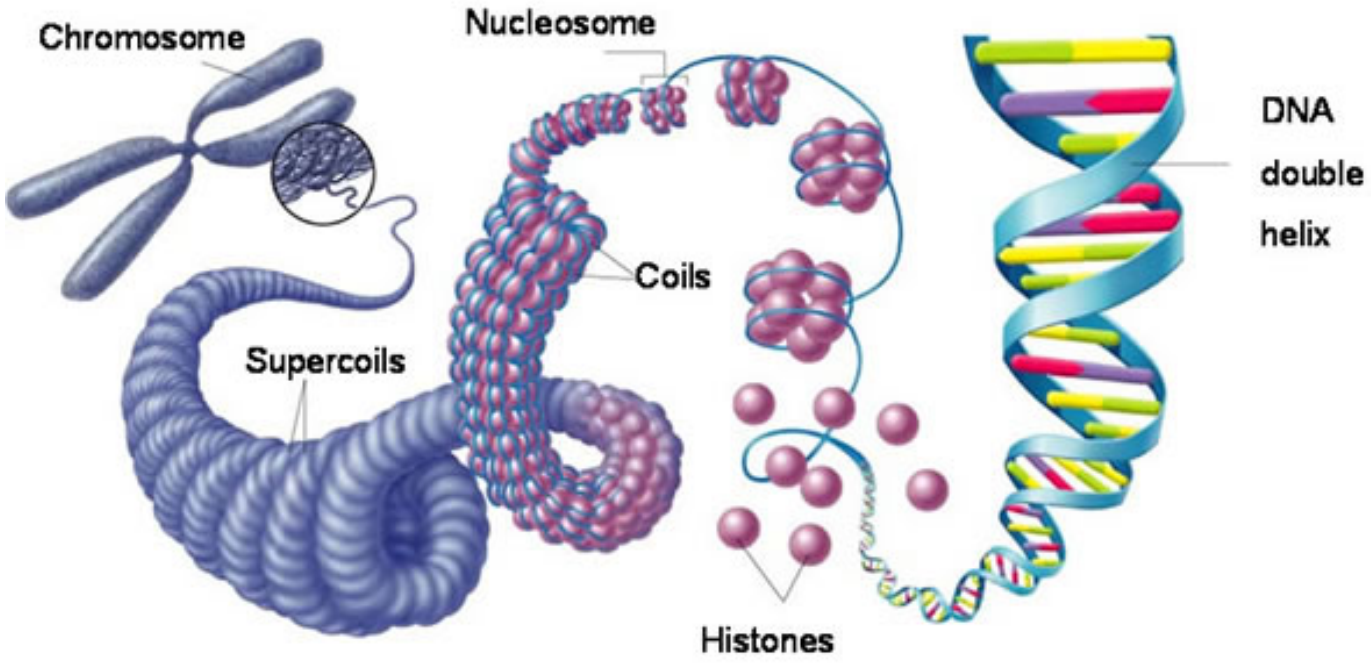
Екатерина Храмеева

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

План лекции

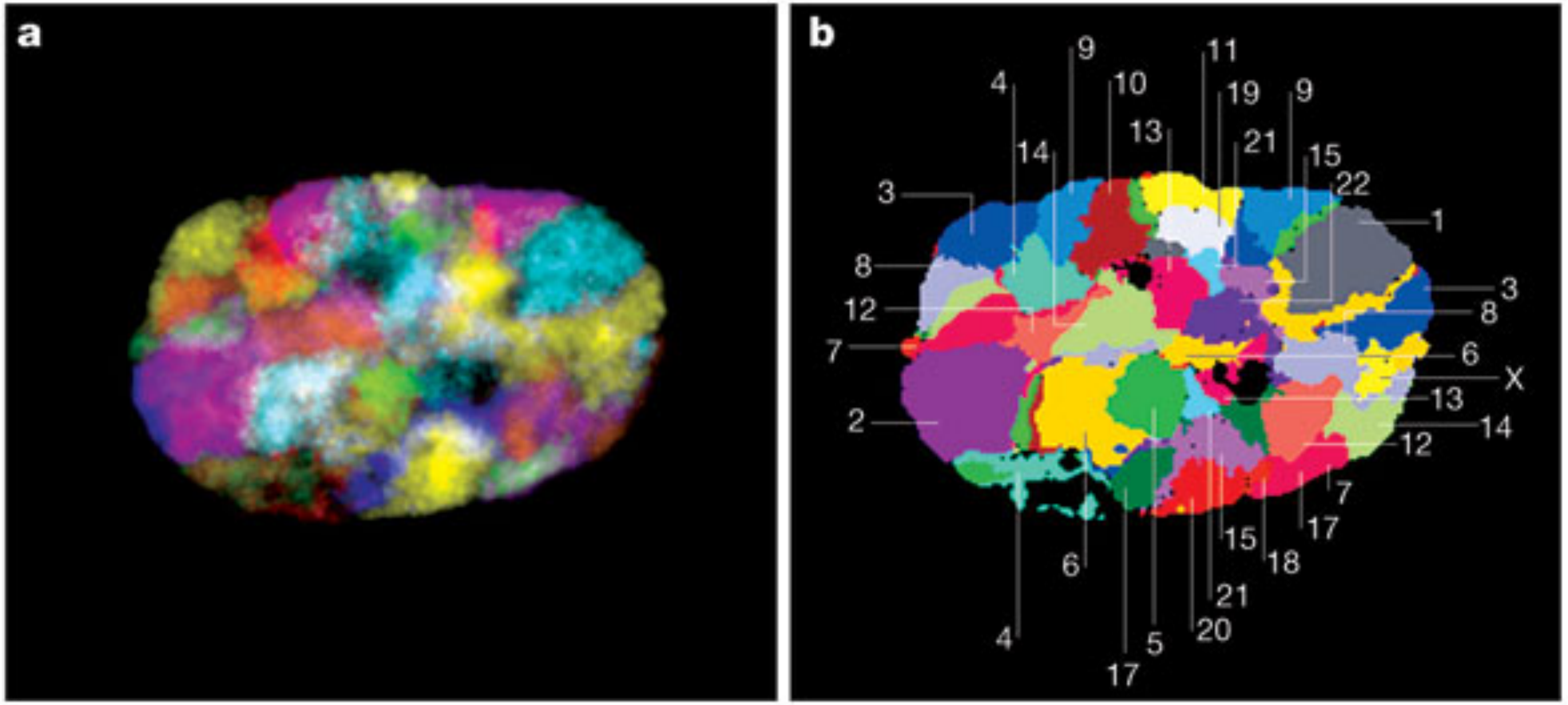
- Принципы организации хроматина
- Экспериментальные методы
  - Обзор 3C-методов
  - Секвенирование
  - Метод Hi-C
- Биоинформатический анализ
  - Анализ HiC данных
  - Построение теплокарт
  - Предсказание топологических доменов
- Какую науку делают с этими данными?

Структура хромосом



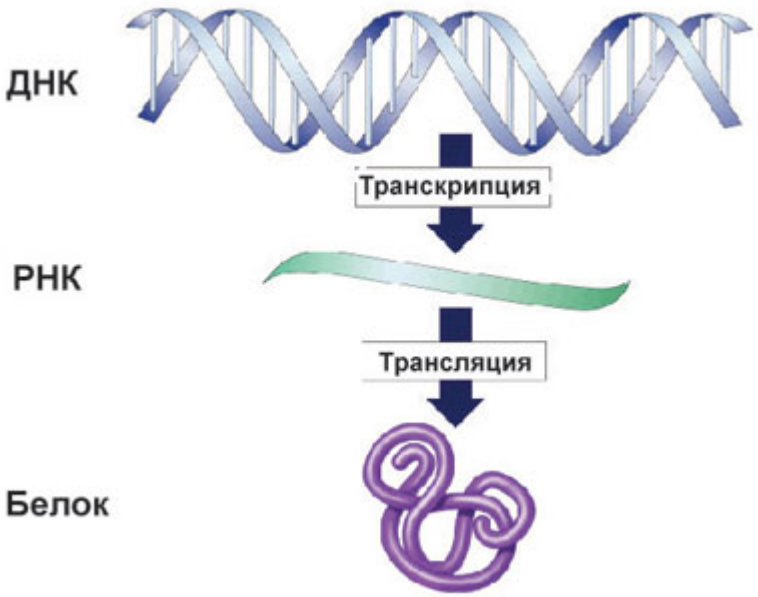
- Хромосомы состоят из ДНК и белков
- ДНК намотана на специальные белки – *гистоны*, из которых формируются *нуклеосомы*
- Нитки нуклеосом скручиваются и суперскручиваются, образуя *хромосомы*

Микрофотография всех хромосом в ядре клетки человека



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

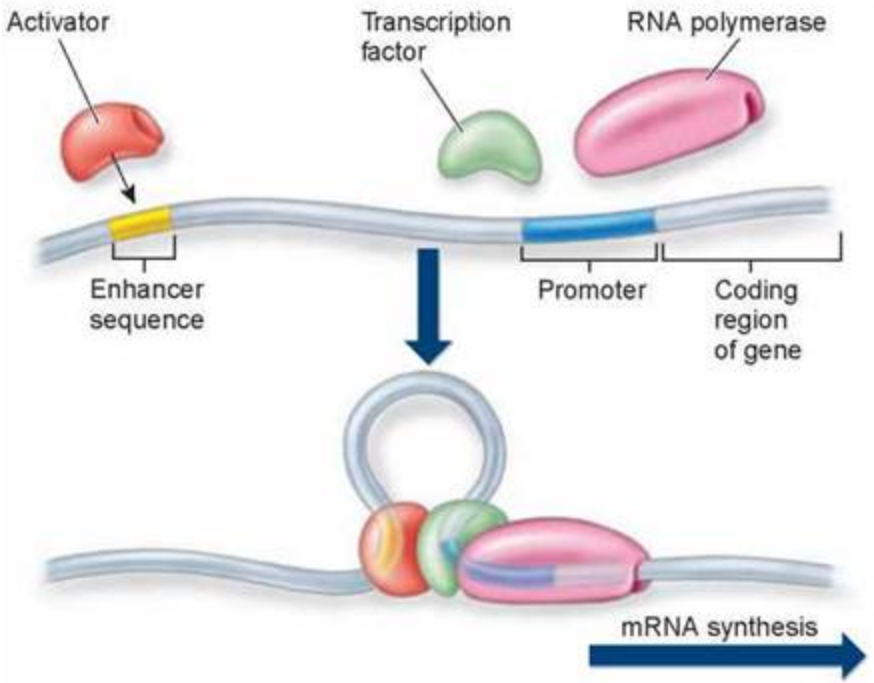
Не только упаковка



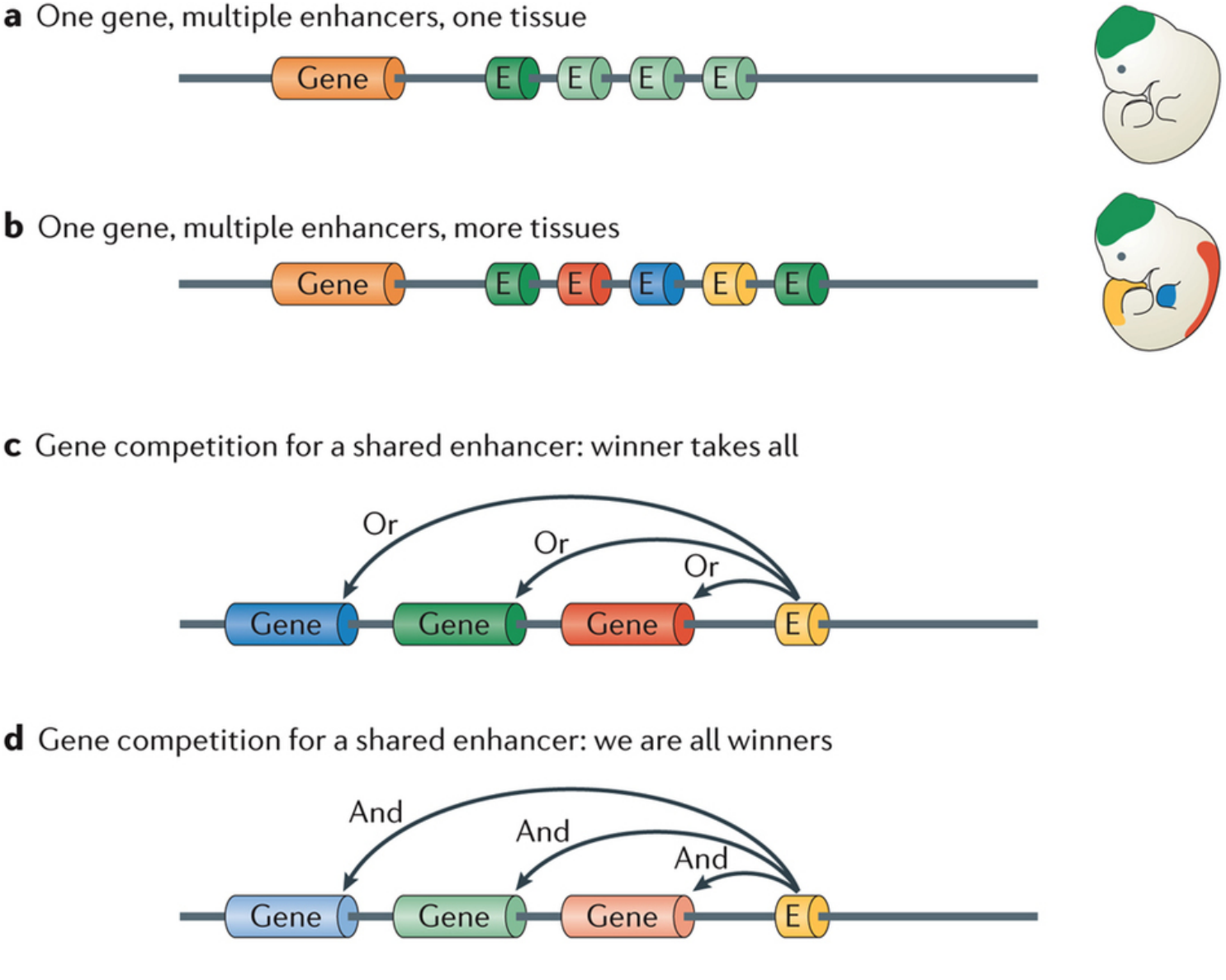
В нашем геноме:

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

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

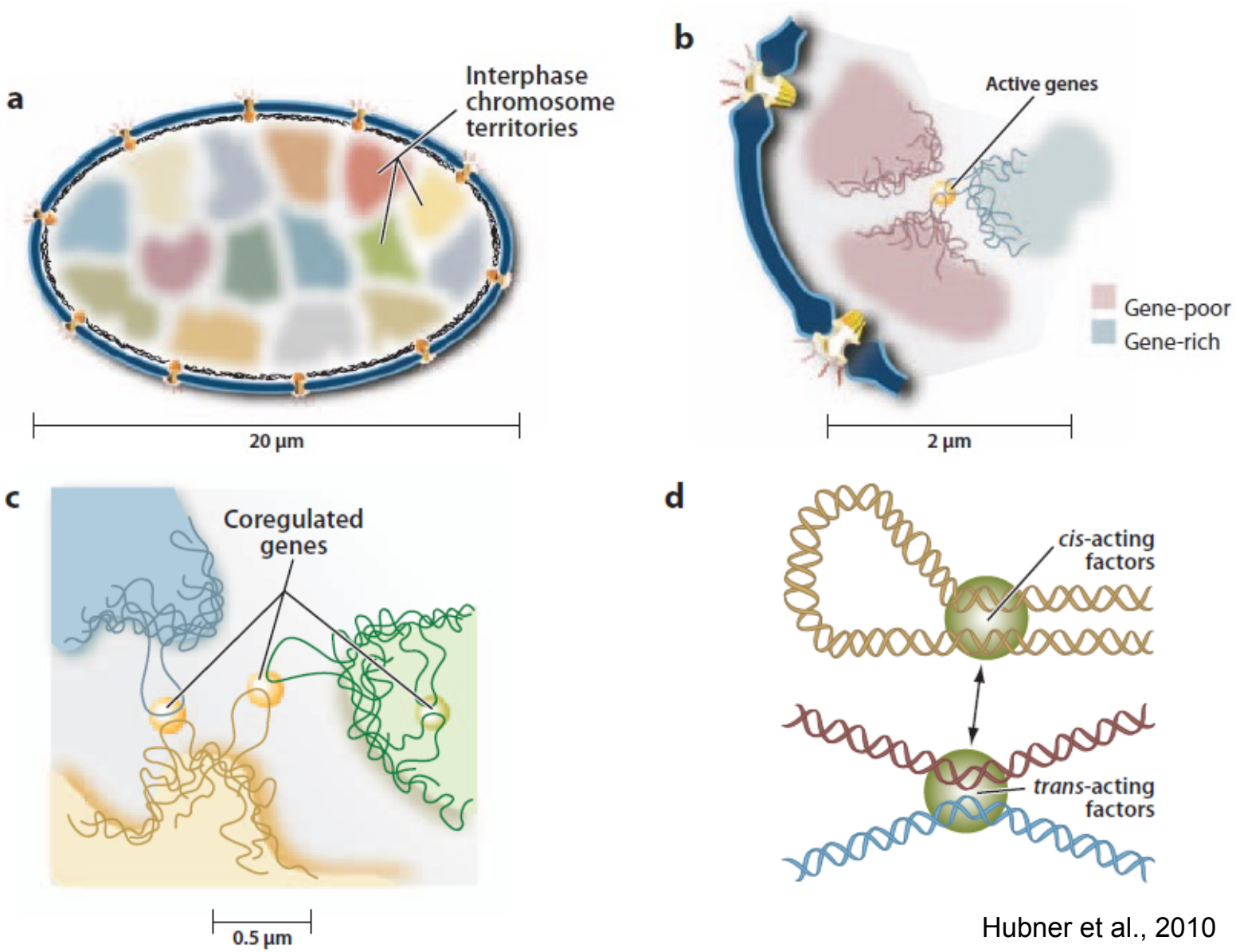


Регуляция работы генов в тканях



Krijger, P. H. L. & de Laat, W. *Nat. Rev. Mol. Cell Biol.* (2016)

Принципы организации хроматина



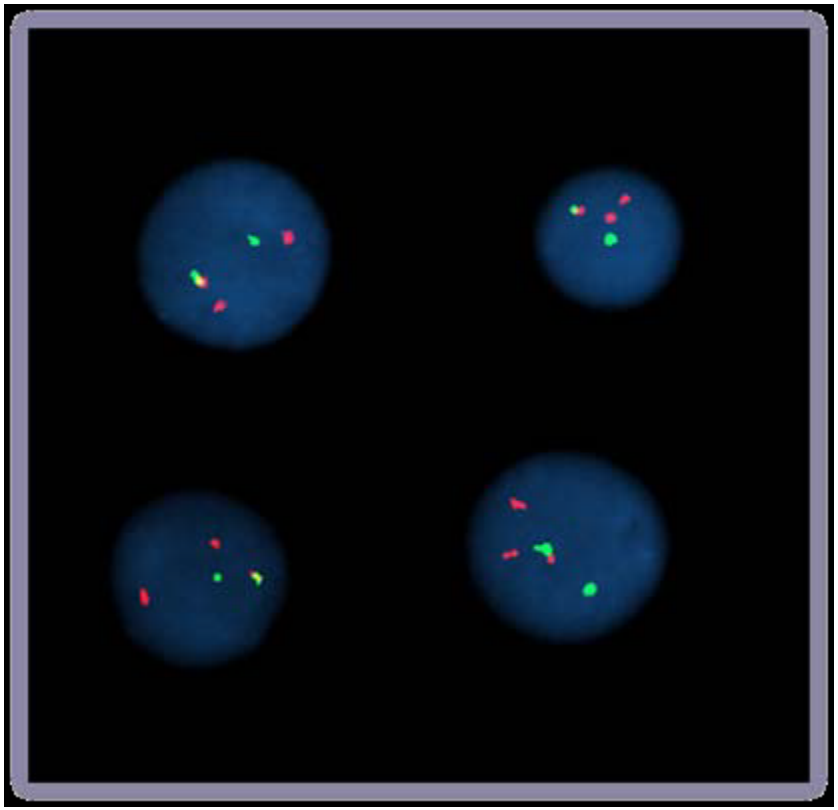
Hubner et al., 2010

Как изучают 3D структуру генома

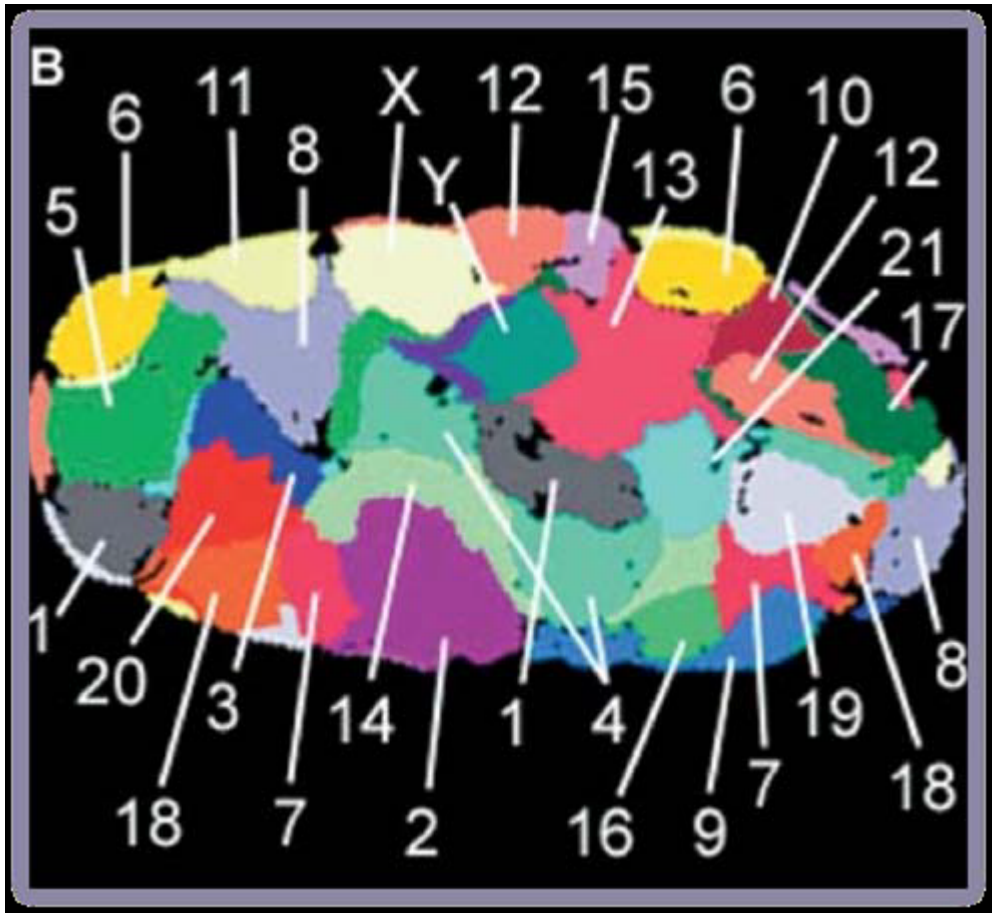
- Микроскопия
- FISH (DNA fluorescence in situ hybridization)
- DamID – позволяет определить, какие участки хроматина находятся на периферии ядра
- 3C-методы:
  - 3C (chromosome conformation capture)
  - 4C (chromosome conformation capture-on-chip)
  - 5C (chromosome conformation capture carbon copy)
  - **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

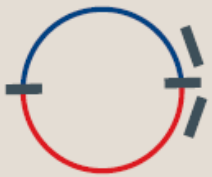
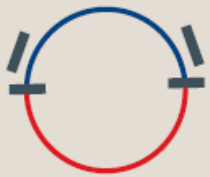
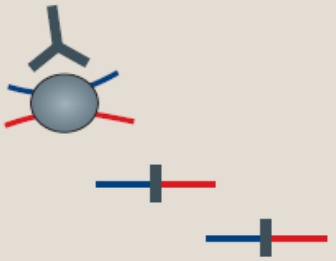
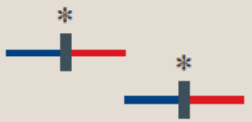


Bolzer et al., PLoS Biol. 2005

3C-методы

a 3C: converting chromatin interactions into ligation products

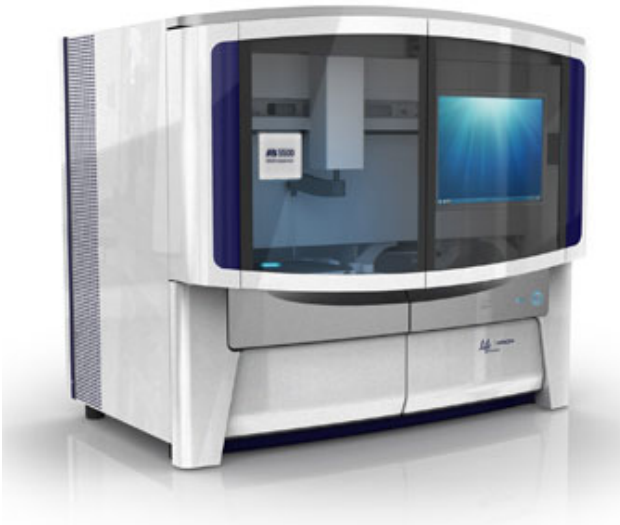
b Ligation product detection methods

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
			• DNA shearing • Immunoprecipitation 	• Biotin labelling of ends • DNA shearing 
PCR or sequencing	Inverse PCR sequencing	Multiplexed LMA sequencing	Sequencing	Sequencing

J. Dekker, M.A. Marti-Renom, L.A. Mirny,
Nature Reviews Genetics, 2013

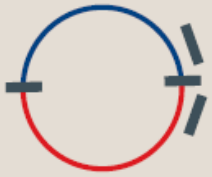
Что такое секвенирование?

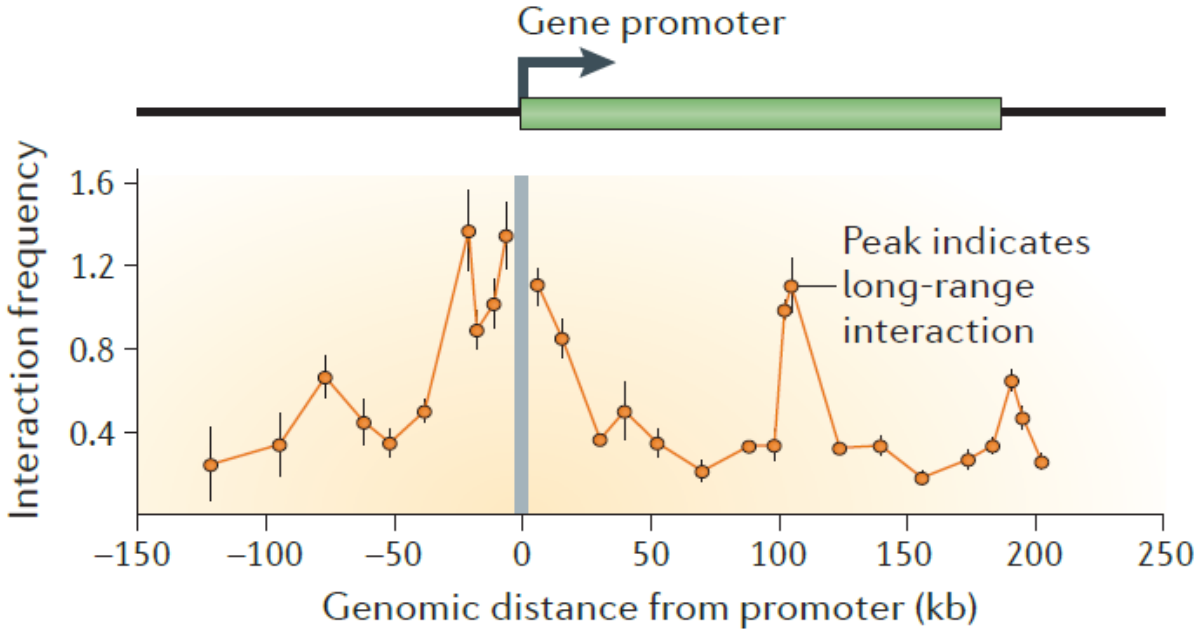
- Секвенировать = установить последовательность нуклеотидов в молекуле ДНК.
- Сначала ДНК надо выделить.
- Нельзя прочесть молекулу ДНК целиком, поэтому ДНК режут на кусочки в случайных местах и «размножают».



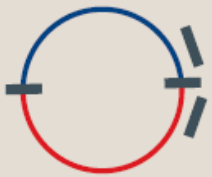
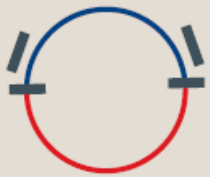
Метод 3C

Пример 3C данных:

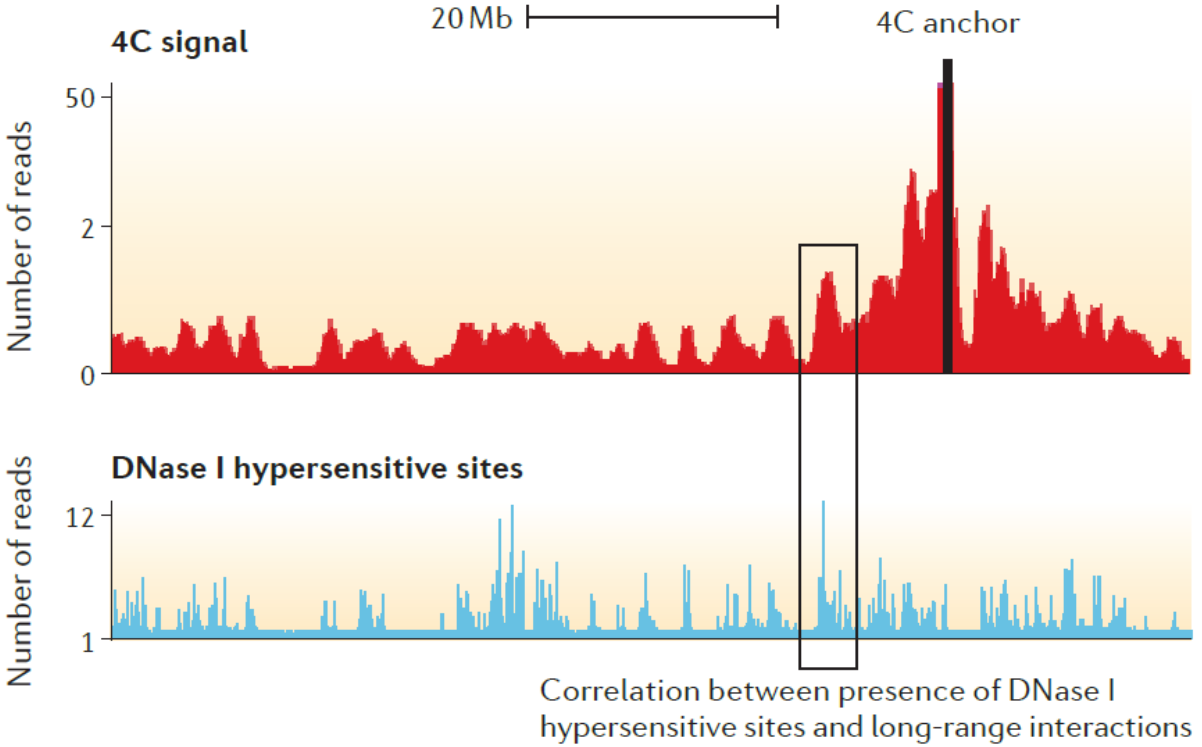
3C
One-by-one All-by-all

PCR or sequencing



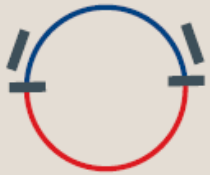
Метод 4C

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

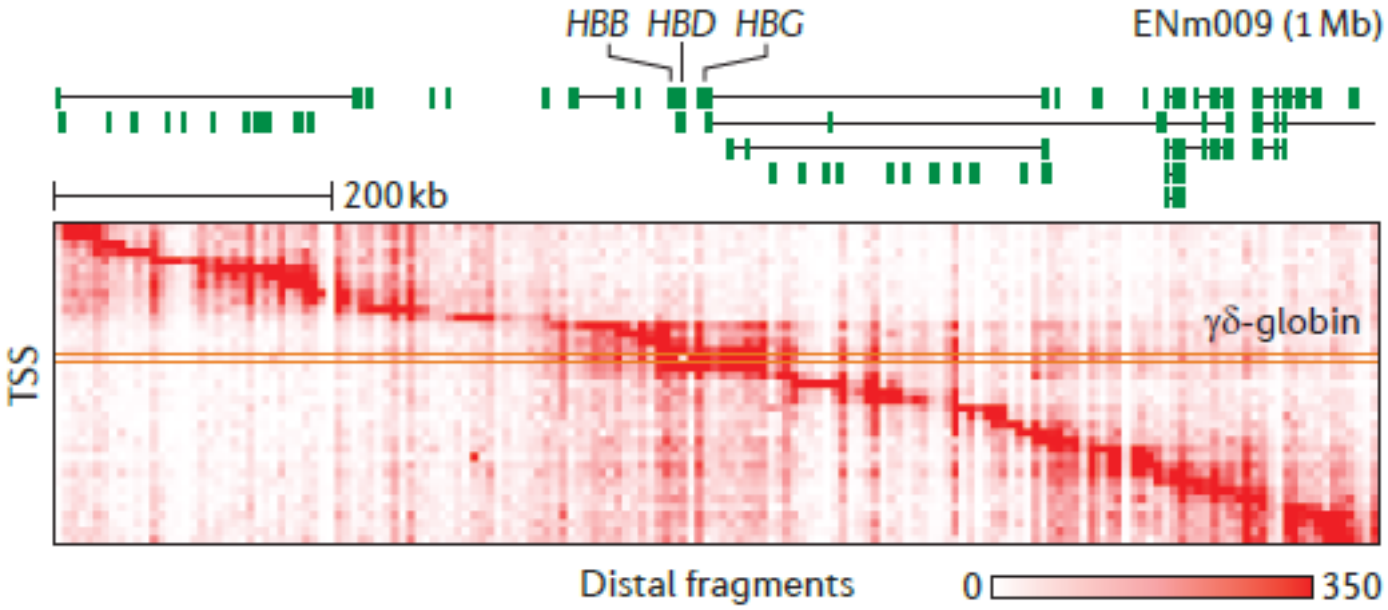
Пример 4C данных:

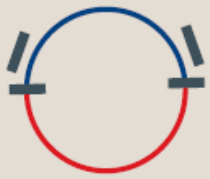


Метод 5C

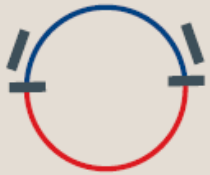
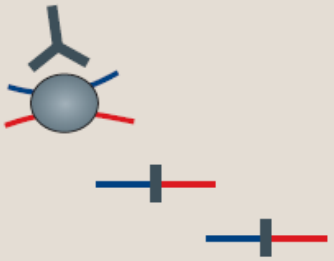
3C	4C	5C
One-by-one All-by-all	One-by-all	Many-by-many
		
PCR or sequencing	Inverse PCR sequencing	Multiplexed LMA sequencing

Пример 5C данных:

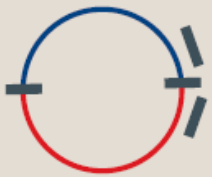
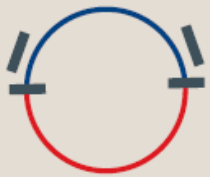
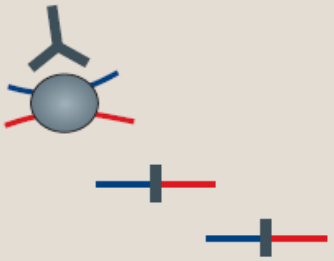
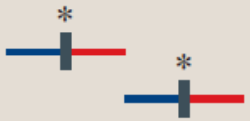


3C	4C	5C
One-by-one All-by-all	One-by-all	Many-by-many
		
PCR or sequencing	Inverse PCR sequencing	Multiplexed LMA sequencing

Метод ChIA-PET

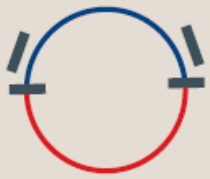
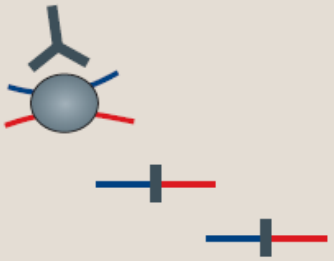
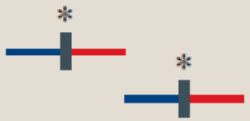
3C	4C	5C	ChIA–PET
One-by-one All-by-all	One-by-all	Many-by-many	Many-by-many
			• DNA shearing • Immunoprecipitation 
PCR or sequencing	Inverse PCR sequencing	Multiplexed LMA sequencing	Sequencing

Метод Hi-C

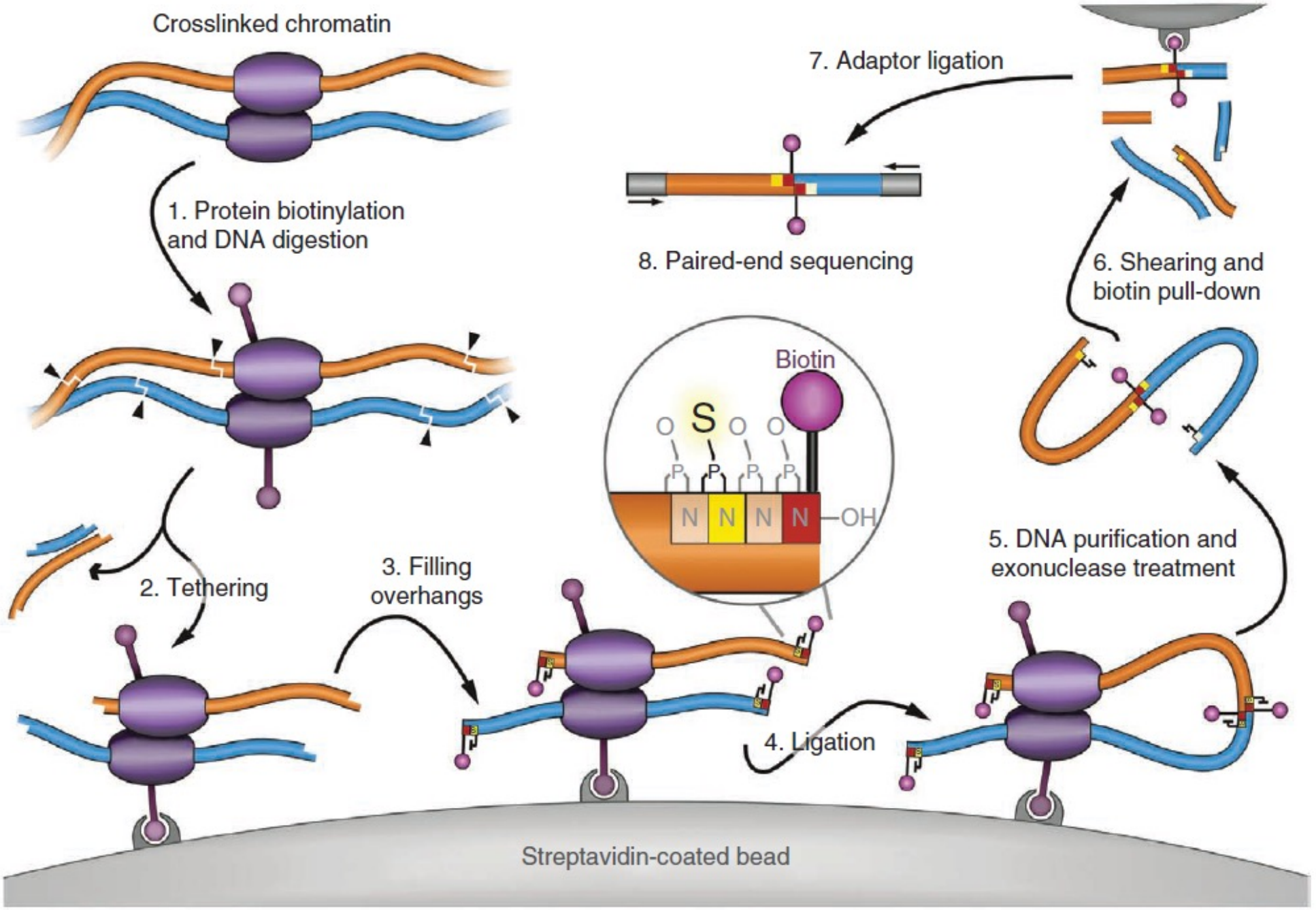
3C	4C	5C	ChIA–PET	Hi-C
One-by-one All-by-all	One-by-all	Many-by-many	Many-by-many	All-by-all
			• DNA shearing • Immunoprecipitation 	• Biotin labelling of ends • DNA shearing 
PCR or sequencing	Inverse PCR sequencing	Multiplexed LMA sequencing	Sequencing	Sequencing

Метод 3C-seq

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

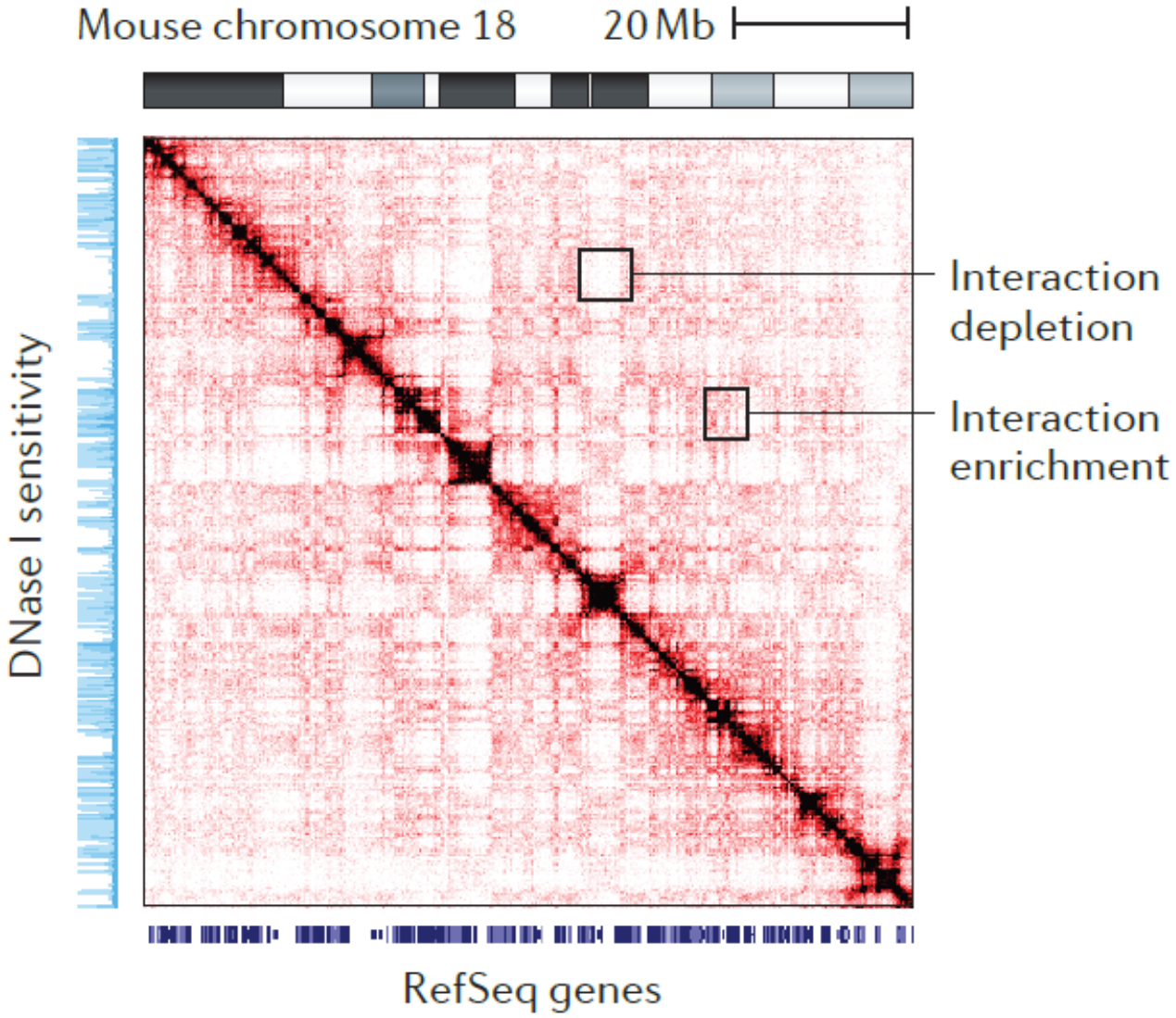
3C	4C	5C	ChIA–PET	Hi-C
One-by-one All-by-all	One-by-all	Many-by-many	Many-by-many	All-by-all
			• DNA shearing • Immunoprecipitation 	• Biotin labelling of ends • DNA shearing 
PCR or sequencing	Inverse PCR sequencing	Multiplexed LMA sequencing	Sequencing	Sequencing

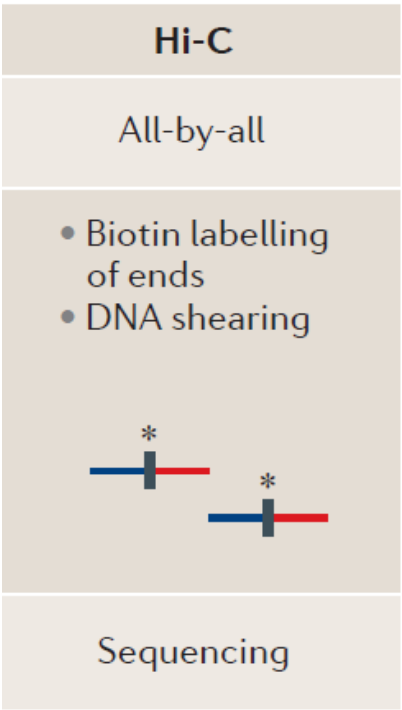
Метод ТСС


Crosslinked chromatin
1. Protein biotinylation and DNA digestion
2. Tethering
3. Filling overhangs
4. Ligation
5. DNA purification and exonuclease treatment
6. Shearing and biotin pull-down
7. Adaptor ligation
8. Paired-end sequencing
Biotin
S
O
P
N
OH
Streptavidin-coated bead

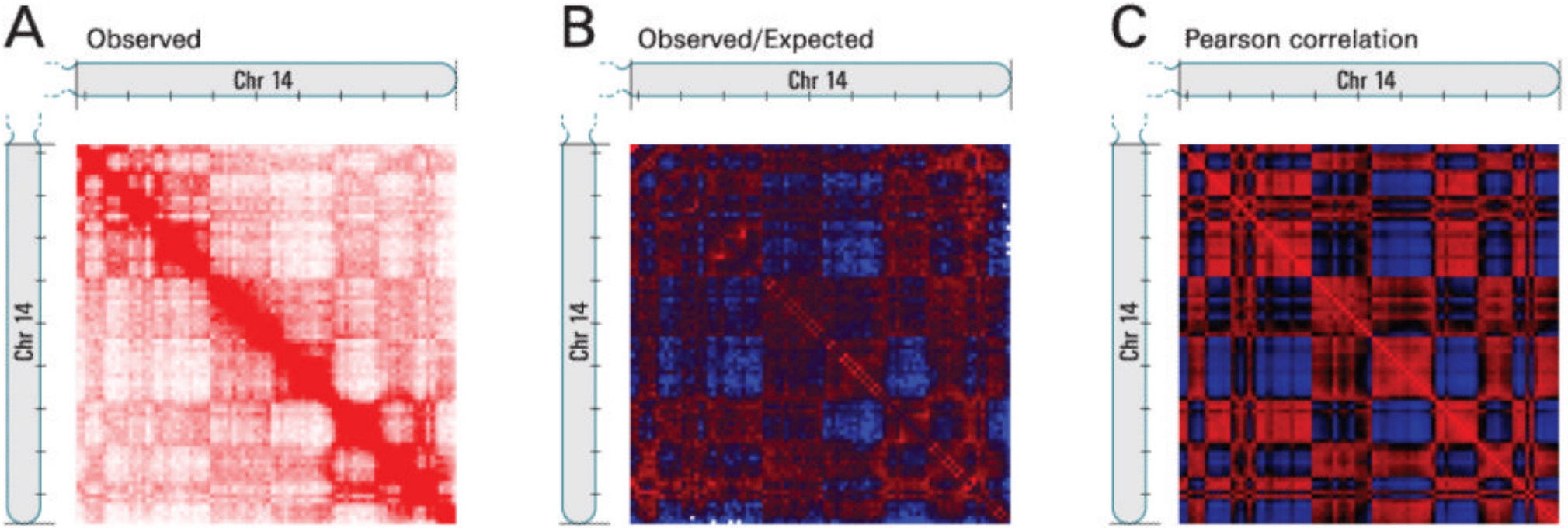
Метод Hi-C

Пример Hi-C данных:

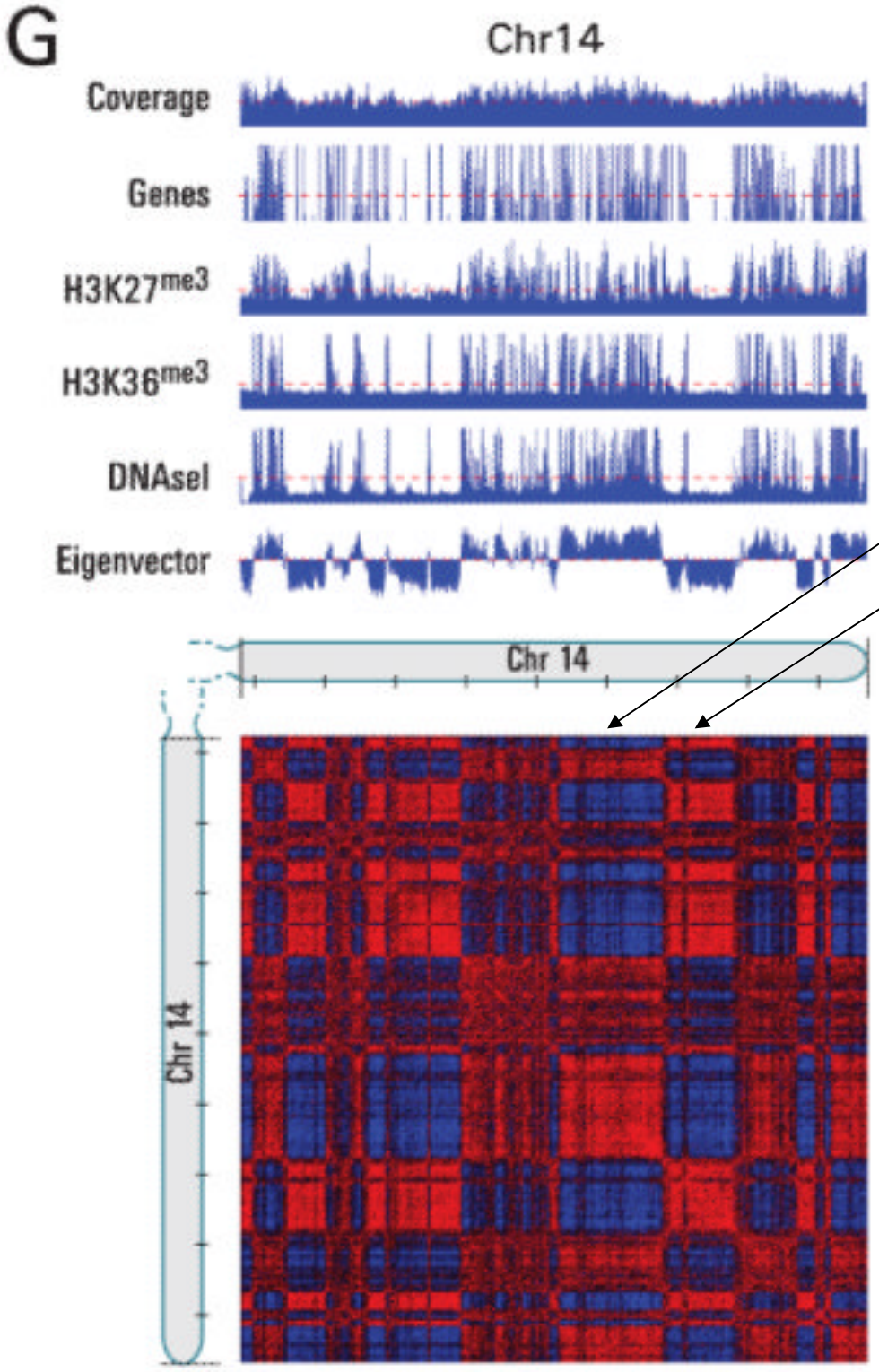




Матрица контактов Hi-C



Lieberman-Aiden et al. 2009



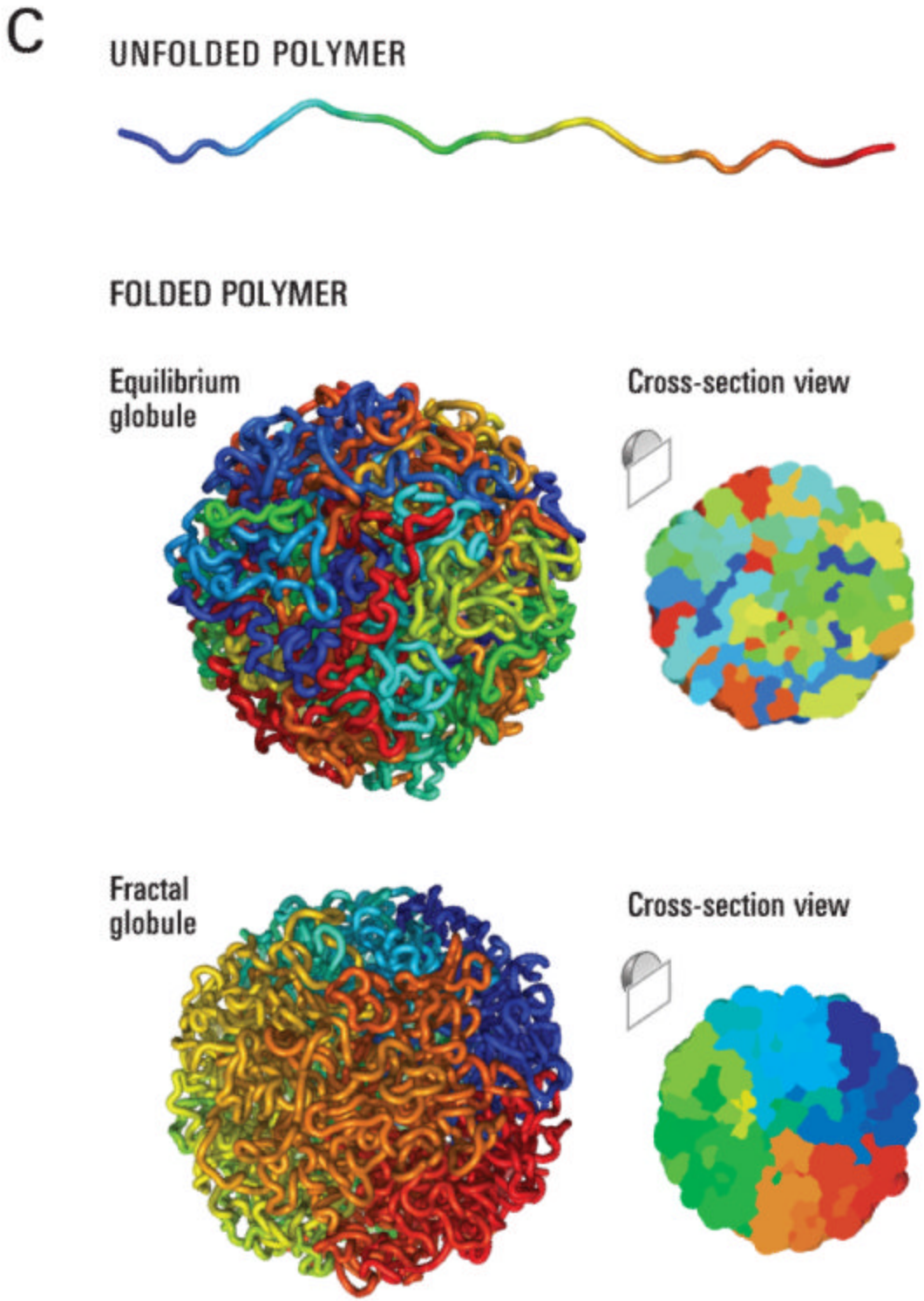
Первая главная компонента (собственный вектор, или eigenvector) коррелирует с плотностью генов и маркерами открытого хроматина.

Компартмент A
Компартмент B

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

Lieberman-Aiden et al. 2009

Фрактальная глобула

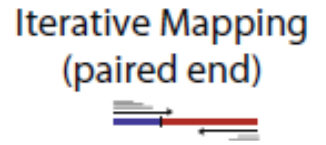


Lieberman-Aiden et al. 2009

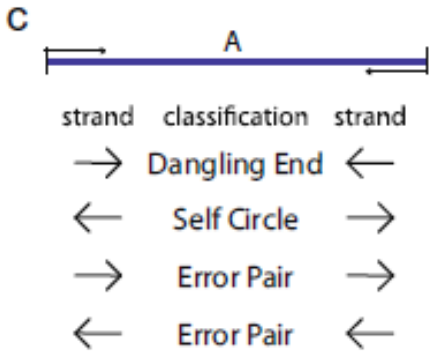
Биоинформатический анализ

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.

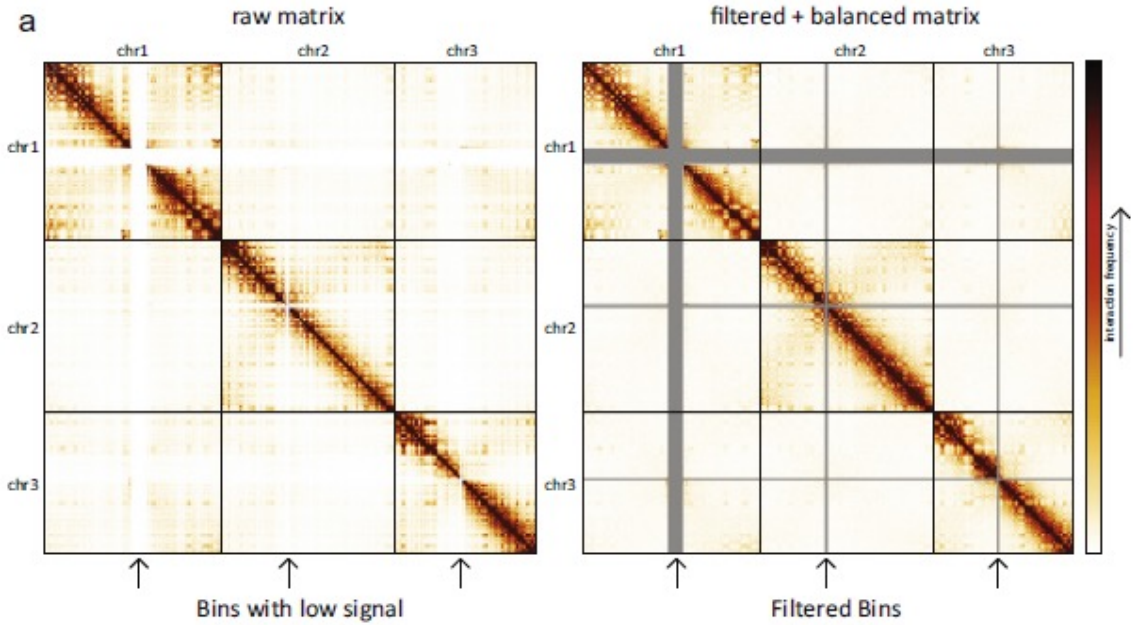


3. Fragment filtering
PCR duplicates, undigested restriction sites.

4. Binning

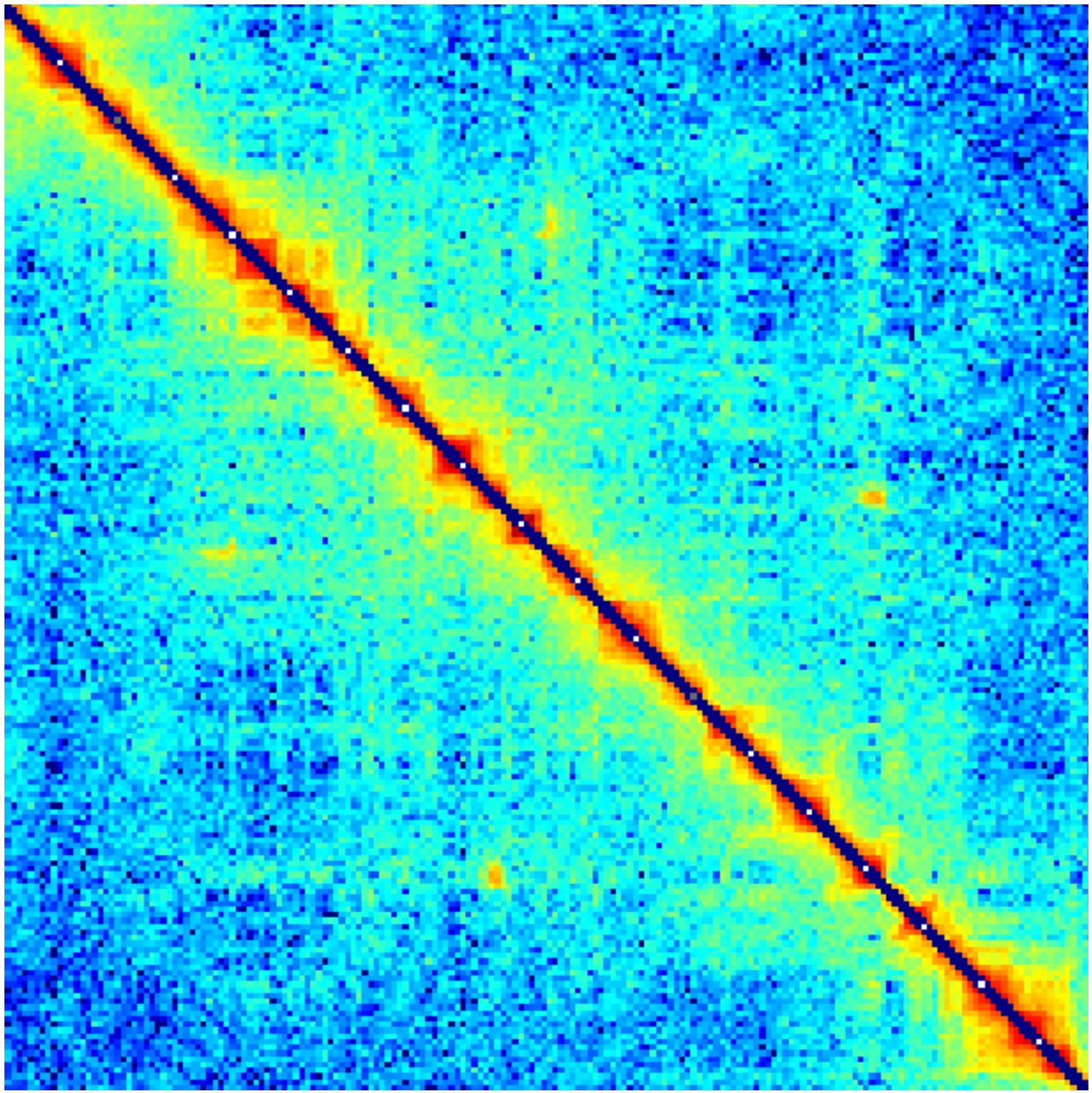
5. Bin level filtering
Remove 1% low signal rows/columns.

6. Balancing
Iterative correction.

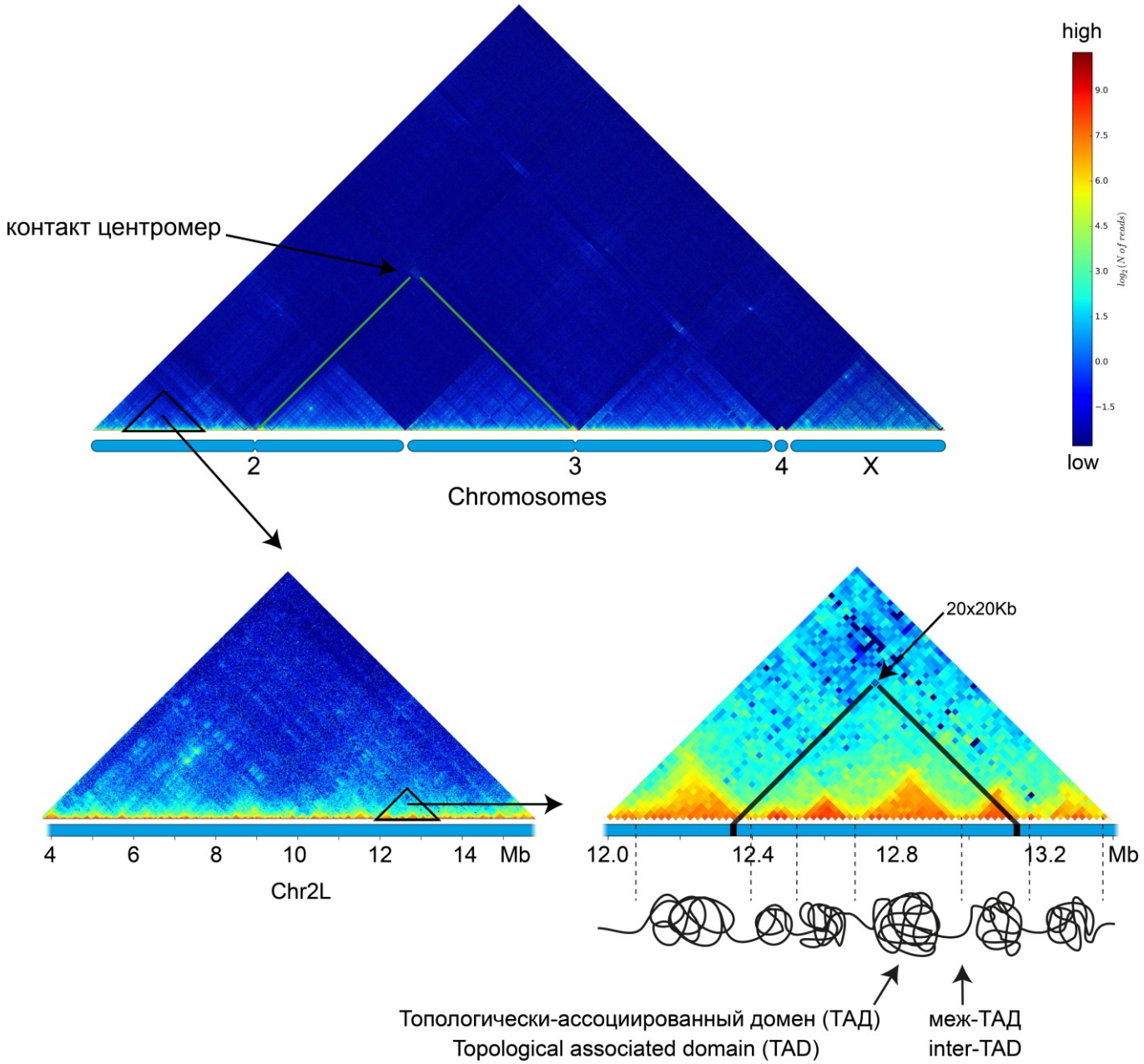


Lajoie, B. R., Dekker, J. & Kaplan, N.
The Hitchhiker's guide to Hi-C analysis: Practical guidelines.
***Methods* 72, 65–75 (2015).**

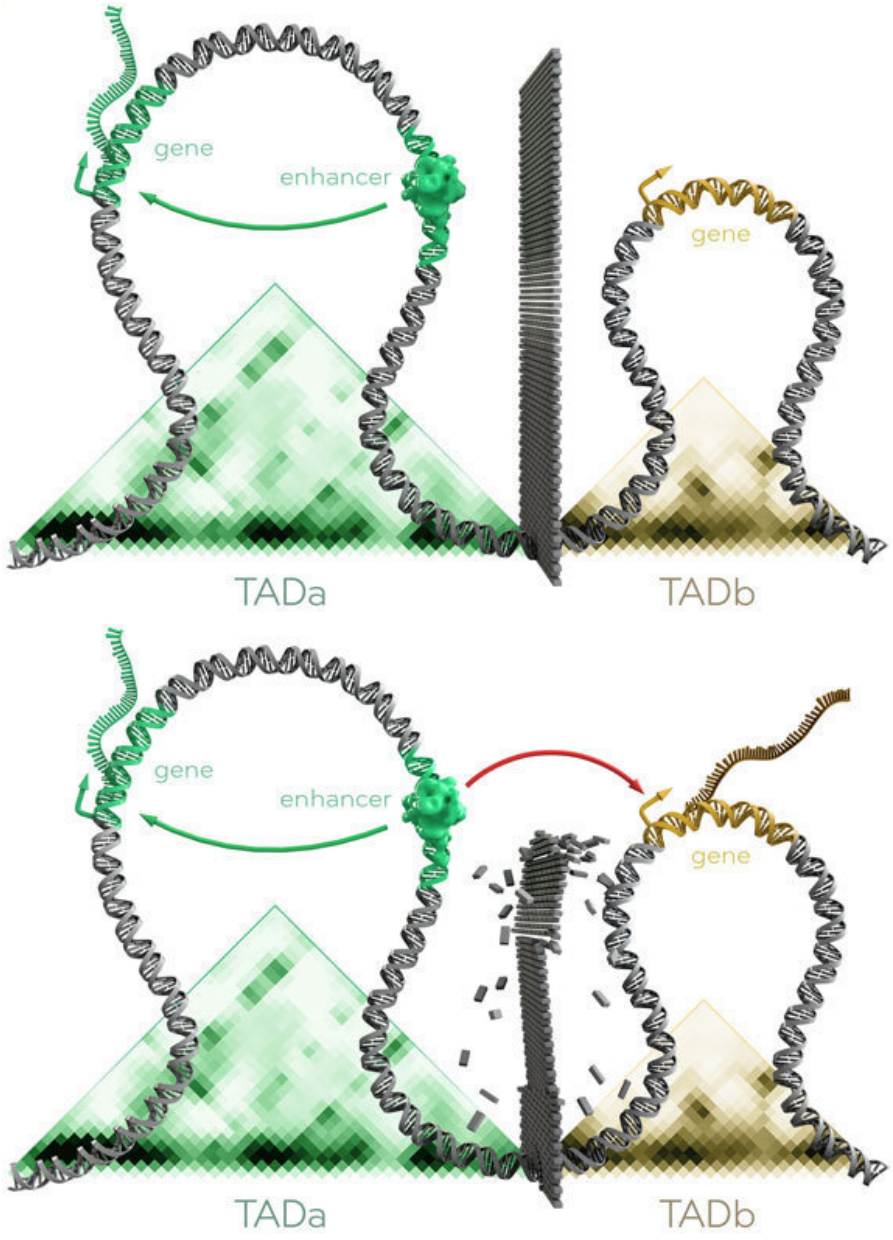
Iteratively corrected heatmap



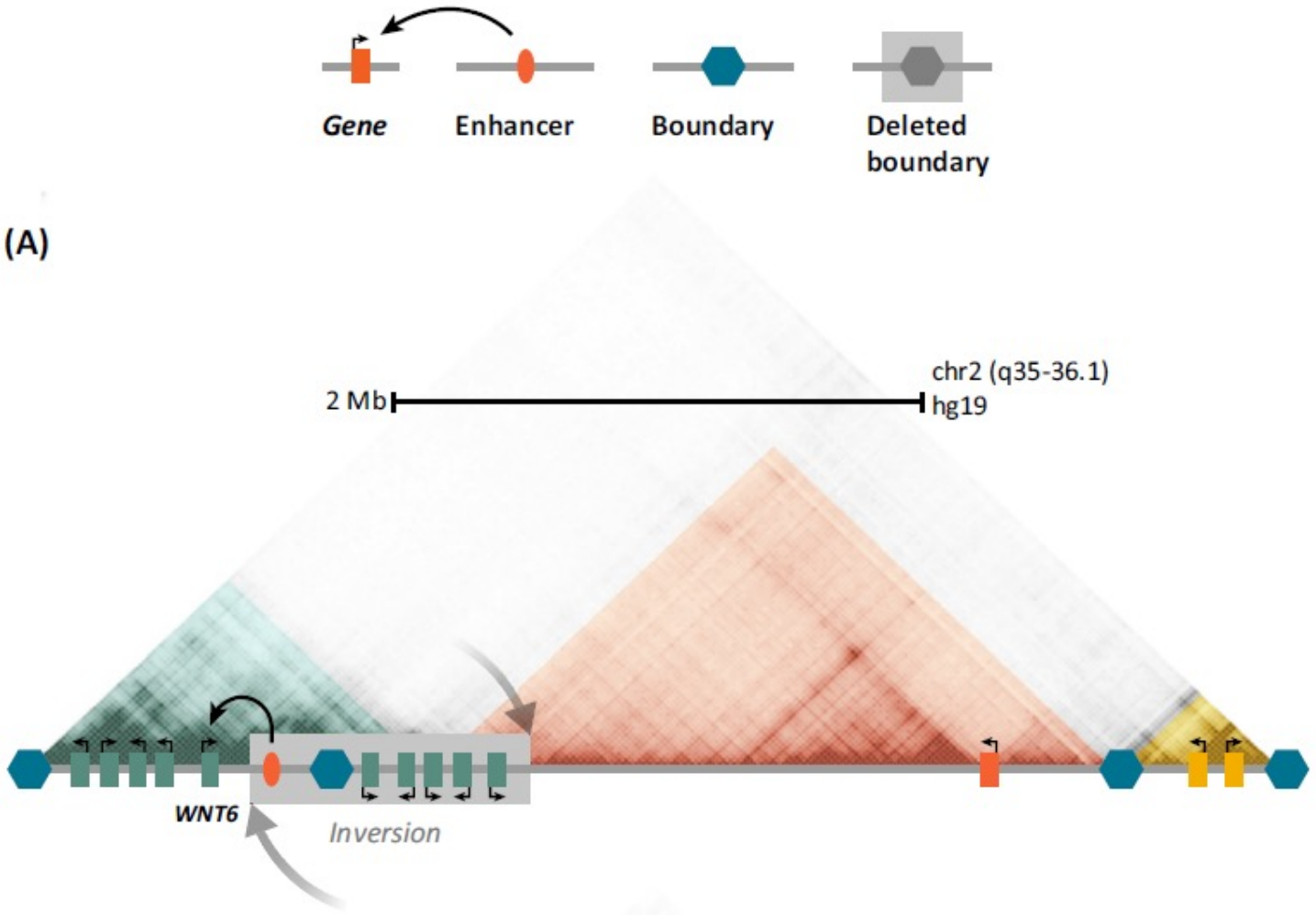
Топологические домены (ТАДы)

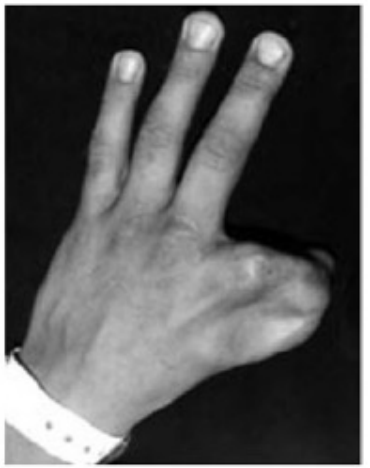


ТАДы и болезни



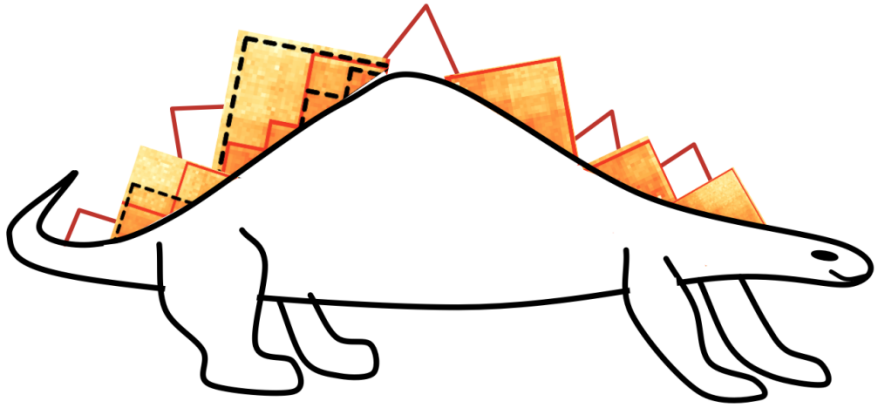
ТАДы и синдактилия





F-syndrome

Автоматизированное предсказание ТАДов

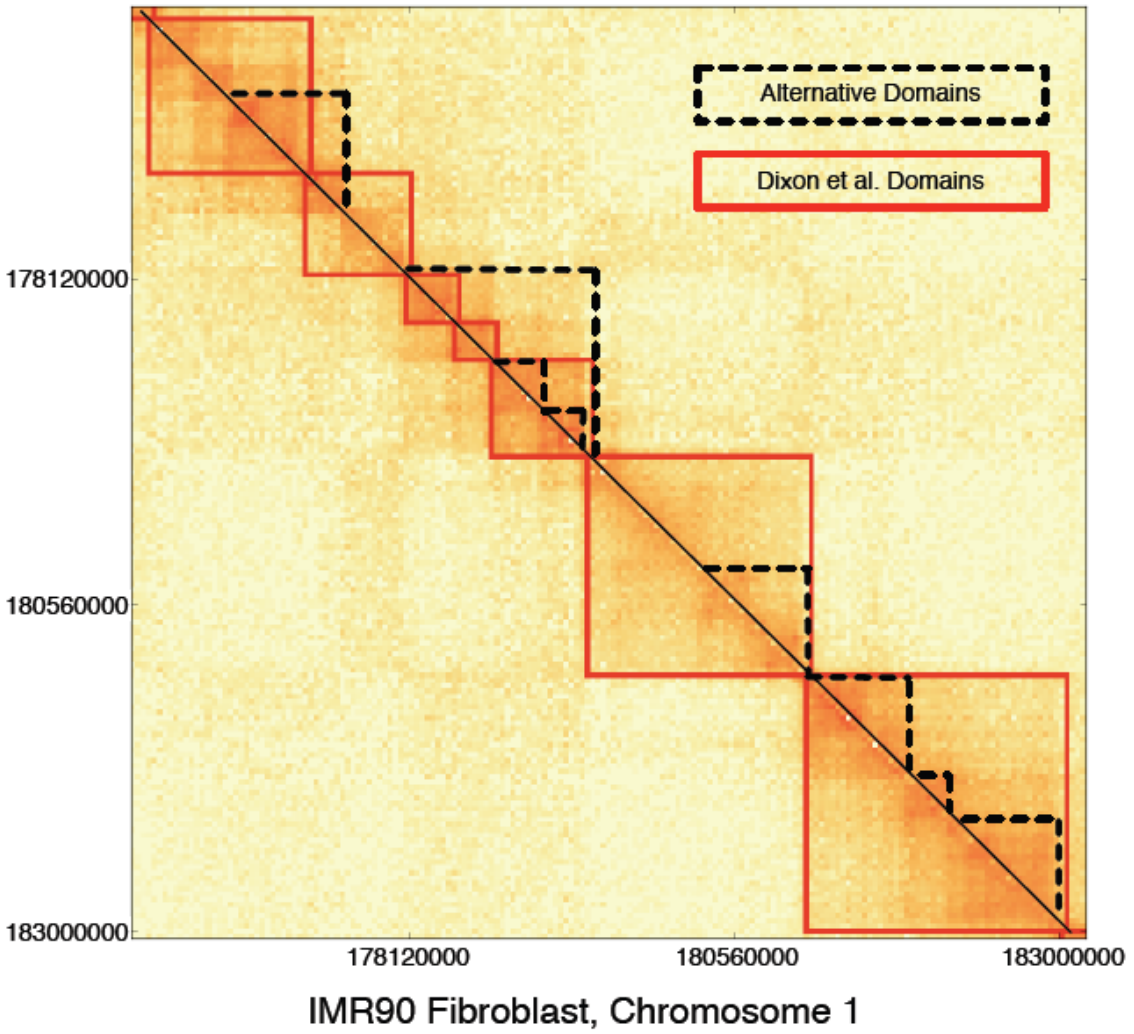


https://github.com/kingsfordgroup/armatus

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

Позволяет получить несколько разметок ТАДов с разной средней длиной ТАДа.

Иерархическая структура ТАДов: ТАДы вложены друг в друга.

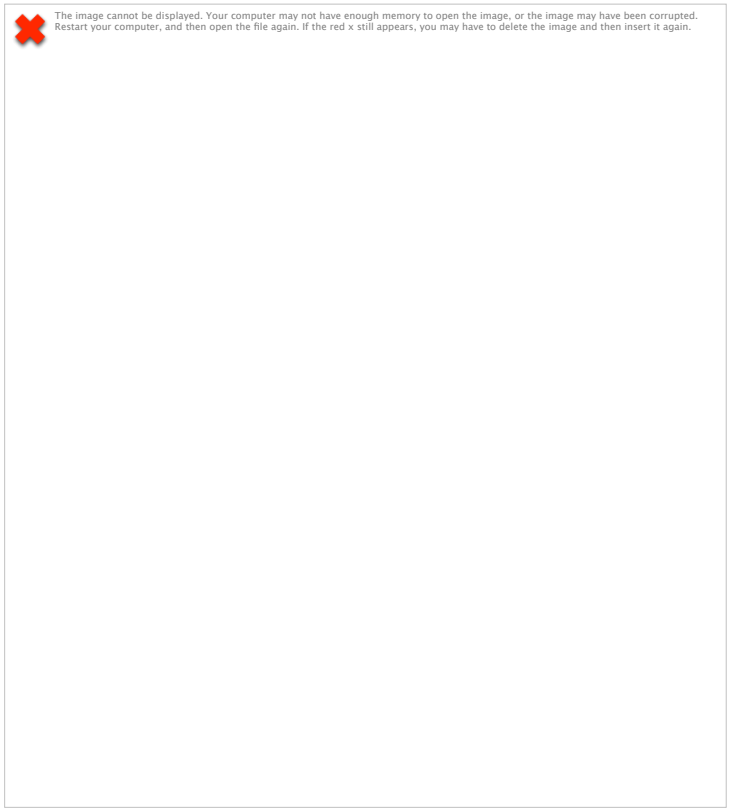


Какую науку делают с этими данными?

Мышь и человек

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

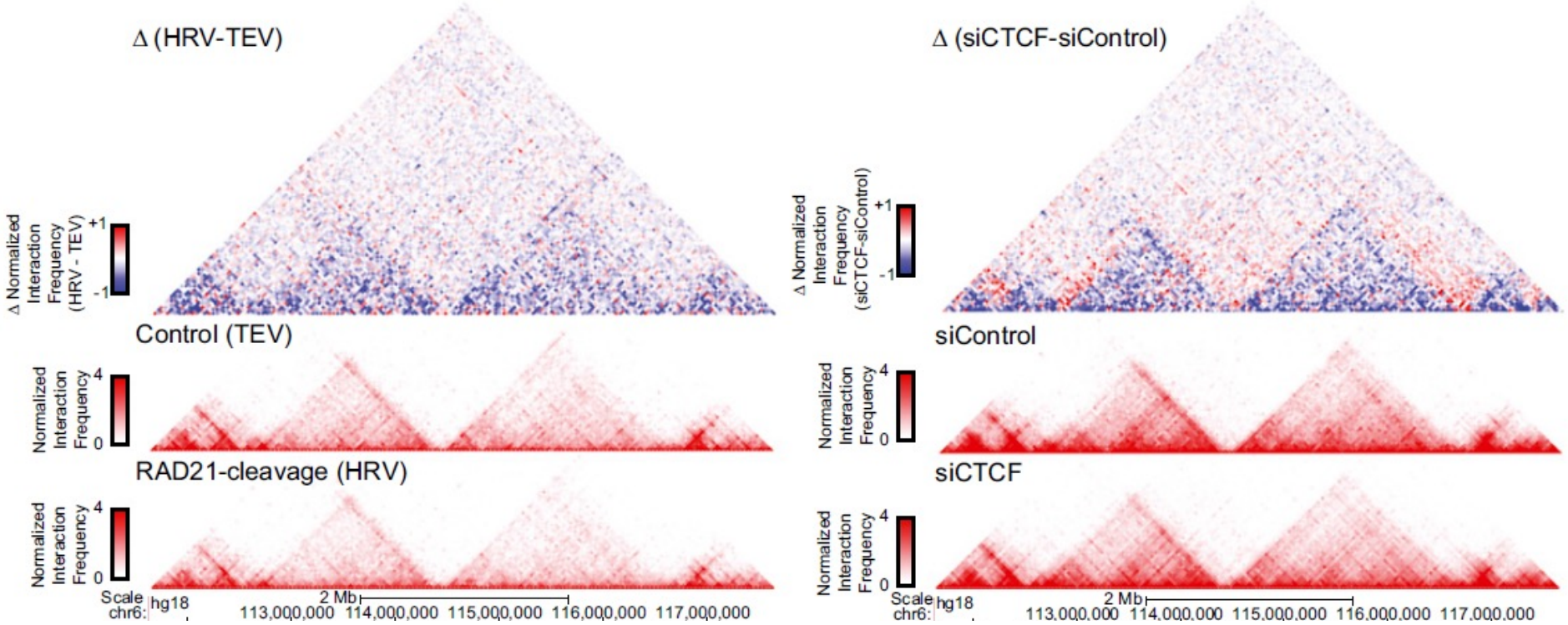
- Human and mouse Hi-C in embryonic and differentiated cell types.
- TAD. “Directionality Index” to quantify the degree of US or DS interaction bias.
- TADs are stable across different cell types and highly conserved across species.
- Boundaries are enriched for CTCF, housekeeping genes, tRNAs, and SINE.



Человек

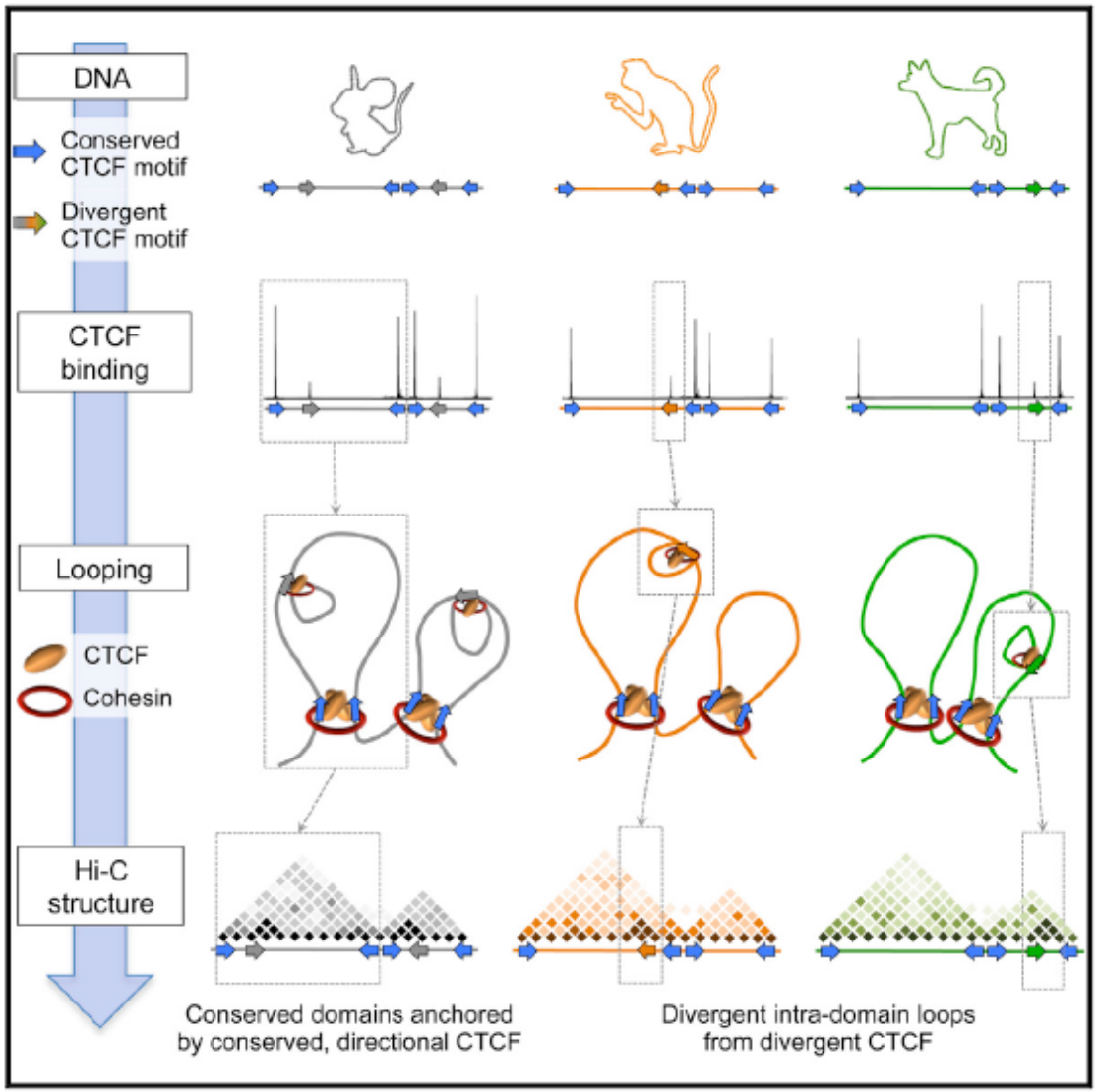
Zuin, J. et al. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc. Natl. Acad. Sci. U. S. A. 111, 996–1001 (2014).

- 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

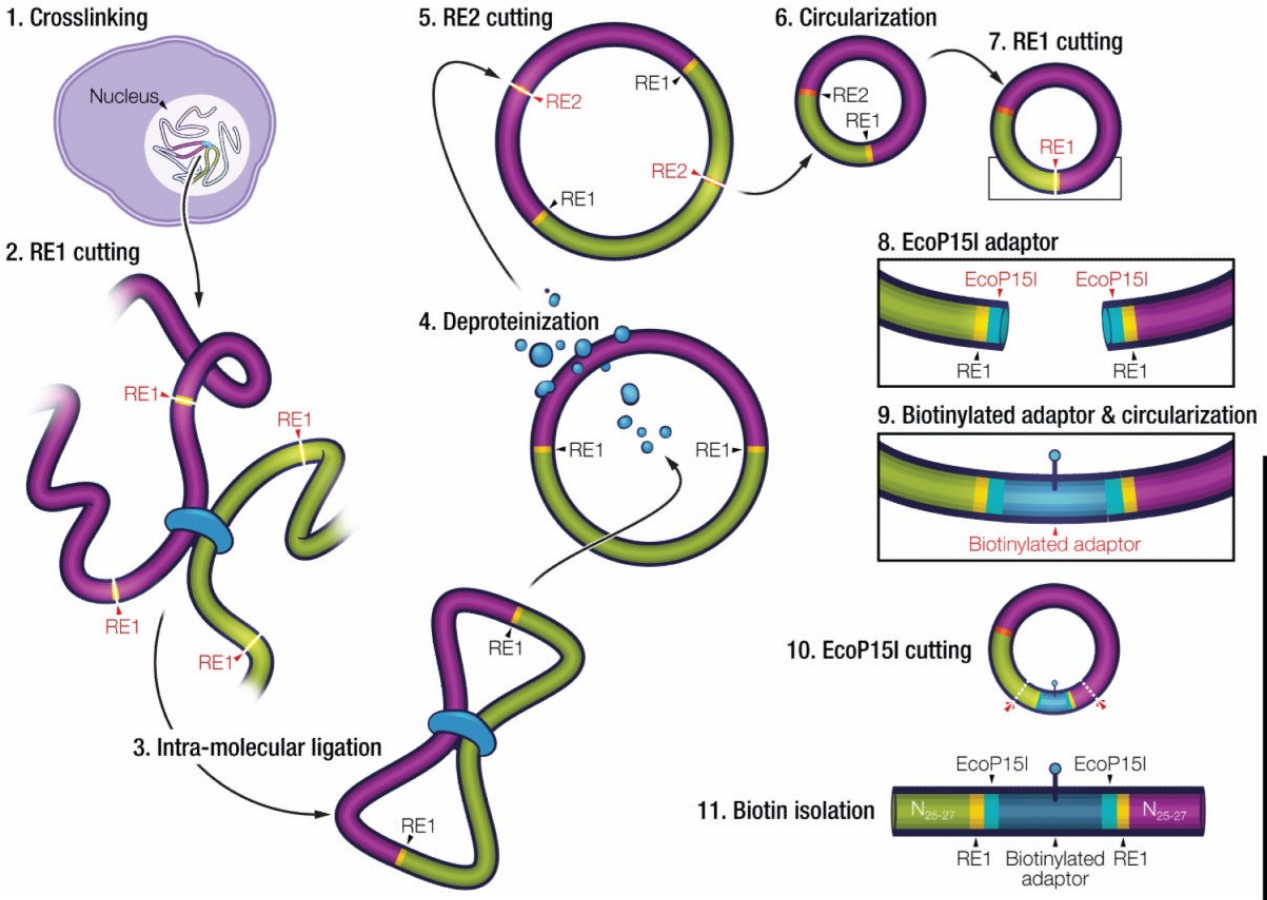
Vietri Rudan, M. et al. Comparative Hi-C Reveals that CTCF Underlies Evolution of Chromosomal Domain Architecture. Cell Rep. **10**, 1297–1309 (2015).



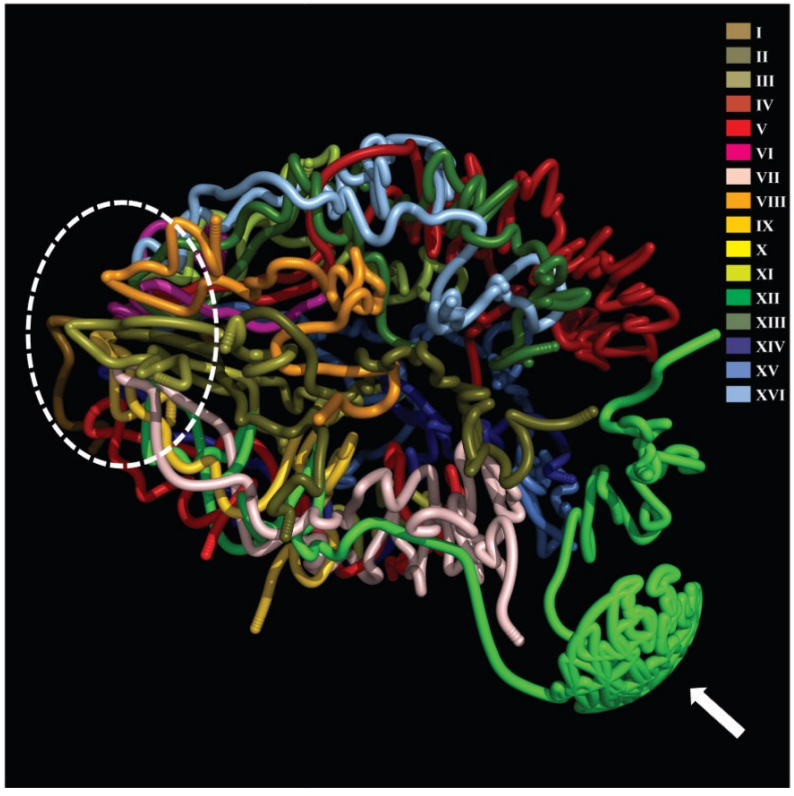
- Multi-species Hi-C comparisons reveal robust conservation of chromosome organization.
- Conserved CTCF sites are co-localized with cohesion and enriched at strong TAD borders
- Divergent CTCF binding between species is correlated with divergence of internal domain structure.
- Large-scale domains are 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).
- Rabl configuration.



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