



# *От организации хроматина к пониманию функционирования геномов эукариот*

Алексей Константинович Шайтан

д.ф.-м.н., профессор, чл.-корр. РАН

кафедра биоинженерии

биологический факультет

МГУ имени М.В.Ломоносова

**Лекция 6.**

**Экспериментальная эпигеномика.**

Апрель 2024

# Содержание лекции

- Методы секвенирования 1,2,3 поколений.
- Проект геном человека.
- Методы секвенирования в эпигенетике: бисульфитное, Nano-pore, Chip-Seq, Chip-Exo, Mnase-seq, ATAC-seq, CUT&TAG, CUT&RUN, Ribo-seq, Hi-C,
- От проекта геном человека к 4Д-нуклеому (включая ENCODE)
- GWAS/EWAS

# Секвенирование по Сенгеру (1-ое поколение)

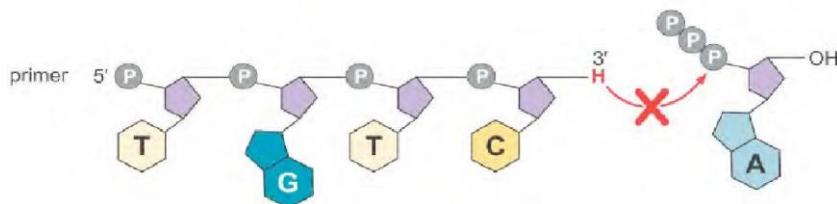
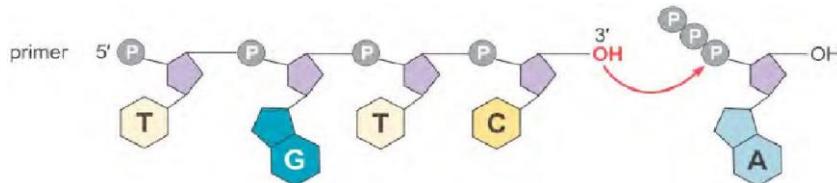
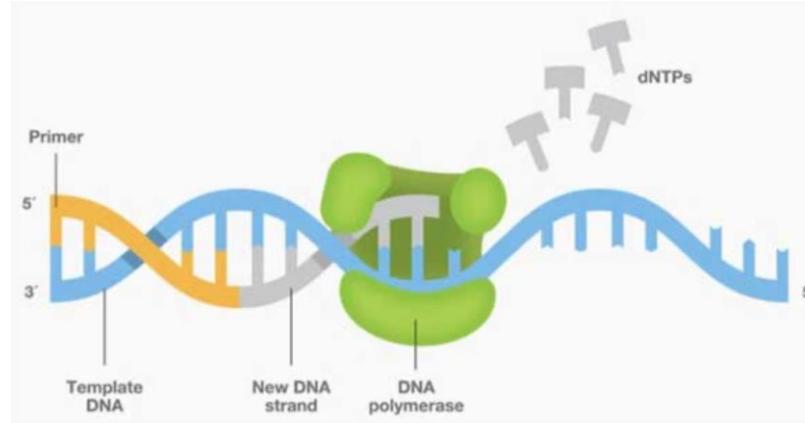
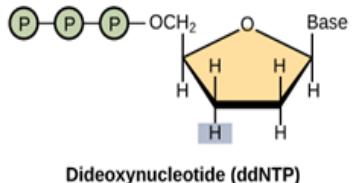
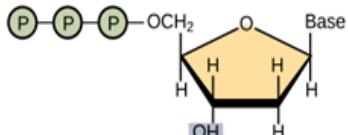
<b>Frederick Sanger</b> OM CH CBE FRS FAA				1958		Frederick Sanger (1918-2013)	 United Kingdom	"for his work on the structure of proteins, especially that of insulin"			
<b>Born</b> <a href="#">Rendcomb</a> , Gloucestershire, England	13 August 1918		1980		Paul Berg (b. 1926)	 United States	"for his fundamental studies of the biochemistry of nucleic acids, with particular regard to recombinant-DNA"				
	<b>Died</b> <a href="#">Cambridge</a> , England <sup>[1]</sup>			Walter Gilbert (b. 1932)	 United States	"for their contributions concerning the determination of base sequences in nucleic acids"					
				Frederick Sanger (1918-2013)	 United Kingdom						

1951-52 – последовательность инсулина

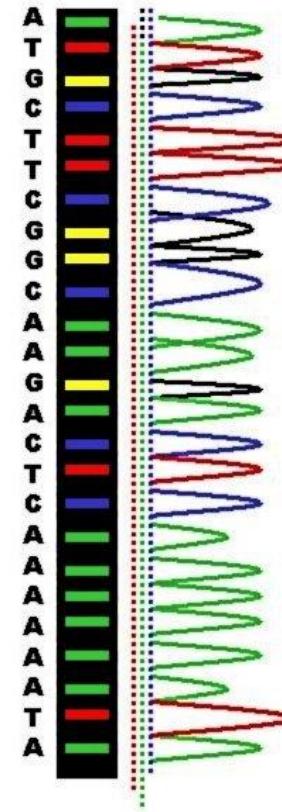
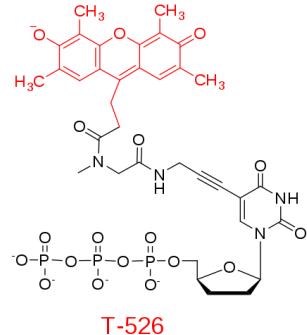
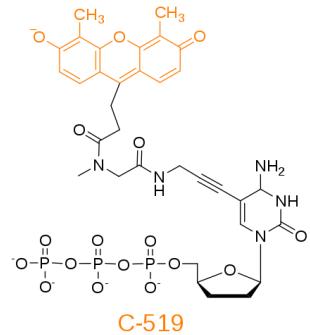
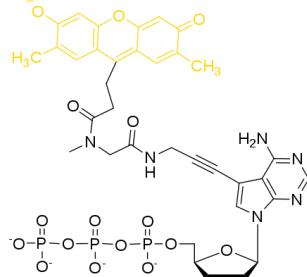
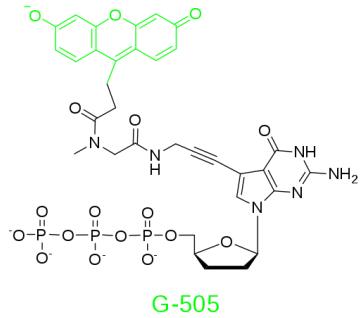
# Метод Сенгера

1977 год

Chain termination method



# dye-terminating sequencing

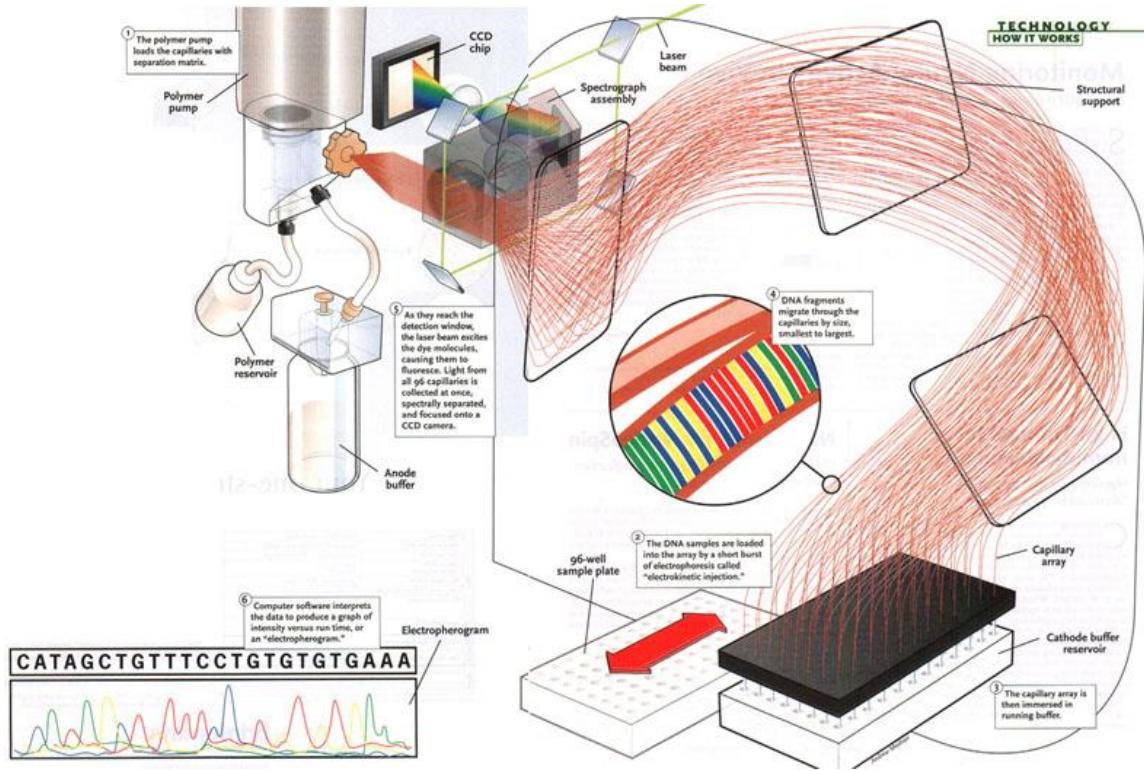


дДНТФ с флюоресцентными метками

хроматограмма

# Первое поколение

- Метод Сэнгера, dye-terminating sequencing, sequencing by synthesis, DNA sequencing by capillary electrophoresis
- Длина прочтений 700-1000 нк. Первые 15-40 нк – низкое качество.

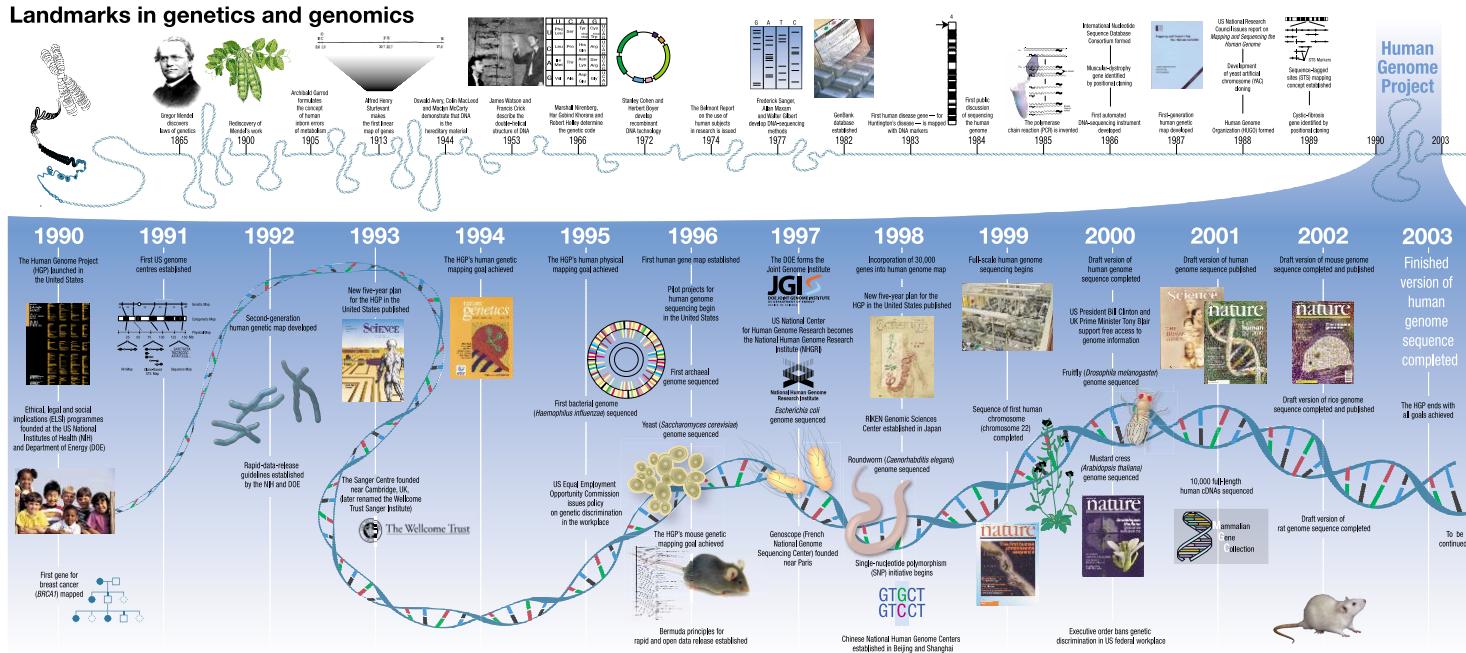


# Проект «Геном человека»

## Human Genome Project (HGP)

1990 - 2003

Рассчитан на 15 лет, с бюджетом \$3 млрд, завершен досрочно



# Проект геном человека: методы и подходы



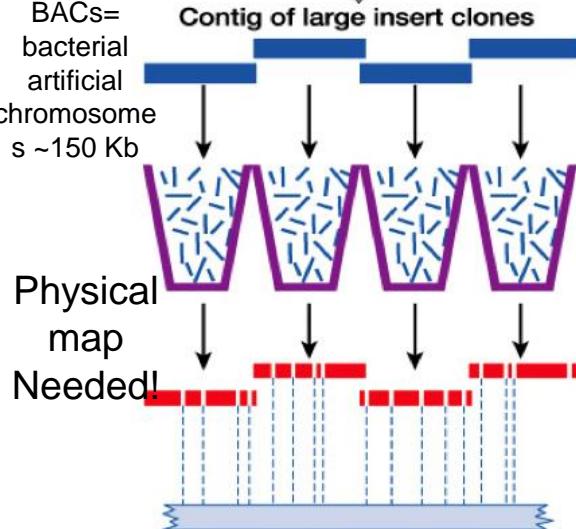
F. Collins

1990

'BAC-by-BAC' approach

Hierarchical shotgun

BACs =  
bacterial  
artificial  
chromosome  
s ~150 Kb



Genome

Random fragmentation

Sequencing and assembly

Genome assembly

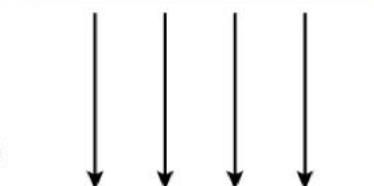


C. Venter

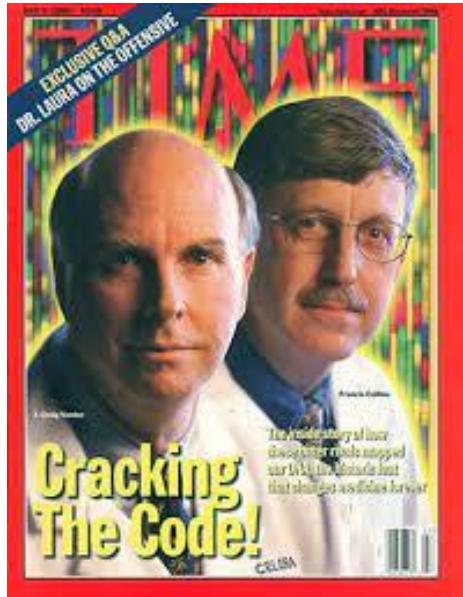
1998



Whole-genome shotgun



2000



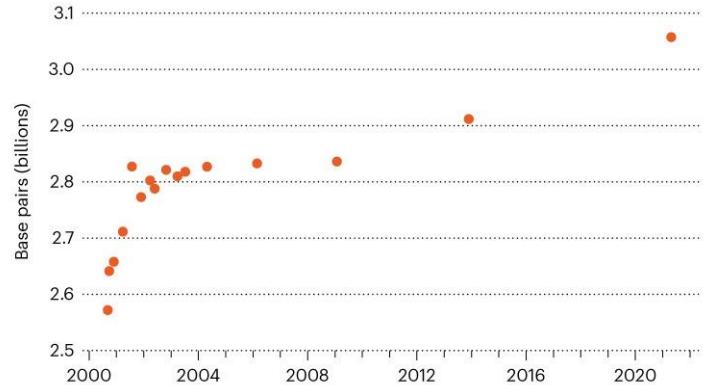
31 марта 2022



## A Truly Complete Human Genome Sequence

### COMPLETING THE HUMAN GENOME

Researchers have been filling in incompletely sequenced parts of the human reference genome for 20 years, and have now almost finished it, with 3.05 billion DNA base pairs.

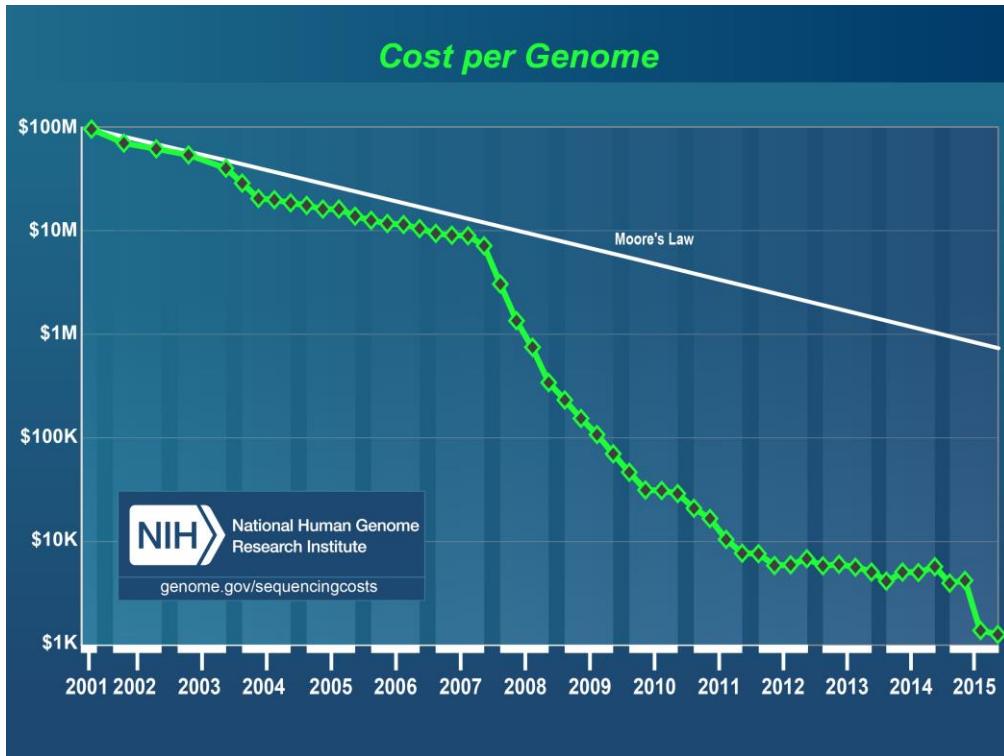


0.3% of sequence might still have errors. Includes X but not Y chromosome. Count excludes mitochondrial DNA.

©nature

Human genome size ~ 3,117 billion base pairs (bp)

# NGS sequencing



nature

Vol 452 | 17 April 2008 doi:10.1038/nature06884

## LETTERS

### The complete genome of an individual by massively parallel DNA sequencing

David A. Wheeler<sup>1\*</sup>, Maithreyan Srinivasan<sup>2\*</sup>, Michael Egholm<sup>2\*</sup>, Yufeng Shen<sup>1</sup>, Lei Chen<sup>1</sup>, Amy McGuire<sup>3</sup>, Wen He<sup>3</sup>, Yi-Ju Chen<sup>2</sup>, Vinod Makhijani<sup>2</sup>, G. Thomas Roth<sup>4</sup>, Xavier Gomes<sup>3</sup>, Karrie Tartaro<sup>1,†</sup>, Faheem Niazi<sup>2</sup>, Cynthia L. Turcotte<sup>2</sup>, Gerard P. Izyk<sup>2</sup>, James R. Lupski<sup>4,5,6</sup>, Craig Chinault<sup>1</sup>, Xing-zhi Song<sup>1</sup>, Yue Liu<sup>1</sup>, Ye Yuan<sup>1</sup>, Lynne Nazareth<sup>1</sup>, Xiang Qin<sup>1</sup>, Donna M. Muzny<sup>1</sup>, Marcel Margulies<sup>2</sup>, George M. Weinstock<sup>1,4</sup>, Richard A. Gibbs<sup>1,4</sup> & Jonathan M. Rothberg<sup>2,†</sup>

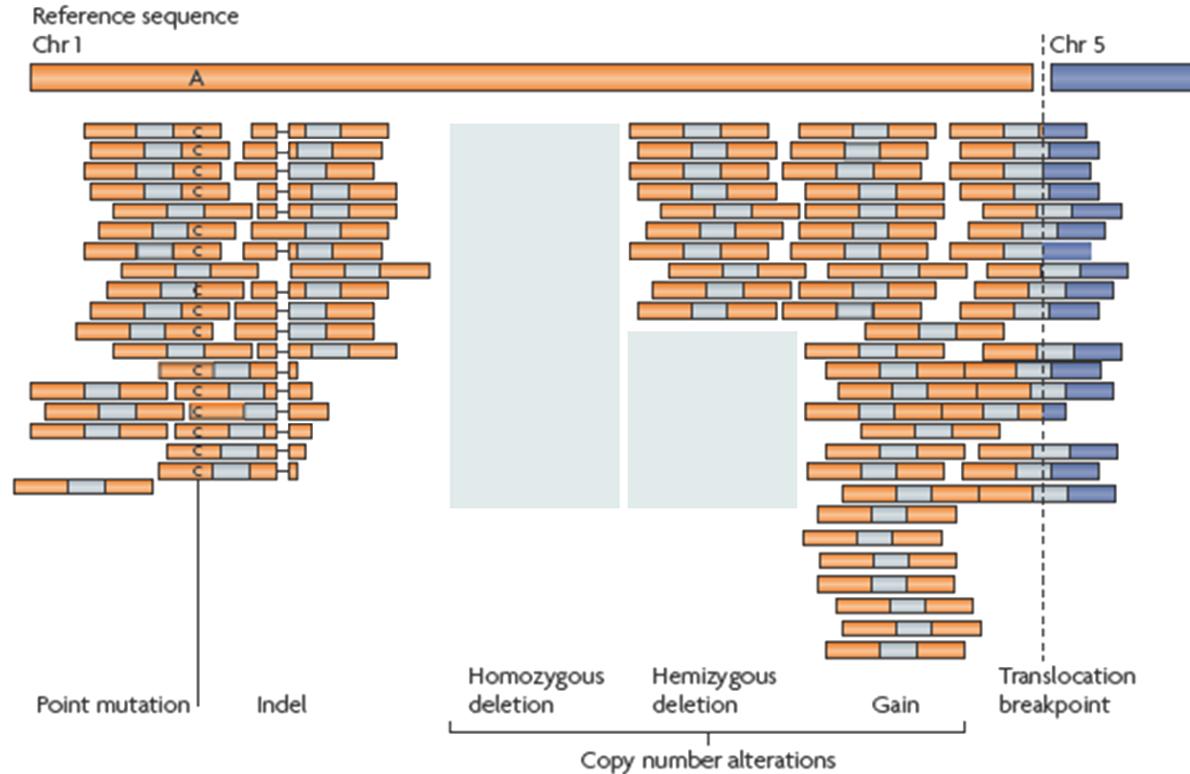
The association of genetic variation with disease and drug response, and improvements in nucleic acid technologies, have given great optimism for the impact of 'genomic medicine'. However,

### James Watson's genome sequenced at high speed



2008  
454 Life Sciences  
4 months  
\$1.5 mln

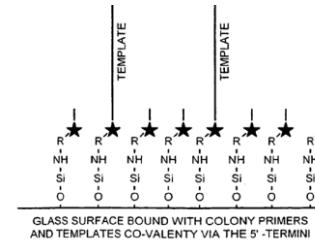
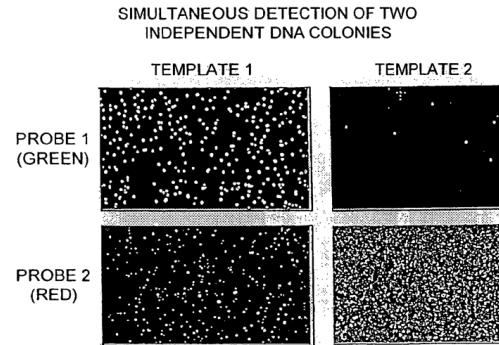
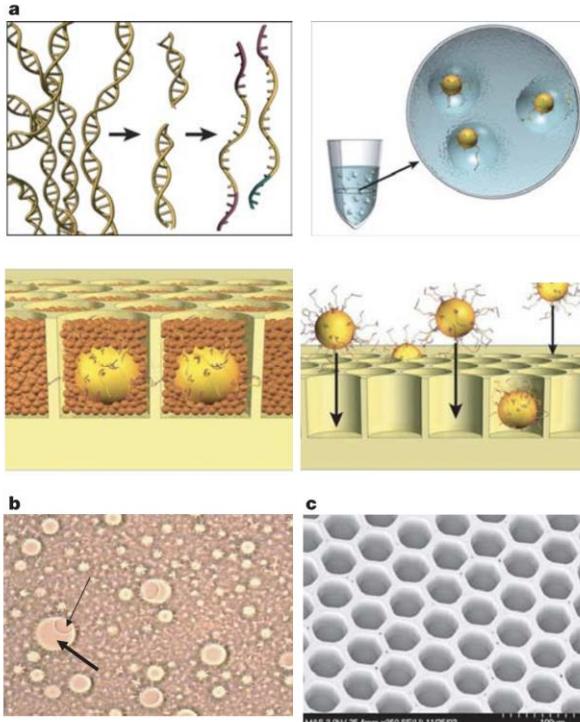
# Ресеквенирование



Nature Reviews Genet. 2010 Oct ;11(10):685-96.

Для достоверности в медицинских целях 30-100-кратное покрытие,  
30-100-кратная глубина секвенирования

# Технологии микрореакторов, массивно-параллельной работы с клональными библиотеками ДНК, «колониями» ДНК



(12) United States Patent  
Adessi et al.

(10) Patent No.: US 7,115,400 B1  
(45) Date of Patent: \*Oct. 3, 2006

(54) METHODS OF NUCLEIC ACID AMPLIFICATION AND SEQUENCING

(75) Inventors: Celine Adessi, Ambilly (FR); Eric Kawashima, Nyon (CH); Pascal Mayer, Eloise (FR); Jean-Jacques Mermoud, Geneva (CH); Gerardo Turcatti, Geneva (CH)

(73) Assignee: Solexa Ltd., Essex (GB)

(56) References Cited

U.S. PATENT DOCUMENTS

5,514,539 A \* 5/1996 Buch et al. .... 435/5  
(Continued)

FOREIGN PATENT DOCUMENTS

EP 0543484 A2 5/1993

(Continued)

Технологии секвенирования нового (второго) поколения

NGS – next generation sequencing

Массивно-параллельное секвенирование

Massively parallel sequencing

### Sequencing by synthesis

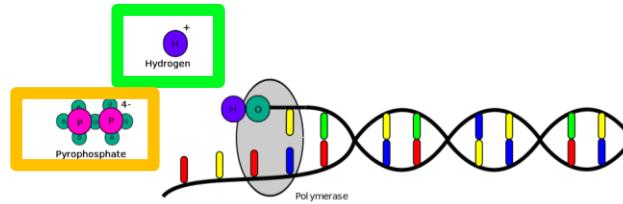
- Пиросеквенирование – по технологии компании 454 Life Sciences/Roche
  - Первая технология на рынке. С 2016 года не поддерживается.
- Секвенирование на основе обратимой терминации синтеза ДНК – по технологии компаний Solexa/Illumina.
  - 90% всего секвенирования выполняется сейчас на этой технологии.
- Полупроводниковое секвенирование - Ion Torrent
- DNBSEQ метод от компании MGI на основе ДНК-наношаров (CoolMPS версия)

### Sequencing by ligation

- SOLiD от Applied Biosystems. Не поддерживается
- DNBSEQ метод от компании MGI на основе ДНК-наношаров (cPAS версия)

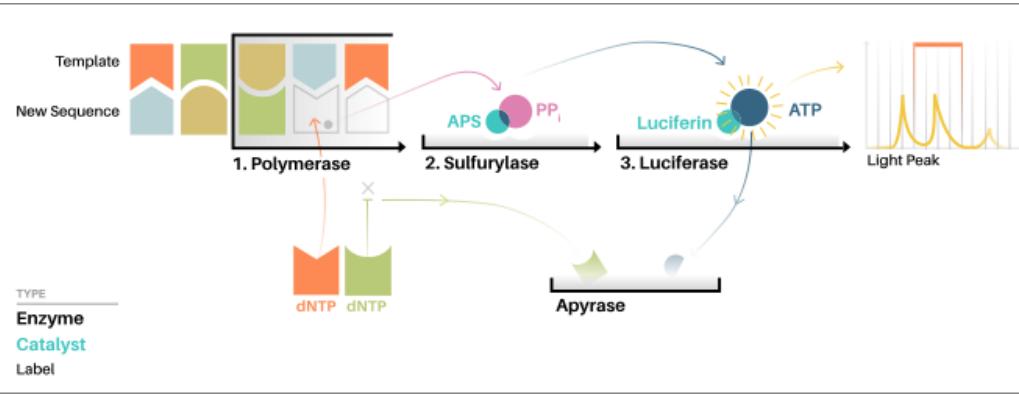
### Sequencing by binding (PacBio announced)

# Пиросеквенирование/ Ионное полупроводниковое секвенирование

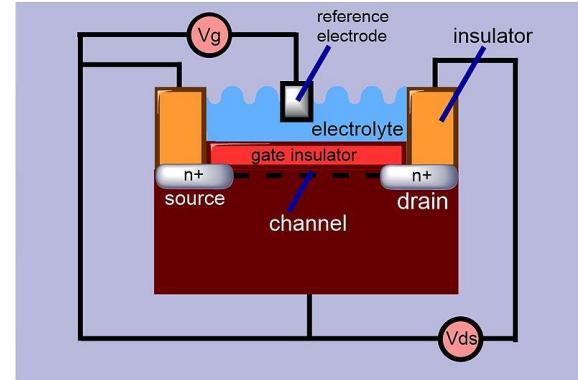


Hydrogen and pyrophosphate are released.

Пирофосфат детектируется через активность люциферазы



**ion-sensitive field-effect transistor (ISFET)**  
Ион чувствительный полевой транзистор  
рН-чувствительный транзистор



# Пиросеквенирование от 454 Life Sciences

- 2005 год – первый секвенатор (GS20 – 25 млн. нт за прогон)
- 2005 секвенированы геномы Mycoplasma и Streptococcus
- 2007 год – улучшенный секвенатор 454 GS-FLX
- 2008 – секвенирован геном James Watson
- Компания 454 LS куплена Roche в 2007 году.
- 2012 – FLX+
- 2006 – 2013 - ... – секвенирование генома неандертальца
- С 2016 года технология закрыта, неконкурентноспособна (по сравнению с Illumina, IonTorrent)



Jonathan Rothberg

Rothberg in 2008  
Born April 28, 1963 (age 59)  
New Haven, Connecticut

- FLX+: 1 млн прочтений, до 700 нт, 0.7 Гб, \$8К, 10 часов за прогон
- Хорошая точность
- Ошибки в гомополимерных повторах





#### **Применение:**

- секвенирование de novo коротких фрагментов ДНК (180 п.о.);
- ресеквенирование любого участка генома человека, модельных объектов, микроорганизмов или вирусов;
- количественный анализ частот аллельной встречаемости;
- мутационный анализ: идентификация точечных однонуклеотидных замен (SNP), вставок и делеций;
- количественный анализ метилирования ДНК (может быть совмещен с SNP типированием);
- верификация и валидация результатов полногеномного анализа.

#### **Технические характеристики:**

- производительность — от 1 до 48 образцов;
- пробоподготовка — автоматизированная на самом приборе;
- сенсорный дисплей управления;
- размеры, Ш × Г × В, см — 25 × 30 × 30;
- вес, кг — 8,5.

# Полупроводниковое секвенирование от Ion Torrent

- 2007 год – Ion Torrent Systems – основана компания
- 2010/11 - первый секвенатор Personal Genome Machine (PGM)
- 2010 - Компания приобретена Life Technologies, позднее ThermoFisher
- 2012 - секвенатор Ion Proton (в 1000 раз мощнее PGM), геном человека за \$1000 (не за один прогон)



- Длина ридов до 400 нт, высокая скорость, нет сложной химии, дешевизна прибора,
- Позиционируется для задач диагностики – каждому доктору в кабинет
- Ошибки в гомополимерных повторах, нет возможности парноконцевого секвенирования



## Ion GeneStudio S5 Systems



**Ion Torrent Genexus System**

- Быстрое время от эксперимента до результата
- Высокий уровень автоматизации
- Хорош для медицинских исследований
- Чипы дают ограниченное число прочтений

## SOLiD (Sequencing by Oligonucleotide Ligation and Detection) от ABI

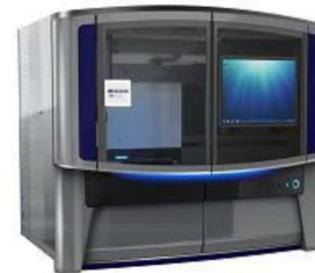
- 2003 – polony sequencing technology (G. Church)
- 2006/08 год – первый секвенатор по технологии SOLiD



ThermoFisher  
SCIENTIFIC

- 2015 ... - сняты с производства
- 2017 - закончена техподдержка

- Короткие риды (50-75 нт), очень дешево (\$0.13/млн нт)
- Высокая точность 99.94%. Использовался для ресеквенирования, анализа вариаций
- Медленно – 1 запуск – до 14 дней
- Нет проблем с гомополимерными повторами
- Проблемы с палиндромными сиквенсами



Applied Biosystems™

5500xl Genetic Analyzer

# NGS платформа Illumina/Solexa

- 1998: Shankar Balasubramanian, David Klenerman (Cambridge) founded Solexa company
- 2005 секвенирован геном phiX-174
- 2006 – первый секвенатор (Genome Analyzer – 1 млрд. нт. за прогон)
- 2007 год – acquired by Illumina
- 2007 - ... - эволюция приборов и технологии. Технология на четырех флуоресцентных метках, на двух, на одной + полупроводниковая детекция

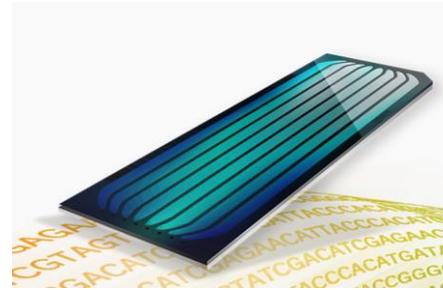


- Линейка секвенаторов
- Длина прочтений: до 150-300 нт.
- Возможность прочтения прямой и обратной цепей (парноконцевое секвенирование, paired-end sequencing)

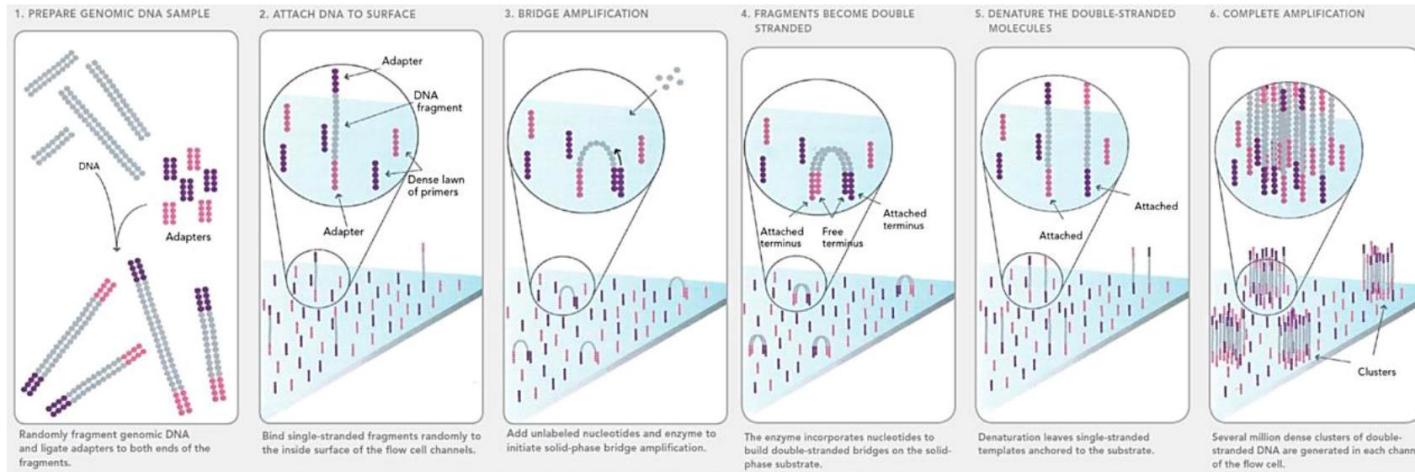
Genome Analyzer

# NGS платформа Illumina/Solexa

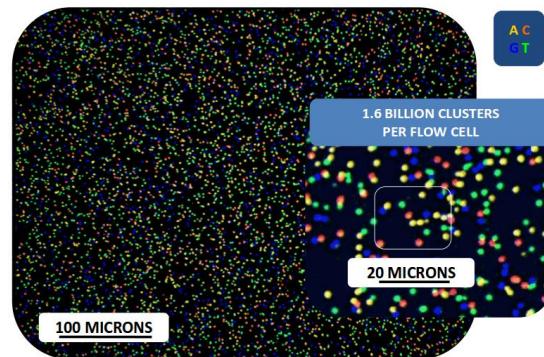
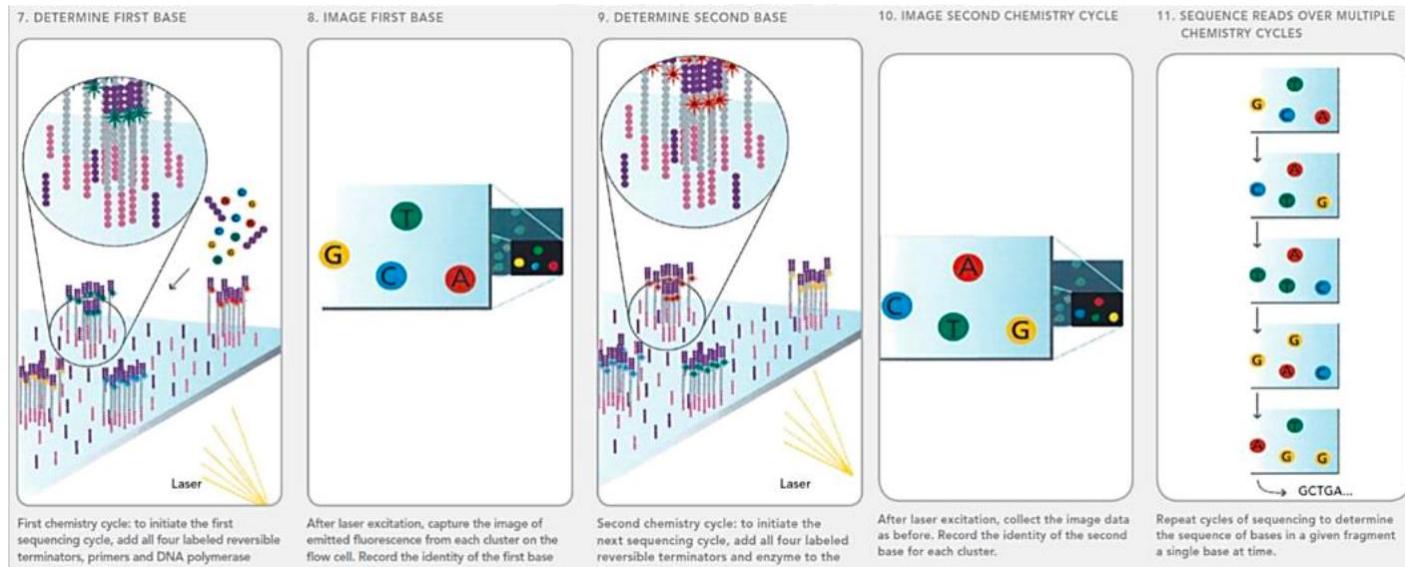
Sequencing by synthesis  
Bridge amplification  
Dye terminator sequencing



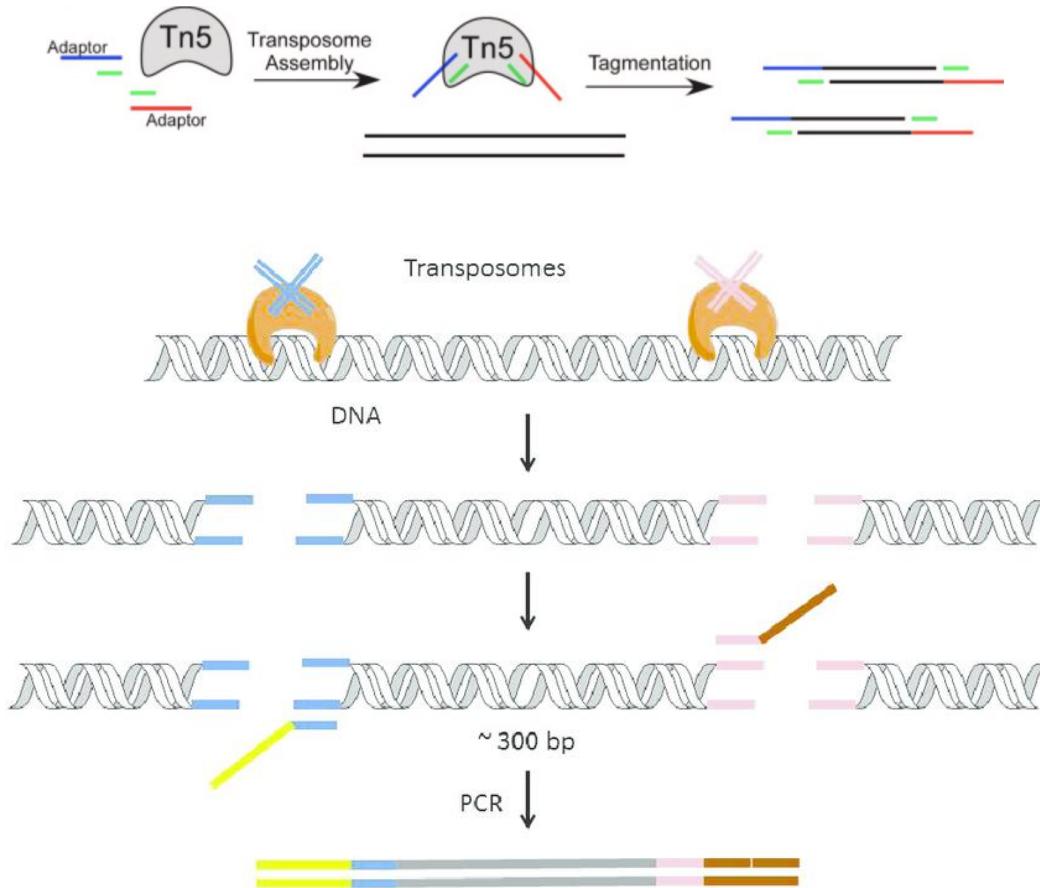
Flow cell  
Проточная ячейка



# NGS платформа Illumina/Solexa



# Подготова библиотеки (Illumina library prep)



# Технология DNBSEQ от MGI

- 1999 – основан Beijin Genomic Institute => BGI Group
- 2009 – Complete Genomics отсеквенировала геном человека с помощью своей проприетарной технологии (статья в Science)
- 2013 – BGI приобретена компания Complete Genomics
- 2014 – BGISEQ-1000 первый секвенатор
- 2014 - ... дальнейшее развитие технологии



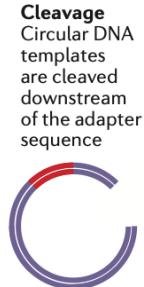
- Позиционируется как альтернатива Illumina, которая дешевле и быстрее
- Длина ридов до 150 нт
- Парноконцевое секвенирование
- performances of the MGISEQ-2000 and DNBSEQ-T7 platforms are comparable to that of the Illumina NovaSeq 6000

# Library generation through RCA

## d In-solution DNA nanoball generation (Complete Genomics (BGI))



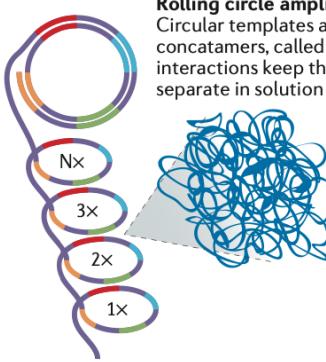
**Adapter ligation**  
One set of adapters is ligated to either end of a DNA template, followed by template circularization



**Cleavage**  
Circular DNA templates are cleaved downstream of the adapter sequence

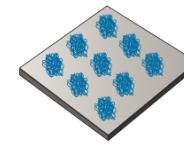


**Iterative ligation**  
Three additional rounds of ligation, circularization and cleavage generate a circular template with four different adapters

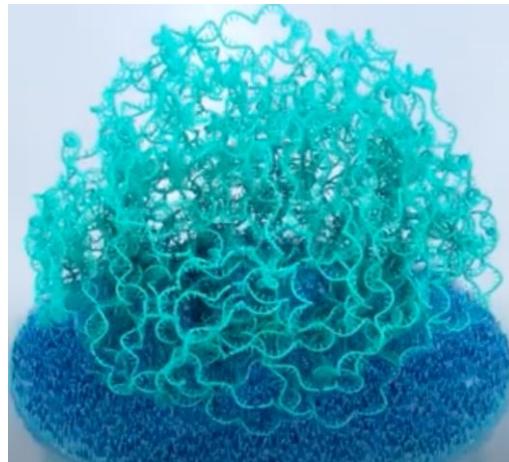


### Rolling circle amplification

Circular templates are amplified to generate long concatamers, called DNA nanoballs; intermolecular interactions keep the nanoballs cohesive and separate in solution



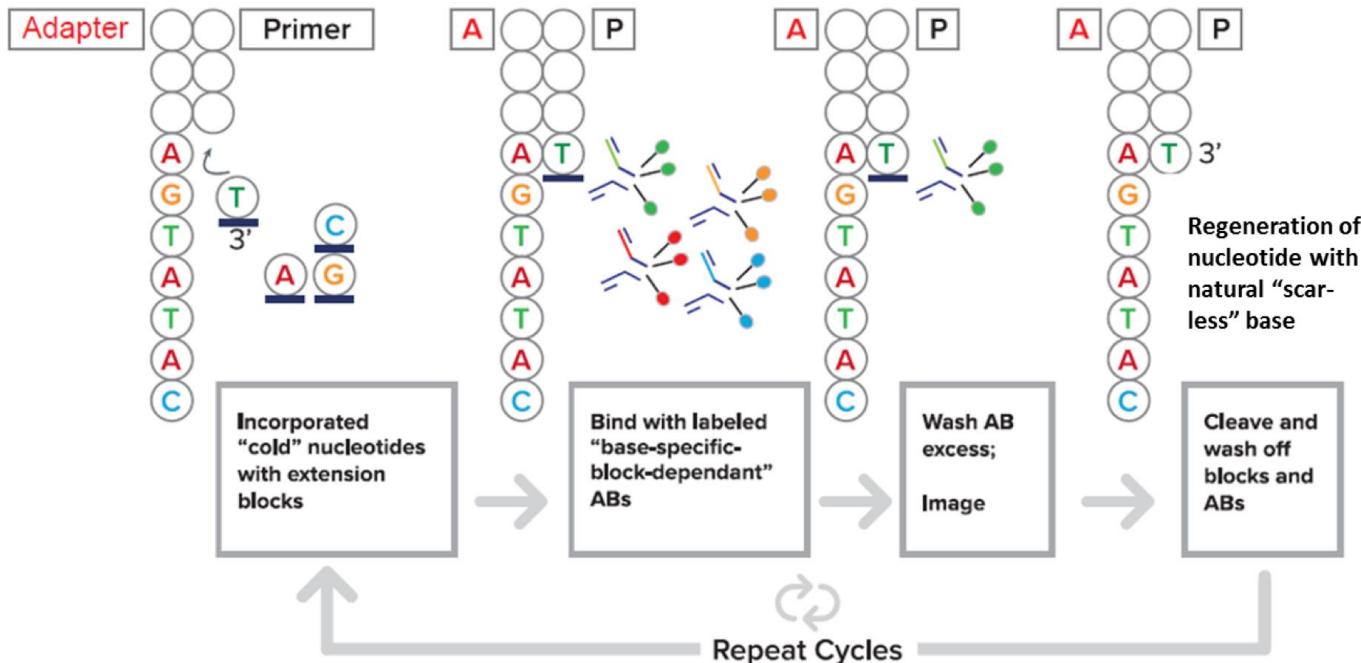
**Hybridization**  
DNA nanoballs are immobilized on a patterned flow cell



DNA nanoball

# CoolMPS

## sequencing by synthesis



**Figure 1:** CoolMPS™ process overview. Bars (—) on the unlabeled ("cold") nucleotides depict removable 3' chemical blocks. Antibodies specific for RTs with natural nucleobase are depicted with three dye molecules to increase fluorescent signal.



# Sequencing by binding

2021: Компания PacBio приобрела Omniome  
Сверхвысокая точность (Q40?)  
2022 октябрь – коммерческий продукт



- [https://youtu.be/i\\_mSaNBOVmQ](https://youtu.be/i_mSaNBOVmQ)

# Методы секвенирования третьего поколения

- Плюсы:

- 1) Можно секвенировать длинные фрагменты ДНК
- 2) Можно получать длинные прочтений (до 20 000 нт)
- 3) Не обязательно амплифицировать для считывания сигнала => более простая пробоподготовка  
=> портативные секвенаторы
- 4) Эпигенетическая разметка ДНК сохраняется (NanoPore)

Минусы:

- 1) Слабый сигнал, высокий процент ошибок

Технологии:

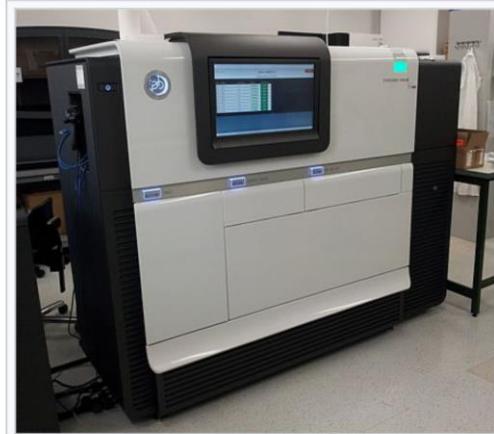
- 1) SMRT by Pacific Biosciences
- 2) Oxford Nanopore
- 3) Illumina Infinity 2022?

# Технология PacBio



- 2004 – основана Pacific Biosciences
- 2010 – первый прибор PacBio RS
- 2013 - PacBio RS II
- 2015 - Sequel System
- 2019 - Sequel II System
- 2020 - Sequel Ile System
- 2022 – Revio System

- Определение структурных вариаций генома
- Может секвенировать эпигенетические модификации
- Может секвенировать РНК напрямую.

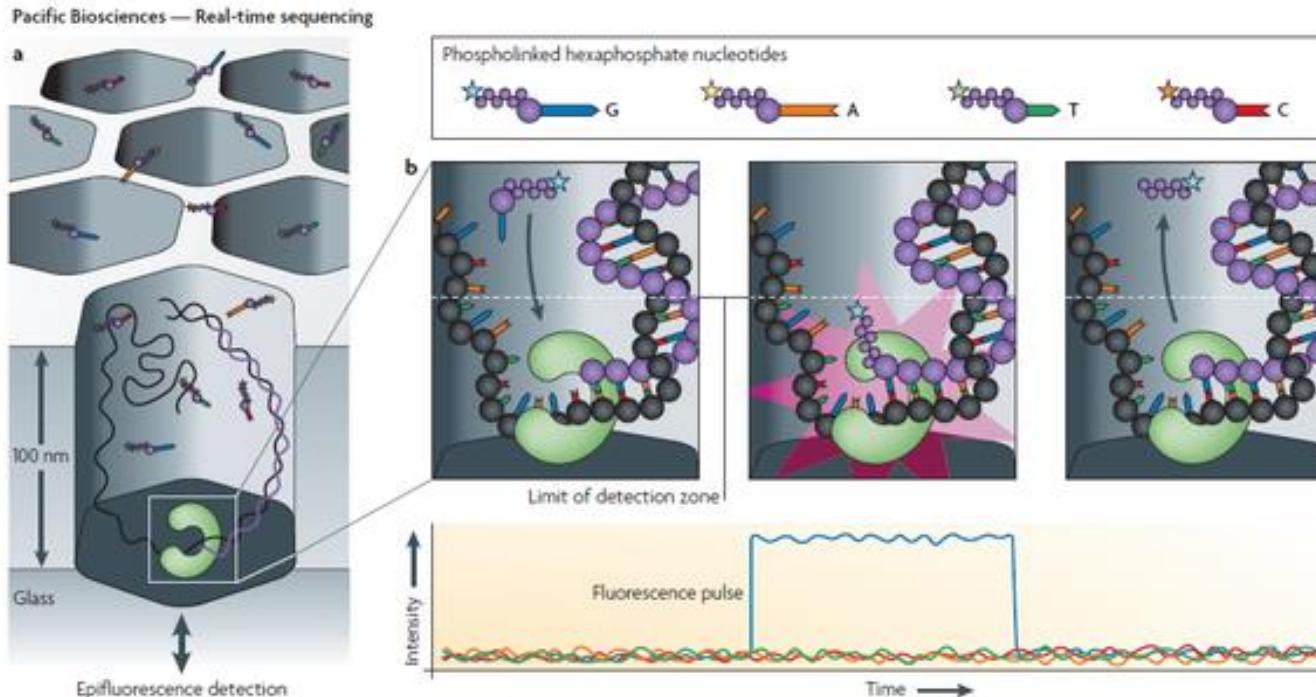


A PacBio RSII sequencer



# Технология PacBio

## Single-molecule real-time (SMRT) sequencing



[https://youtu.be/\\_ID8JyAbwEo](https://youtu.be/_ID8JyAbwEo)

<https://youtu.be/NHCJ8PtYCFc>

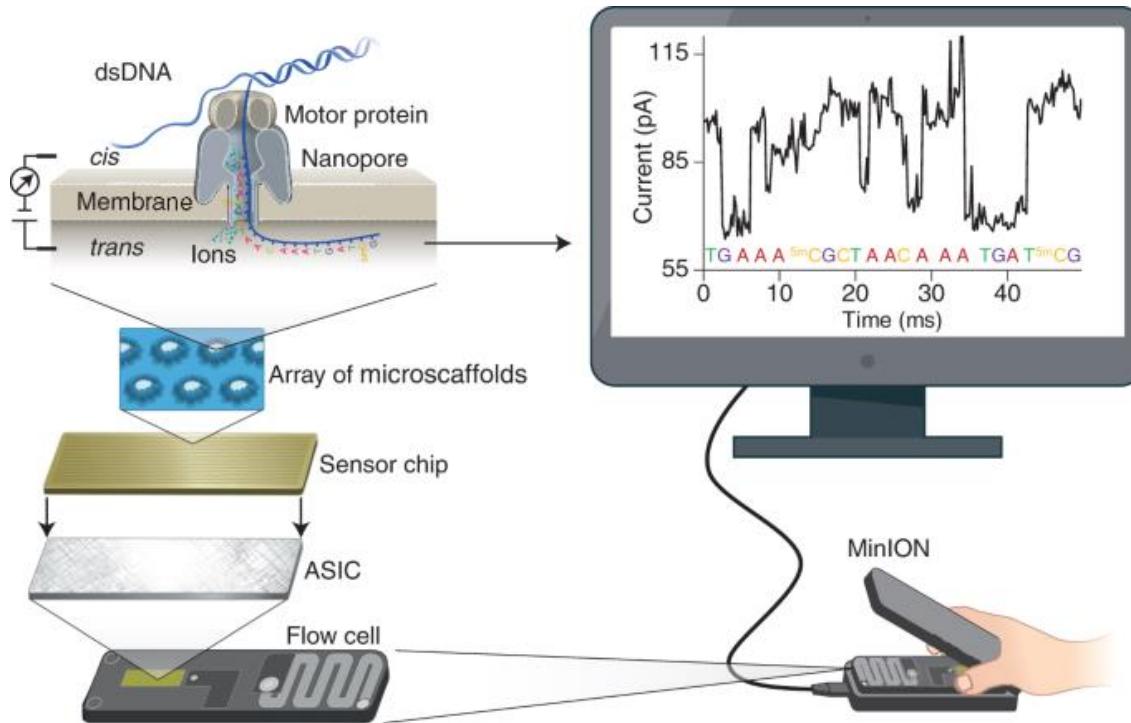
# Технология Oxford Nanopore

- 2005 – основана Oxford Nanopore Technologies
- 2014 – первый прибор MinION

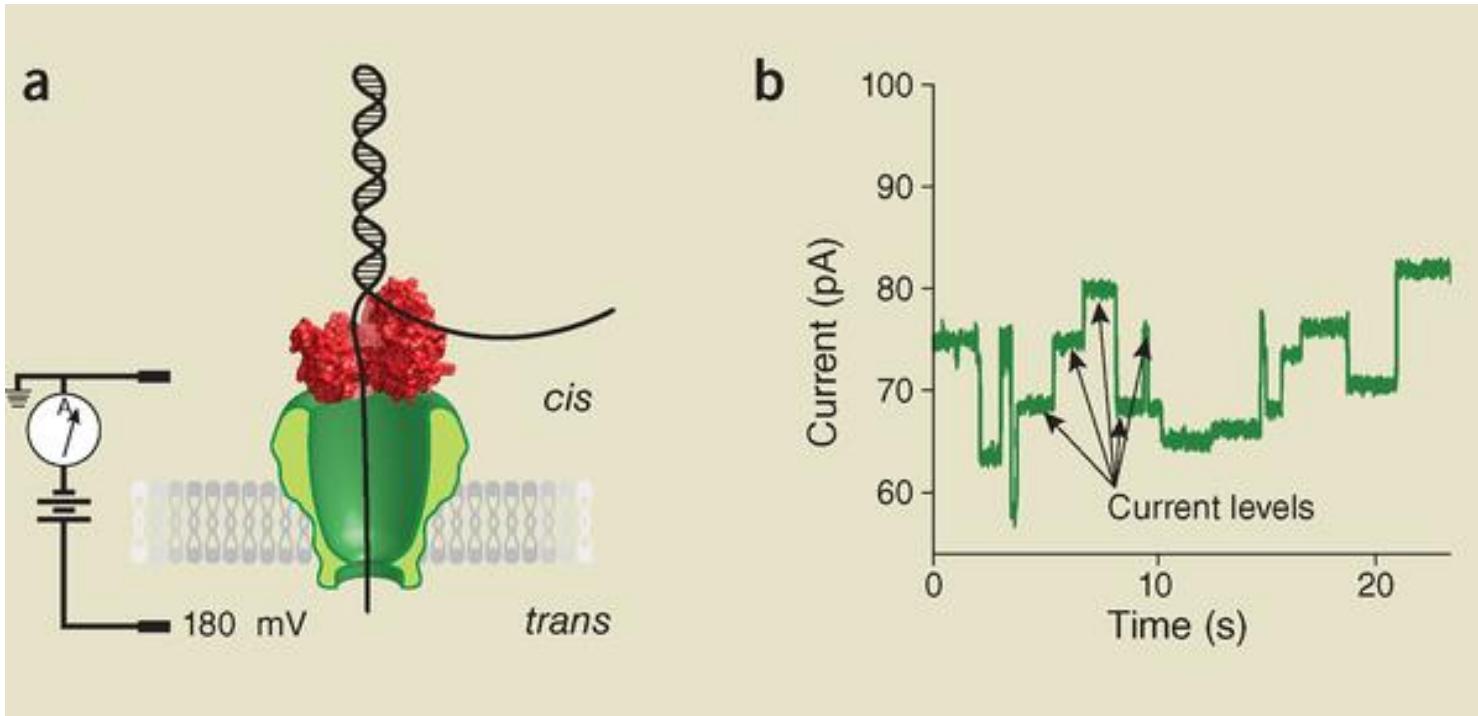


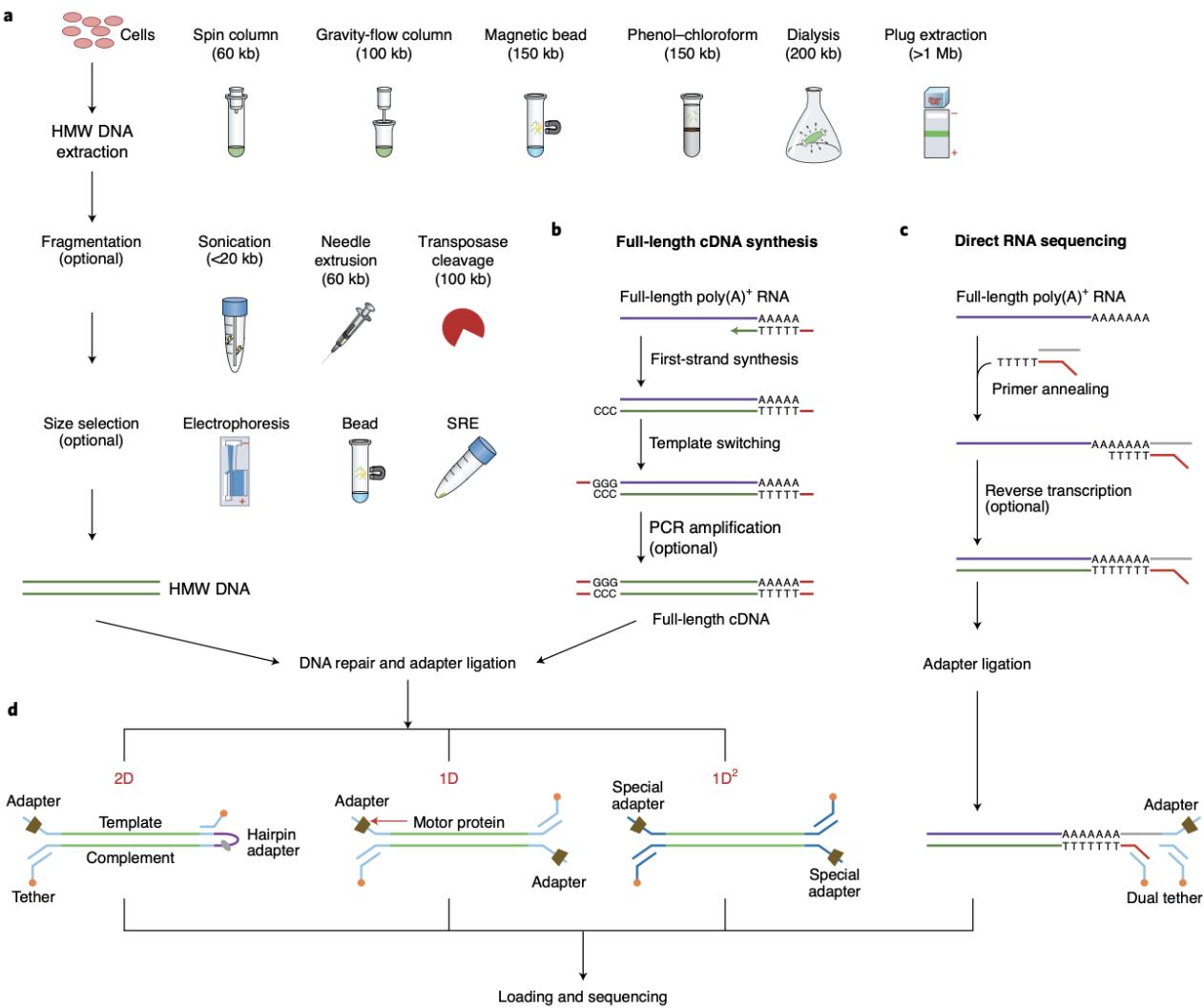
- Портативный прибор
- Точность 99,9%
- Определение структурных вариаций генома
- Может секвенировать эпигенетические модификации
- Может секвенировать РНК напрямую.

# Технология Oxford Nanopore



<https://youtu.be/RcP85JHLmnI>





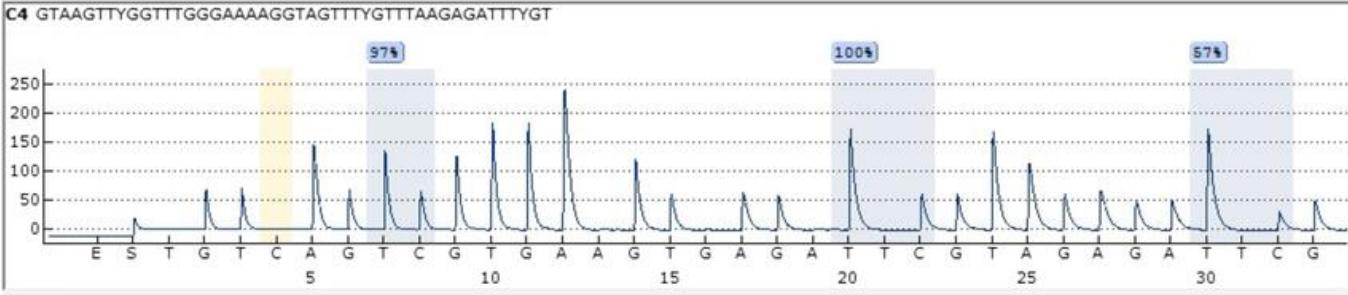
	Flongle	MinION	GridION (5 flow cells)	PromethION (48 flow cells)
				
Maximum run time	16 hours	72 hours	72 hours	64 hours
Theoretical 1D maximum yield	Up to 3.3 Gb	Up to 40 Gb	Up to 200 Gb	Up to 15 Tb
Current 1D maximum yield	Up to 2 Gb	Up to 30 Gb	Up to 150 Gb	Up to 8.6 Tb
Available channels	Up to 126	Up to 512	Up to 2,560	Up to 144,000

## The sequencing players at a glance

Company	Type of tech	Length of read	Technology details
Illumina	Short read	~500 bases	SBS-based
Illumina	Long read	~10,000 bases	Infinity: in early research
Oxford Nanopore	Long read	Up to 1 million bases, usually 5,000-20,000	Nanopore-based
Pacbio	Long read	Up to 50,000 bases, usually ~25,000	SBS-based
Pacbio	Short read	~500 bases	SBB-based, acquired via Omniome

SBS=*sequencing by synthesis*. SBB=*sequencing by binding*. Source: interviews.

# PyroMark



## Кратко обсудим плюсы и минусы разных методов

### Sequencing by synthesis

- Пиросеквенирование – по технологии компании 454 Life Sciences/Roche
  - Первая технология на рынке. С 2016 года не поддерживается.
- Секвенирование на основе обратимой терминации синтеза ДНК – по технологии компаний Solexa/Illumina.
  - 90% всего секвенирования выполняется сейчас на этой технологии.
- Полупроводниковое секвенирование - Ion Torrent
- DNBSEQ метод от компании MGI на основе ДНК-наношаров (CoolMPS версия)

### Sequencing by ligation

- SOLiD от Applied Biosystems. Не поддерживается
- DNBSEQ метод от компании MGI на основе ДНК-наношаров (cPAS версия)

### Sequencing by binding (PacBio)

### Третье поколение

SMRT от PacBio

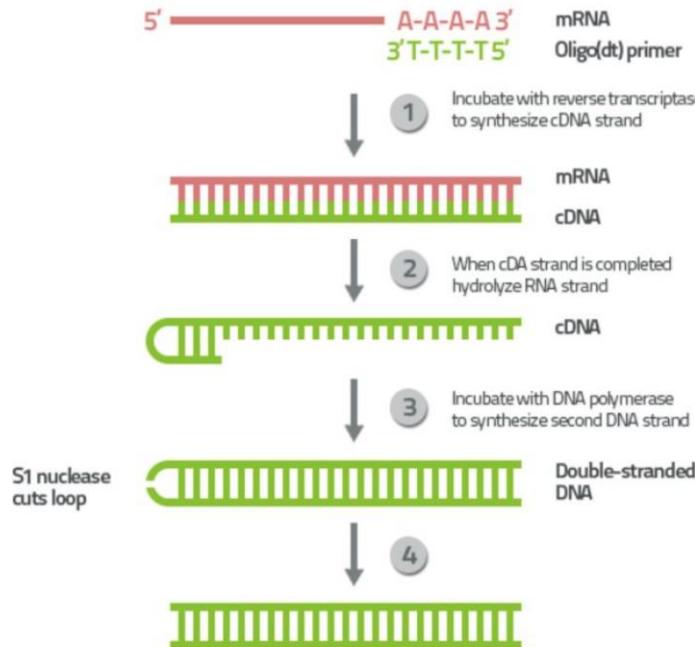
Nanopore Sequencing

# Производные методы секвенирования

- RNA-seq
- Модификации ДНК (бисульфитная конверсия, NanoPore)
- ChIP-seq, ChIP-exo
- Доступность хроматина (DNase-seq, FAIR-seq, ATAC-seq)
- Позиционирование нуклеосом (MNase-seq, Mnase-Exo III)
- Ribo-seq
- DAM-ID
- DAM-метилирование
- 3C-методы (HiC, Micro-C)
- ДНК-РНК
- Single cell

# Секвенирование РНК

Комплементарная ДНК (кДНК, cDNA), кДНК библиотеки

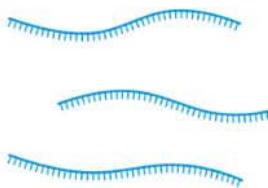


кДНК может быть секвенирована по Сэнгеру или NGS

# RNA-seq

## RNA Sequencing

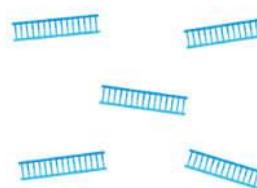
- 1 Isolate RNA from samples



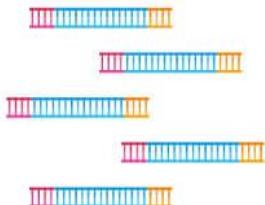
- 2 Fragment RNA into short segments



- 3 Convert RNA fragments into cDNA



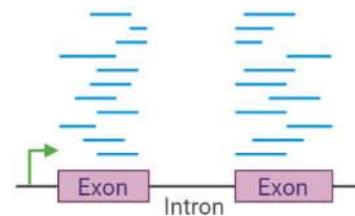
- 4 Ligate sequencing adapters and amplify



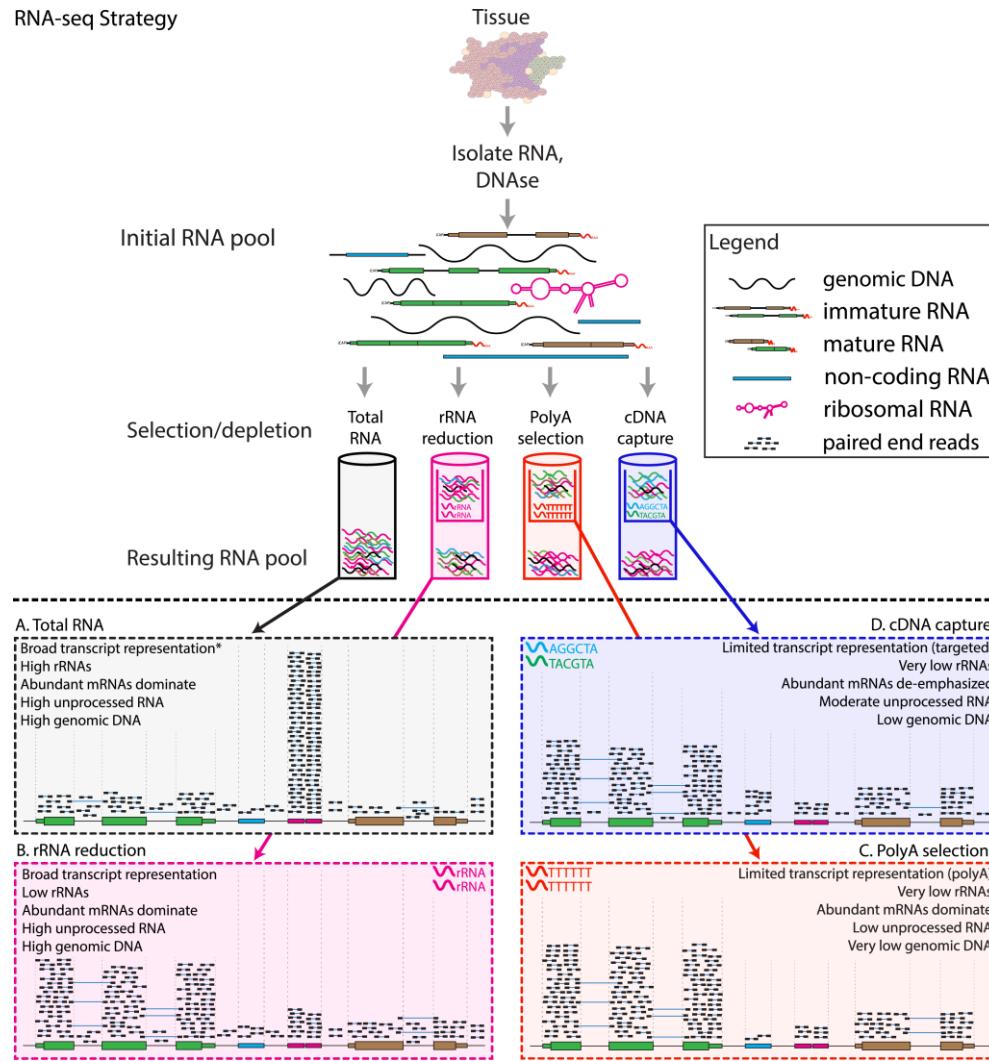
- 5 Perform NGS sequencing

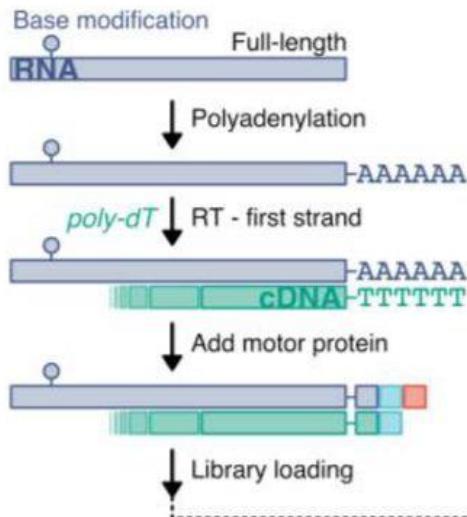
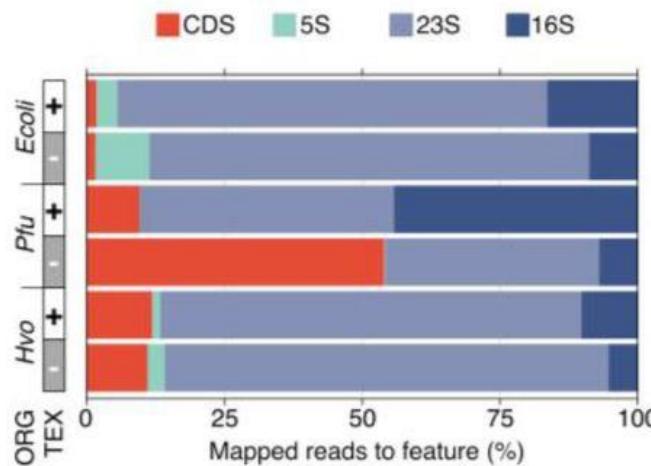


- 6 Map sequencing reads to the transcriptome/genome



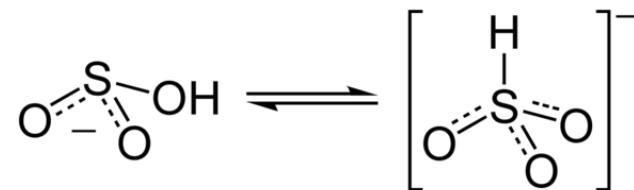
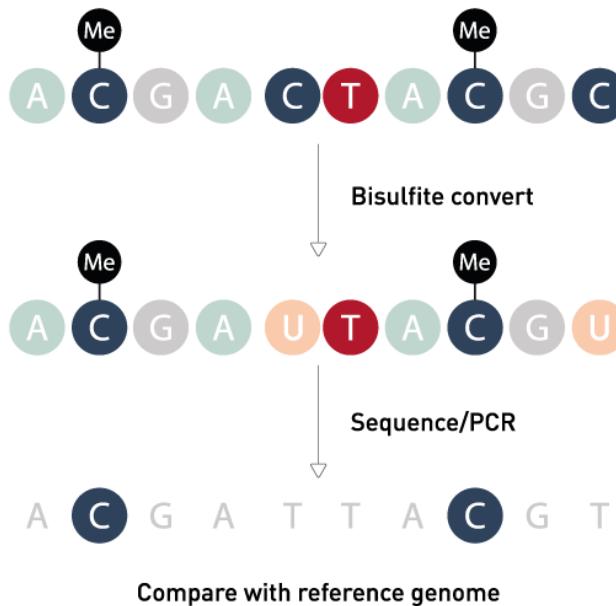
## RNA-seq Strategy



**a****b**

# Анализ модификаций ДНК

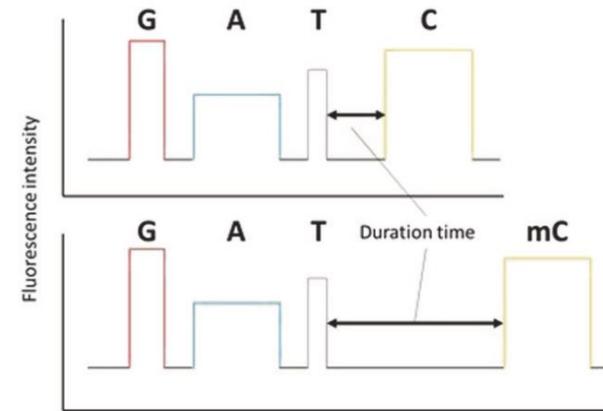
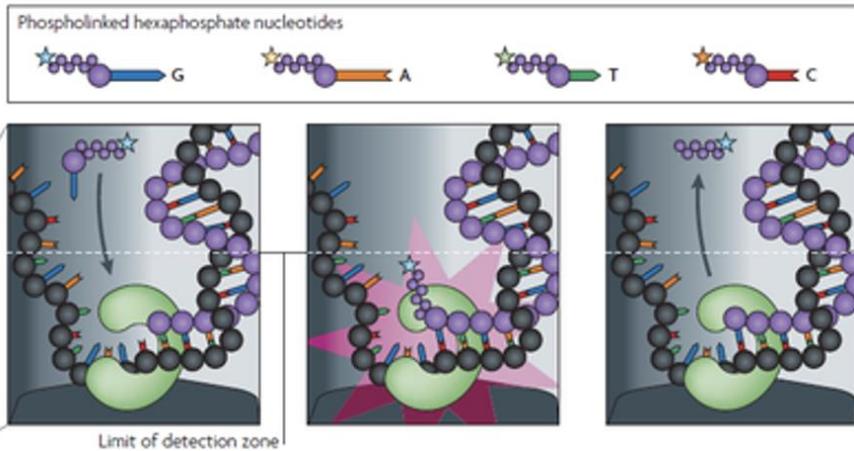
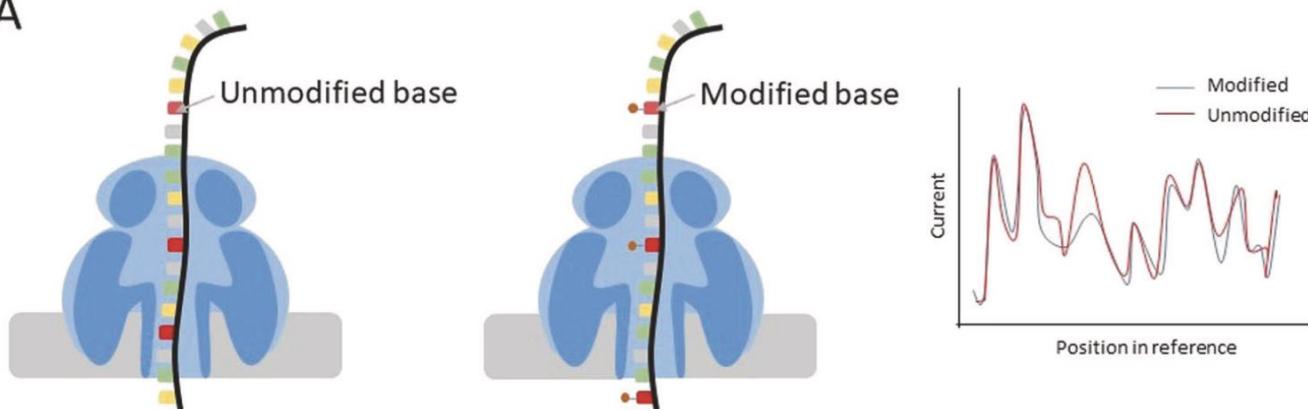
- Пример определение 5m-цитозинов в ДНК, бисульфитное секвенирование



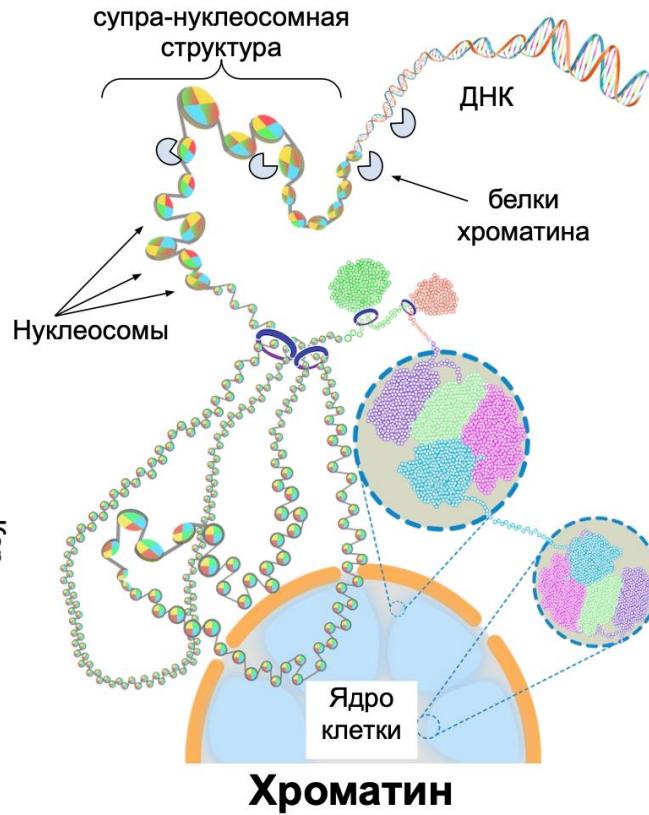
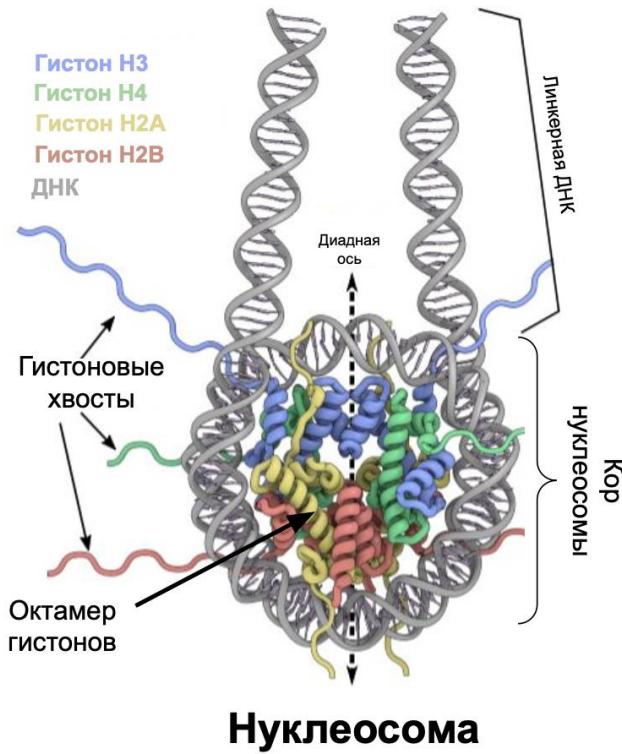
Бисульфит (гидросульфит)

# Oxford Nanopore и PacBio

A

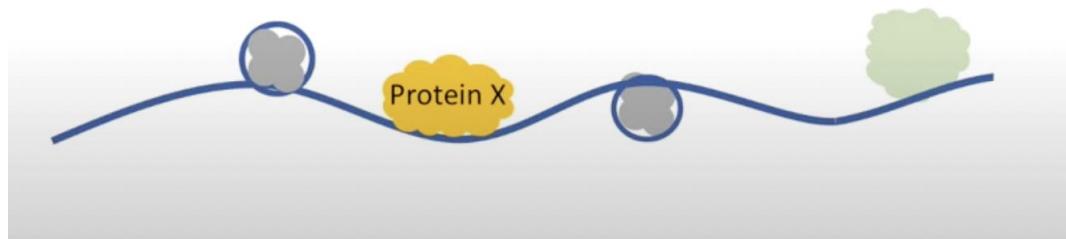


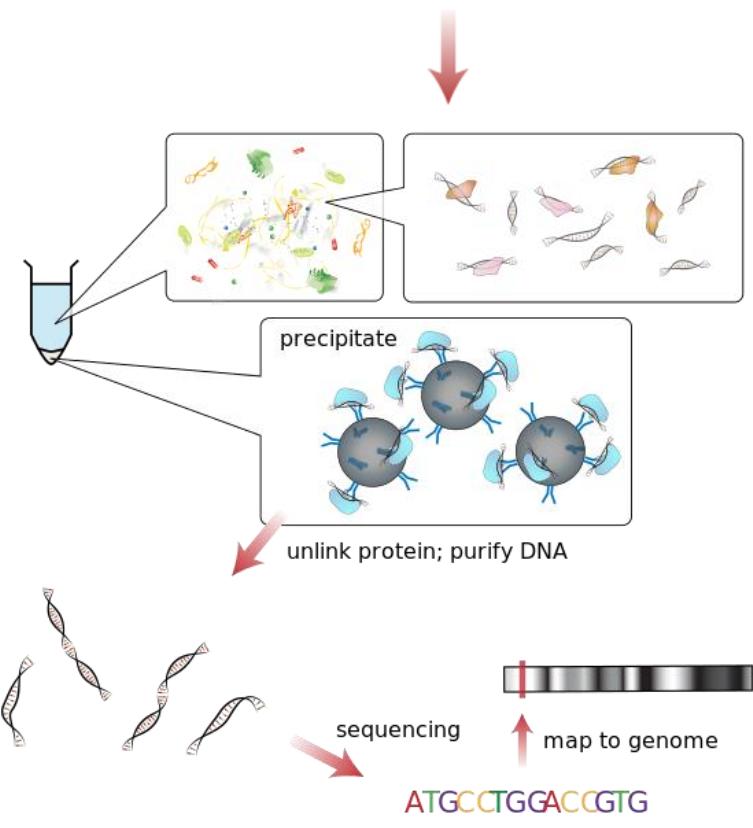
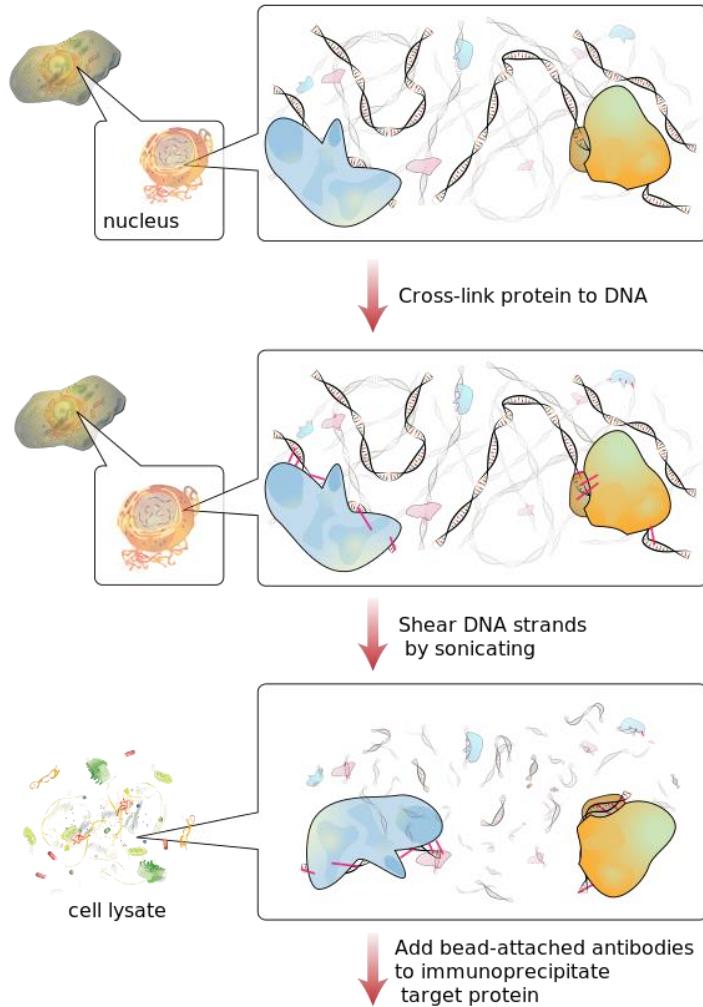
# ChIP-seq (Chromatin immunoprecipitation)



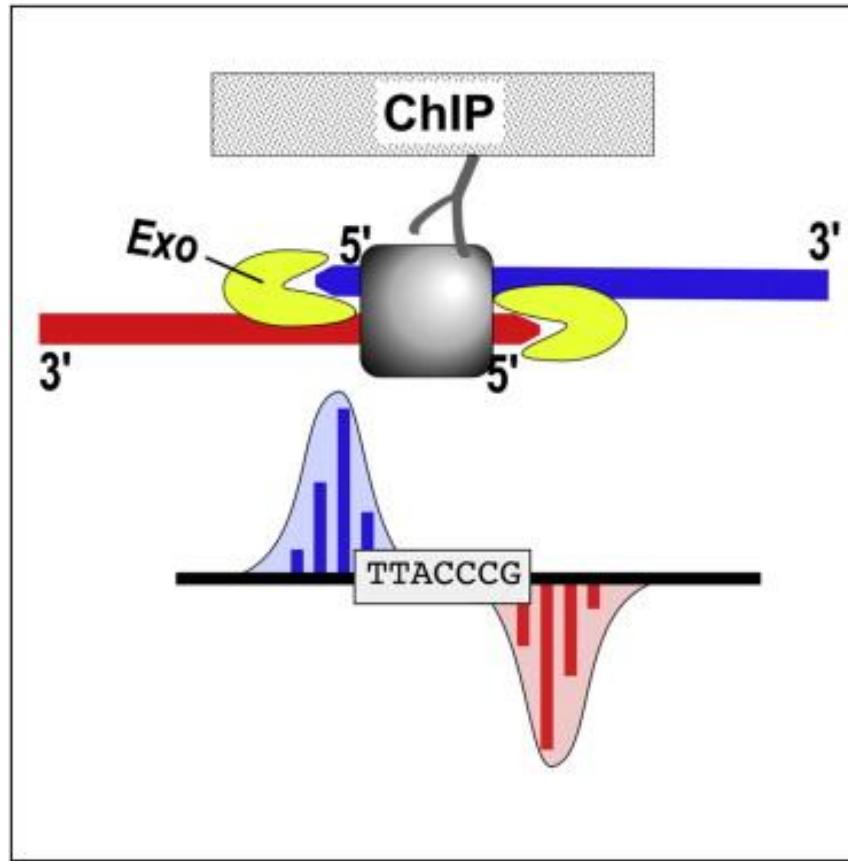
# ChIP-seq

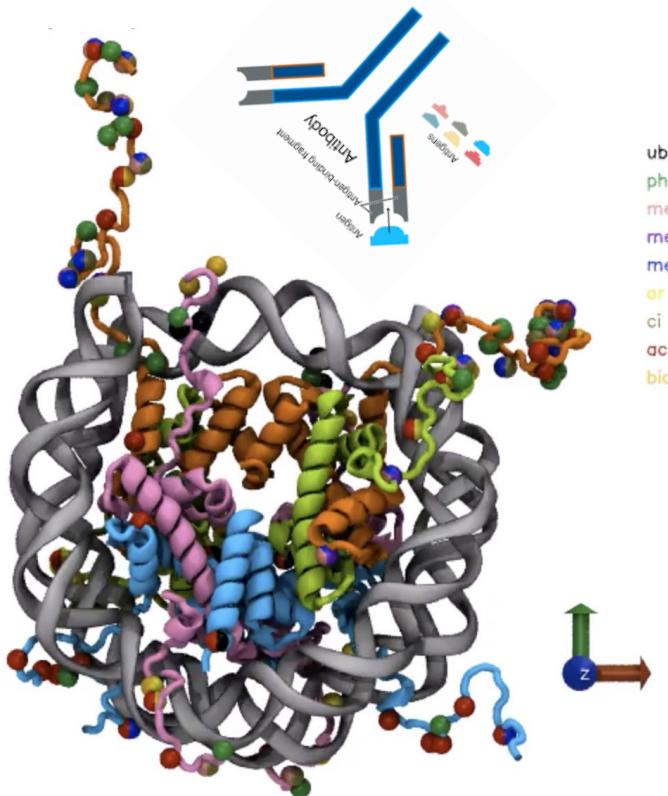
- Chromatin immunoprecipitation (ChIP)



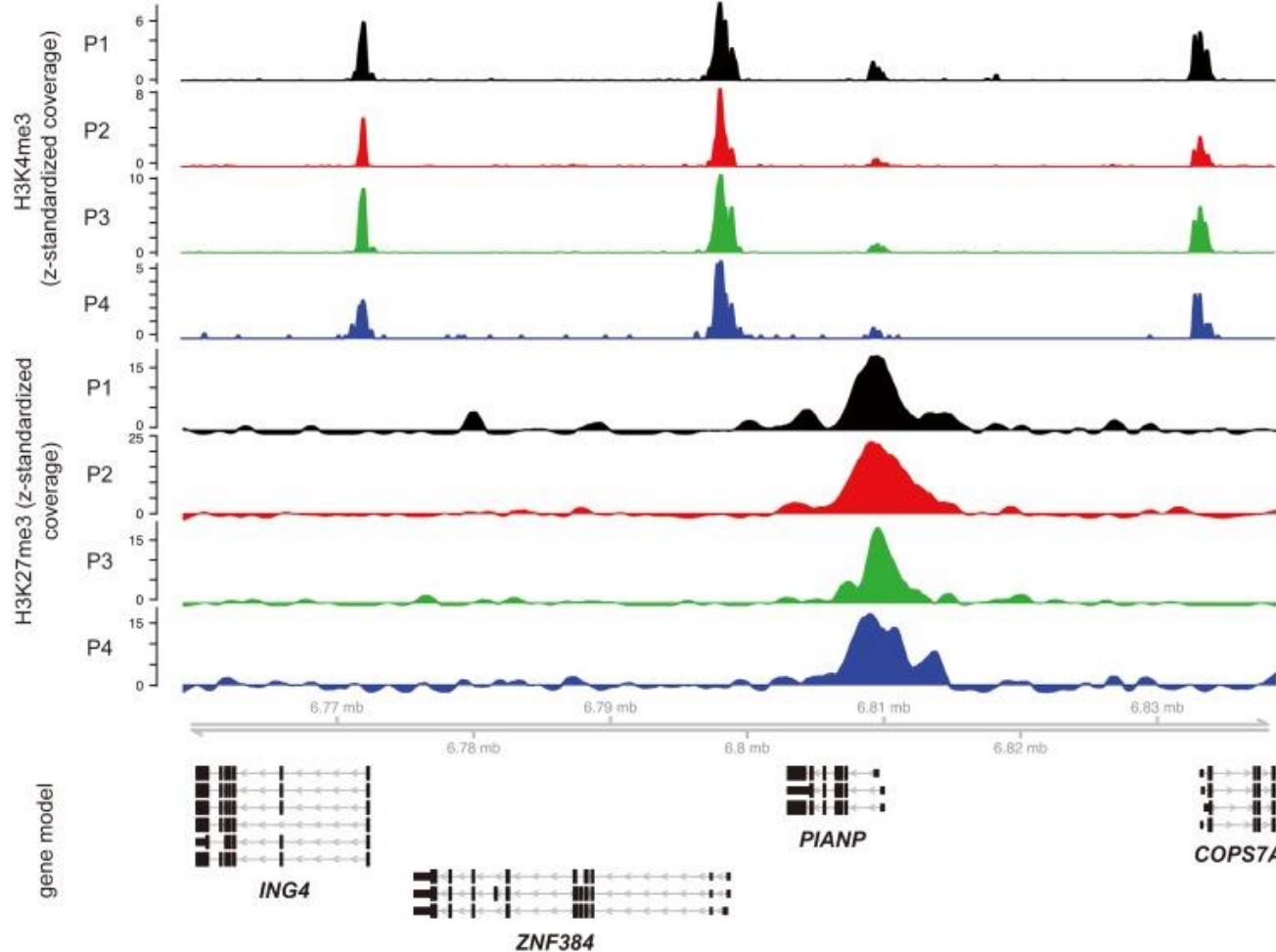


# ChIP-exo-seq



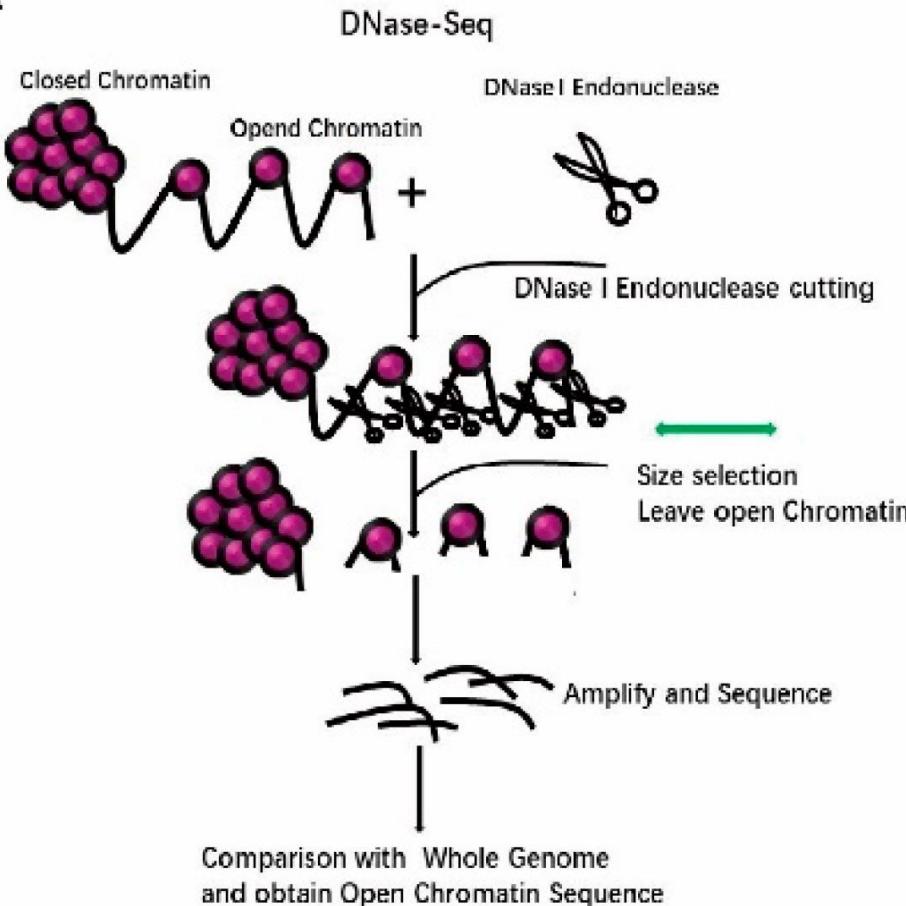


H2A	H2B	H3	H4
H2AS1ph	H2BK5ac	H3R2me1	H4S1ph
H2AR3me2	H2BK5me1	H3R2me2	H4R3me1
H2AR3ci	H2BK12ac	H3R2ci	H4R3ci
H2AK5ac	H2BS14ph	H3T3ph	H4K5ac
H2AK9ac	H2BK15ac	H3K4ac	H4K8ac
H2AK9bio	H2BK16ac	H3K4me1	H4K8bio
H2AK13bio	H2BK20ac	H3K4me2	H4K12ac
H2AK13or	H2BK30ar	H3K4me3	H4K12bio
H2AK19ub	H2BK46ac	H3S6ph	H4K16ac
H2AT120ph	H2BK120ac	H3T6ph	H4K16ar
H2AK121ub	H2BK120ub	H3R8ci	H4K20me1
H2AK125bio		H3K9ac	H4K20me2
H2AK127bio		H3K9me1	H4K20me3
H2AK129bio		H3K9me2	H4K91ac
H2AS137ph		H3K9me3	H4K91ub
H2AS139ph		H3K9bio	
H2AY142ph		H3S10ph	
		H3T11ph	
		H3K14ac	
		H3R17me1	
		H3R17me2	
		H3R17ci	
		H3K18ac	
		H3K18bio	
		H3K23ac	
		H3R26me1	
		H3R26ci	
		H3K27ac	
		H3K27me1	
		H3K27me2	
		H3K27me3	
		H3K27or	
		H3S28ph	
		H3S31ph	
		H3K36ac	
		H3K36me3	
		H3K36me1	
		H3K36me2	
		H3K37ar	
		H3Y41ph	
		H3T45ph	
		H3K56ac	
		H3K79me1	



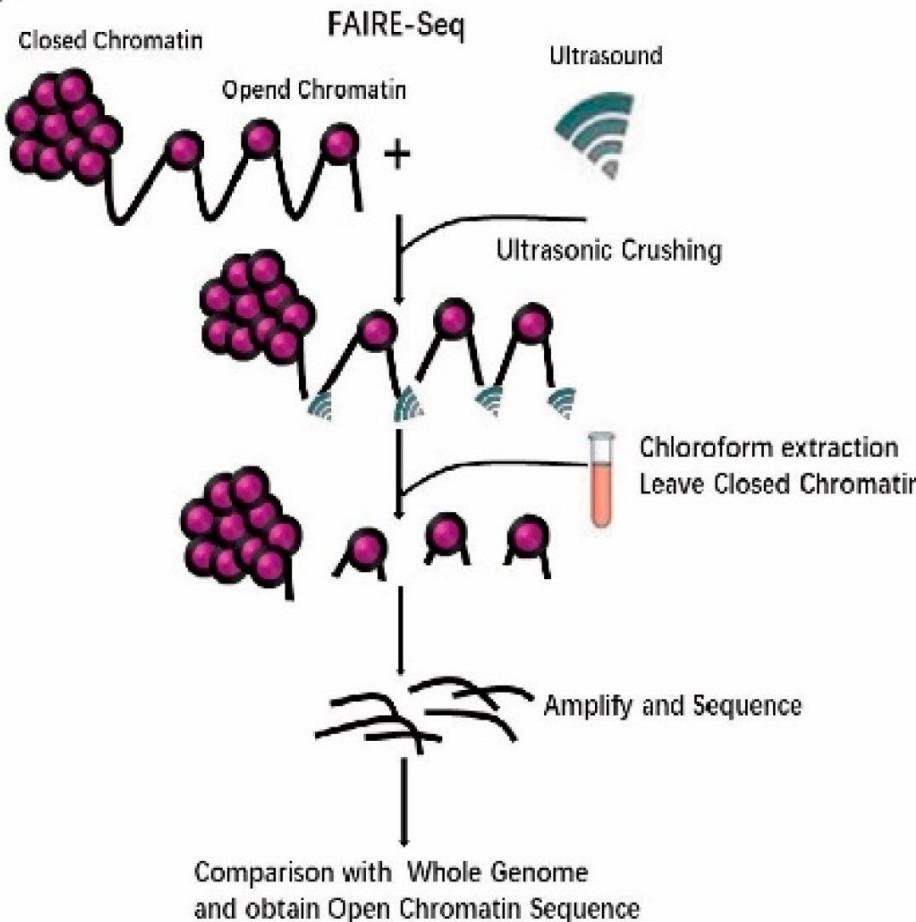
# Определение доступности хроматина

A



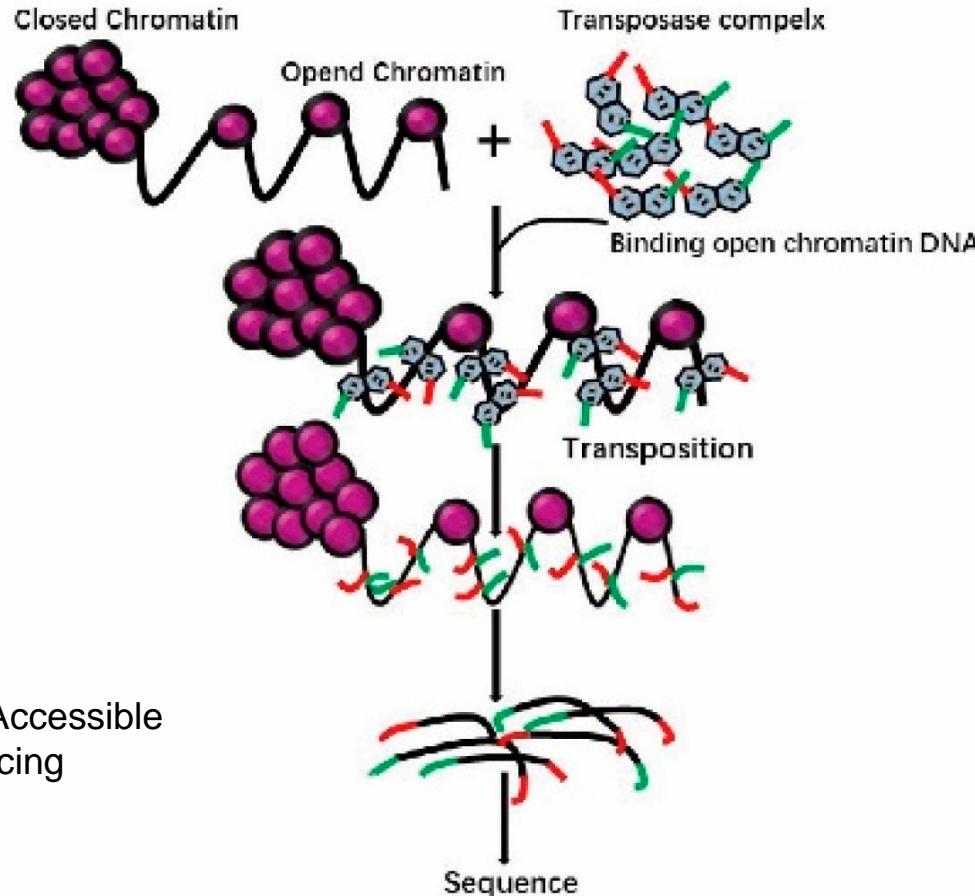
# Определение доступности хроматина

B



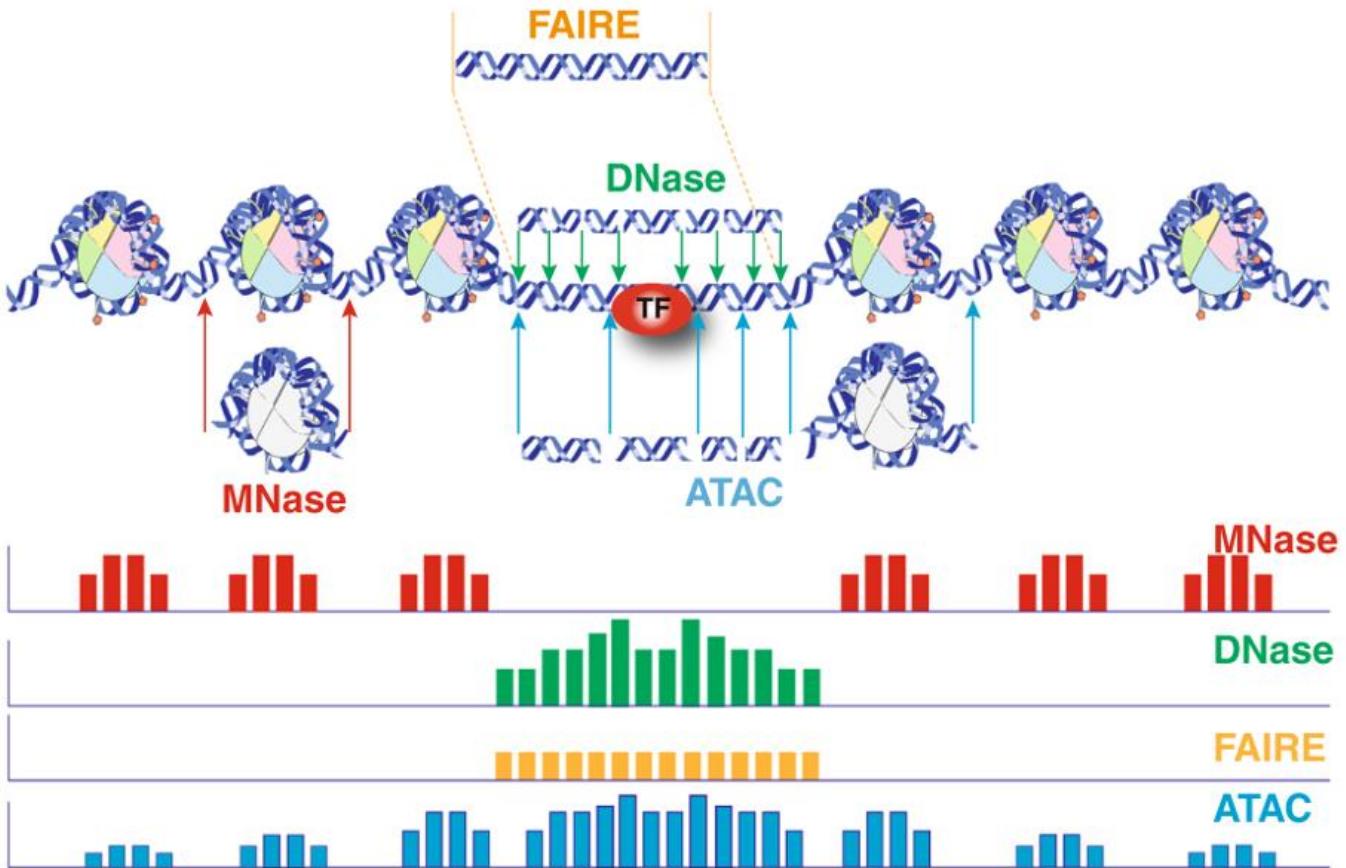
# Определение доступности хроматина

ATAC-Seq

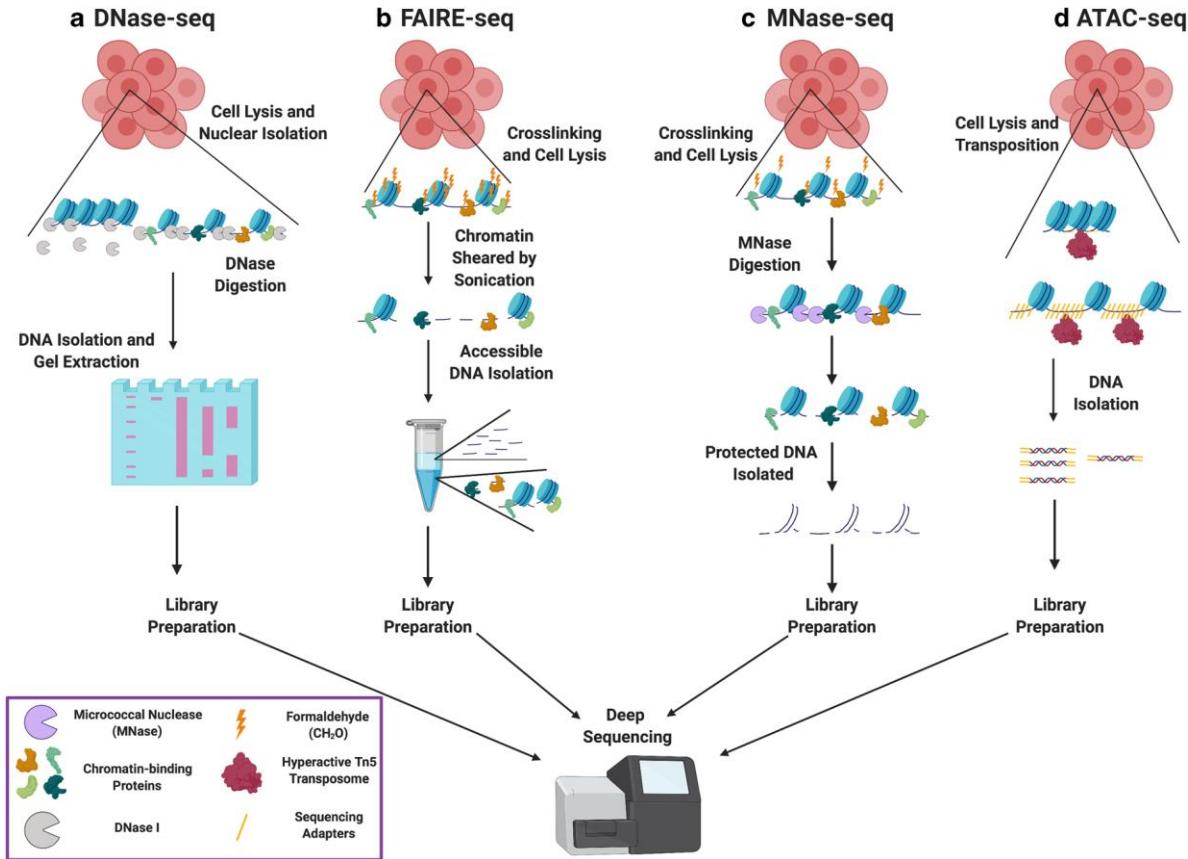


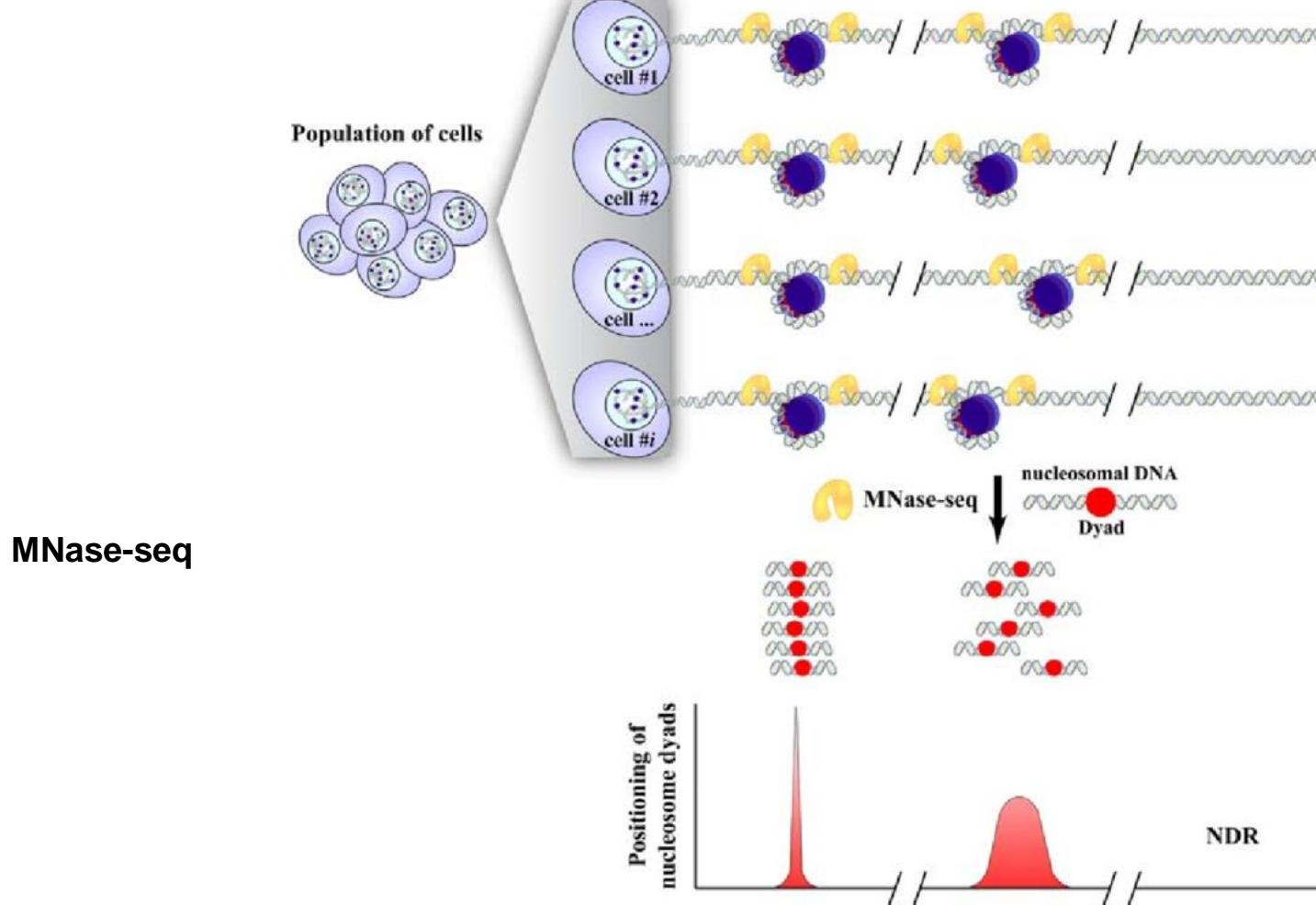
Assay for Transposase-Accessible  
Chromatin using sequencing

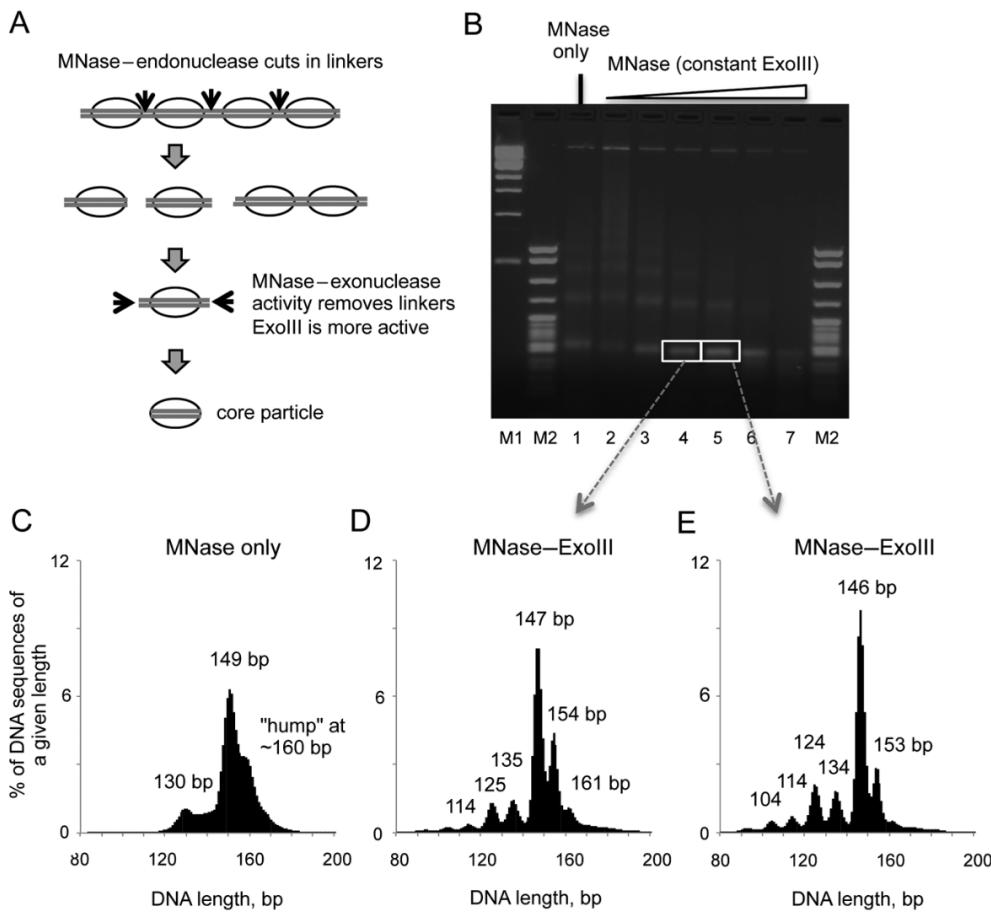
# Определение доступности хроматина



# Определение доступности хроматина

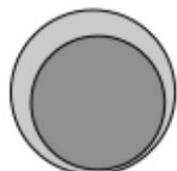




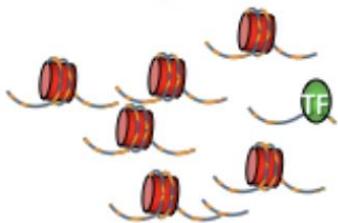


**Figure 1.** Simultaneous digestion of yeast chromatin with MNase and ExoIII results in a set of DNA fragments of discrete lengths, corresponding to the nucleosome core particle and its intermediates. (A) Digestion

- 10.1093/nar/gkv943

**A****ChIP**

Crosslink and shatter



Specific antibody

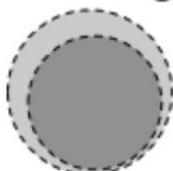
Precipitate

Extract DNA

Adapter ligation

Amplify library

Sequence

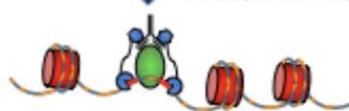
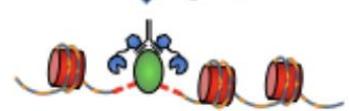
**B****CUT&Tag**

permeabilize

Specific antibody



Protein A-Tn5

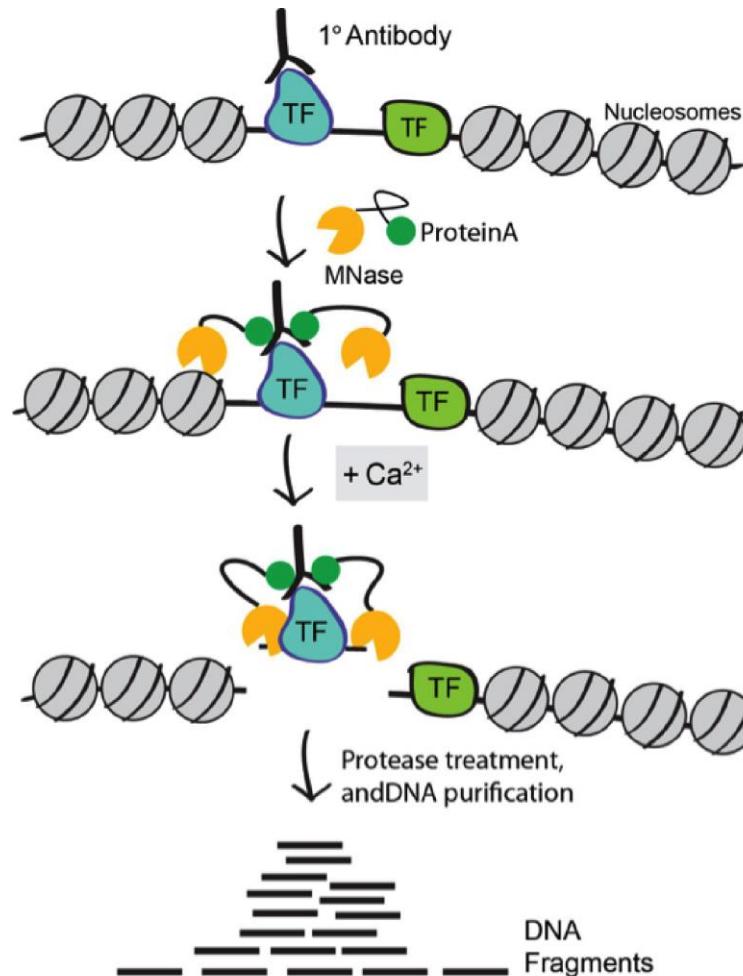
Mg<sup>++</sup> @ 37°C

Extract DNA

Amplify library

Sequence

**CUT&TAG**



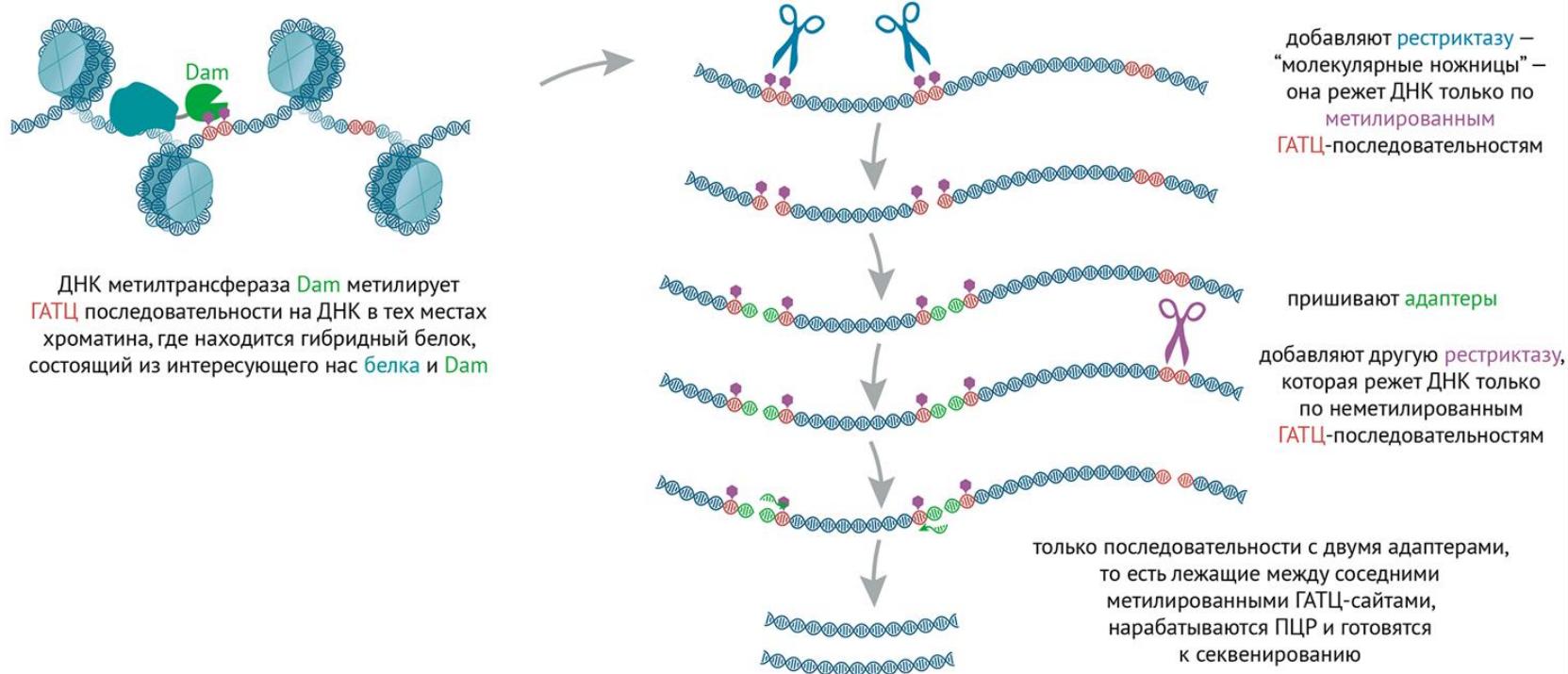
# CUT&RUN

Figure 1. CUT&RUN schematic (see text for details).

# CUT&Tag vs. CUT&RUN vs. ChIP-Seq

	CUT&Tag	CUT&RUN	ChIP-Seq
<b>Performed Under Native Conditions?</b>	Yes	Yes	No
<b>Chromatin Fragmentation Method</b>	Tn5-based tagmentation	MNase digestion	Sonication
<b>Cell Number Requirements</b>	5,000-500,000 cells	500,000 cells	1-10 million cells
<b>Sequencing Depth Required *</b>	2 million reads **	8 million reads	20-50 million reads
<b>Integrated Library Preparation?</b>	Yes, uses tagmentation	No, separate library prep required	No, separate library prep required
<b>Compatible Targets</b>	Primarily histone modifications, some transcription factors and co-factors	Wide range of histone modifications, transcription factors, and co-factors	Wide range of histone modifications, transcription factors, and co-factors
<b>Workflow Length</b>	1-2 days	1-2 days	2-3 days

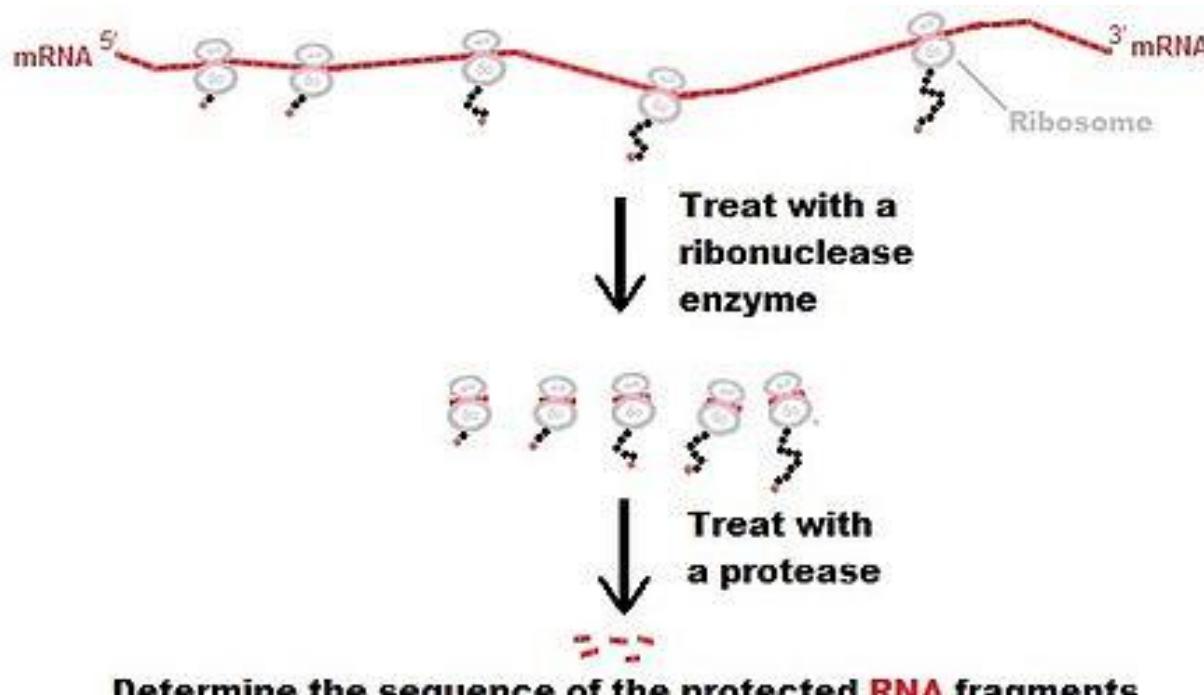
# DamID – альтернатива ChIP-seq *in vivo*



<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6451315/>

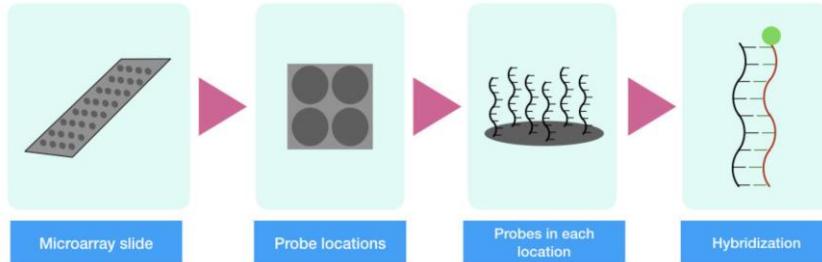
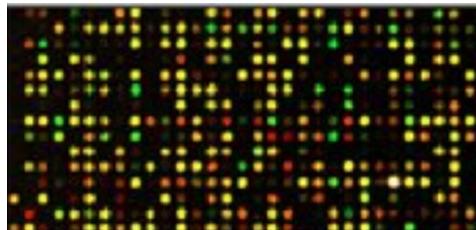
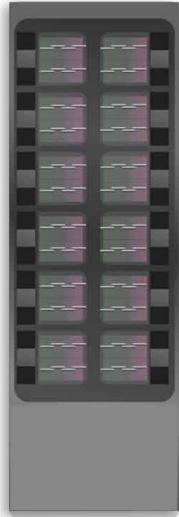
<https://biomolecula.ru/articles/molekuly-i-epigenom#metody-izucheniiia-dnk-i-khromatina>

## Ribo-seq – определение мРНК, которые активно транслируются



Compare those sequences to the reference genome  
to determine EXACTLY where the ribosomes were.

# Методы ДНК микрочипов

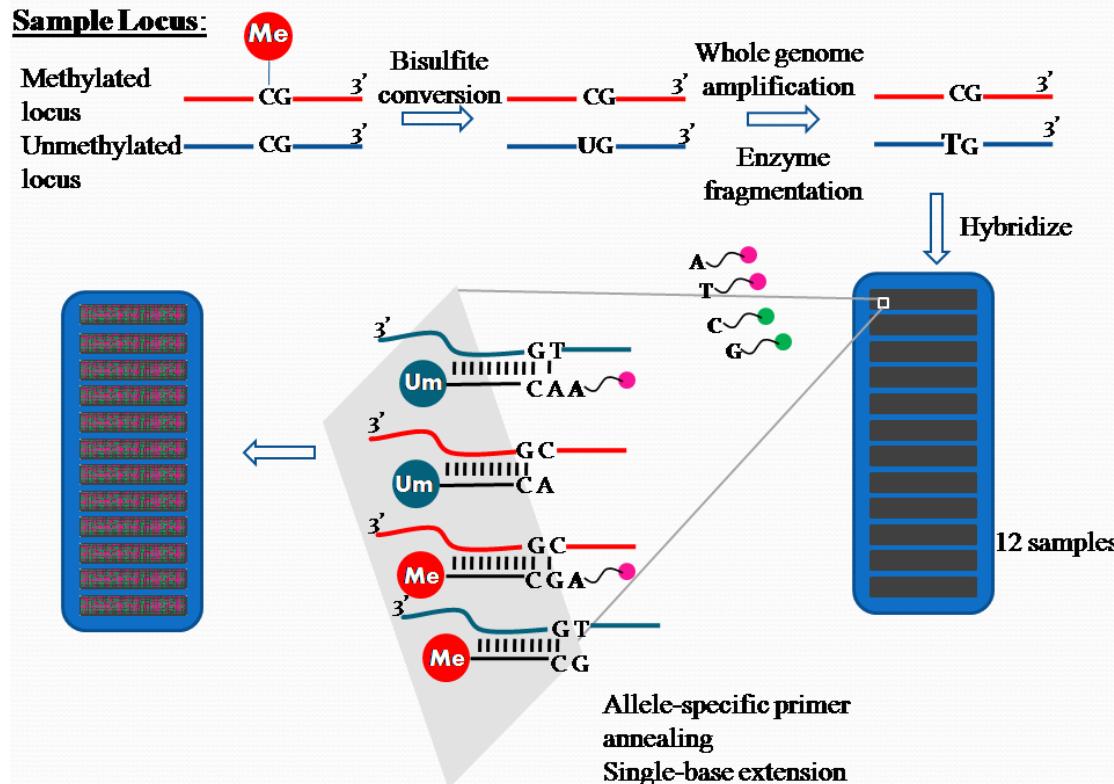
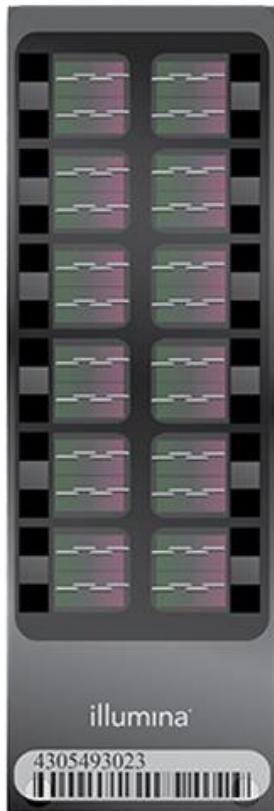


Андрей Дарьевич Мирзабеков



Дата рождения 19 октября 1937  
Место рождения Баку, СССР  
Дата смерти 13 июля 2003 (65 лет)

# Illumina Infiuum Methylation Assay



Infinium MethylationEPIC BeadChip ("EPIC") interrogates over 850,000 methylation sites across the human genome

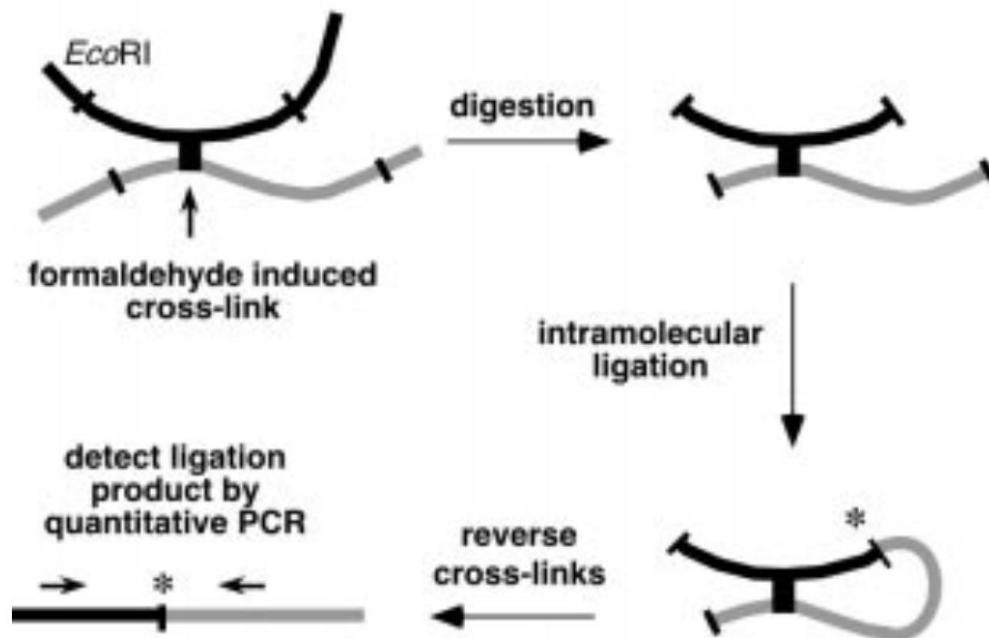
**3-С-методы – определение контактов ДНК в  
геноме (3Д-укладка генома)**

**CHROMOSOME CONFORMATION CAPTURE**

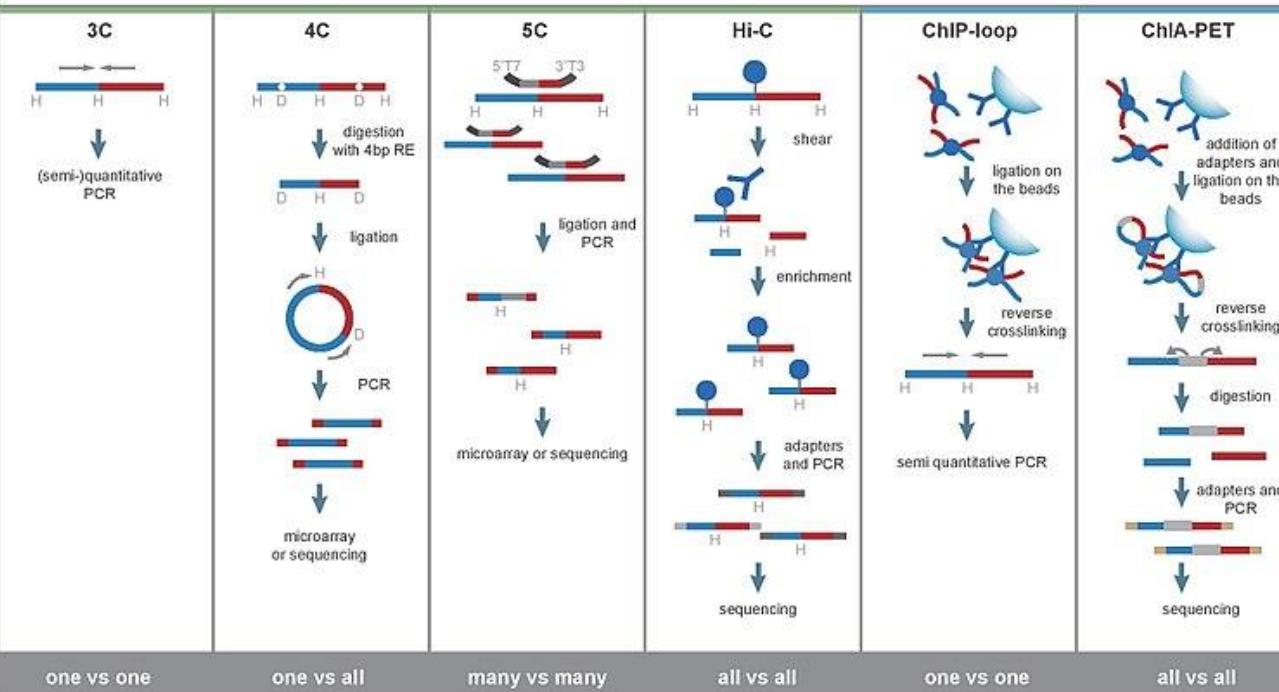
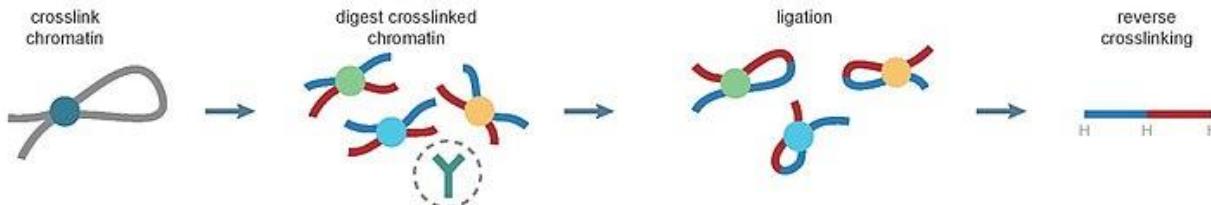
# Capturing Chromosome Conformation

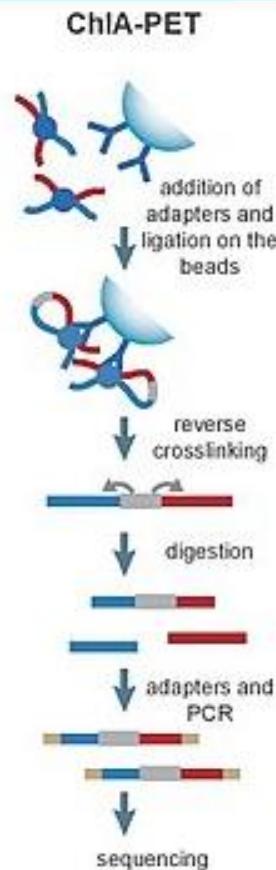
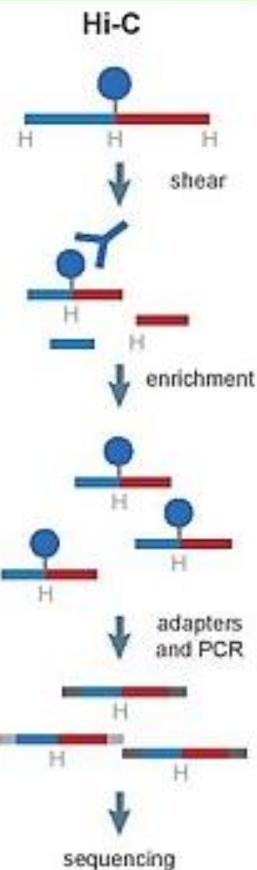
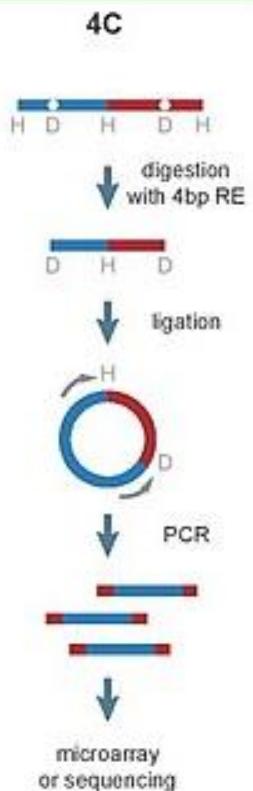
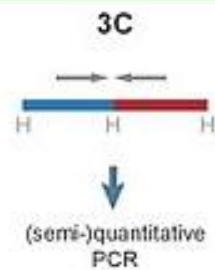
Job Dekker,<sup>1\*</sup> Karsten Rippe,<sup>2</sup> Martijn Dekker,<sup>3</sup> Nancy Kleckner<sup>1</sup>

15 FEBRUARY 2002 VOL 295 SCIENCE



## Chromosome Conformation Technologies



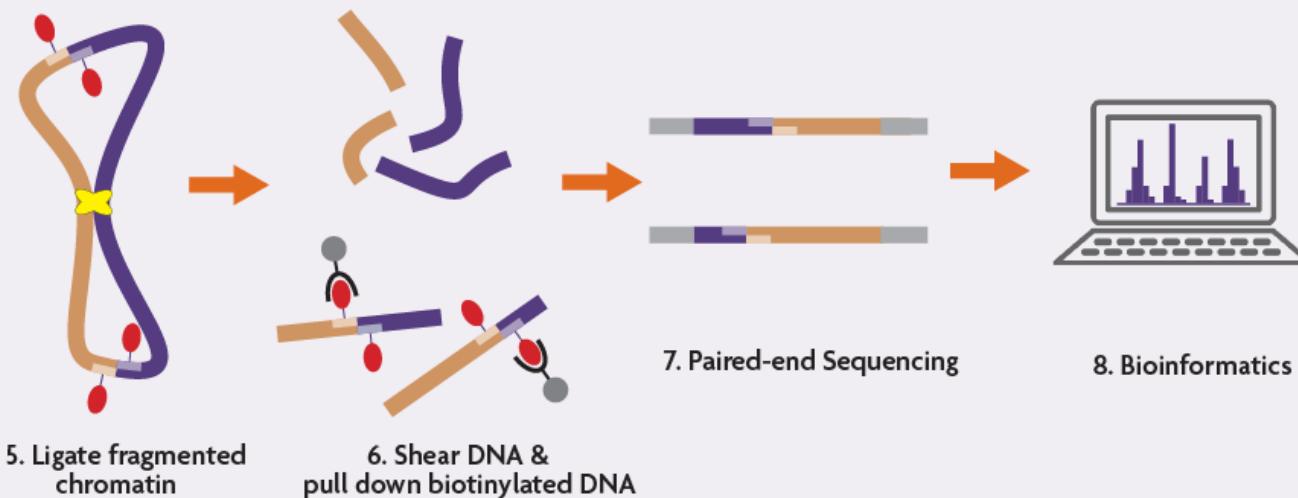
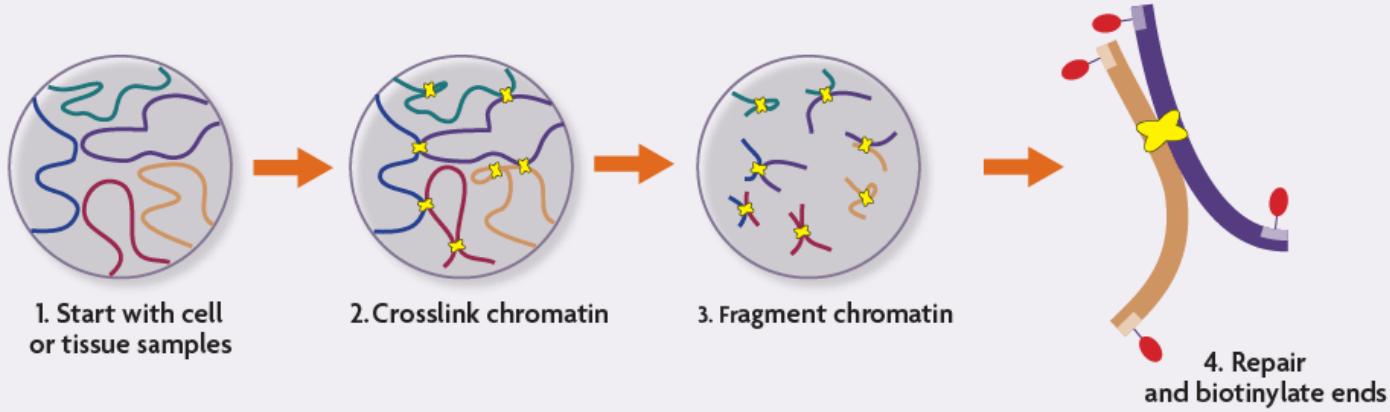


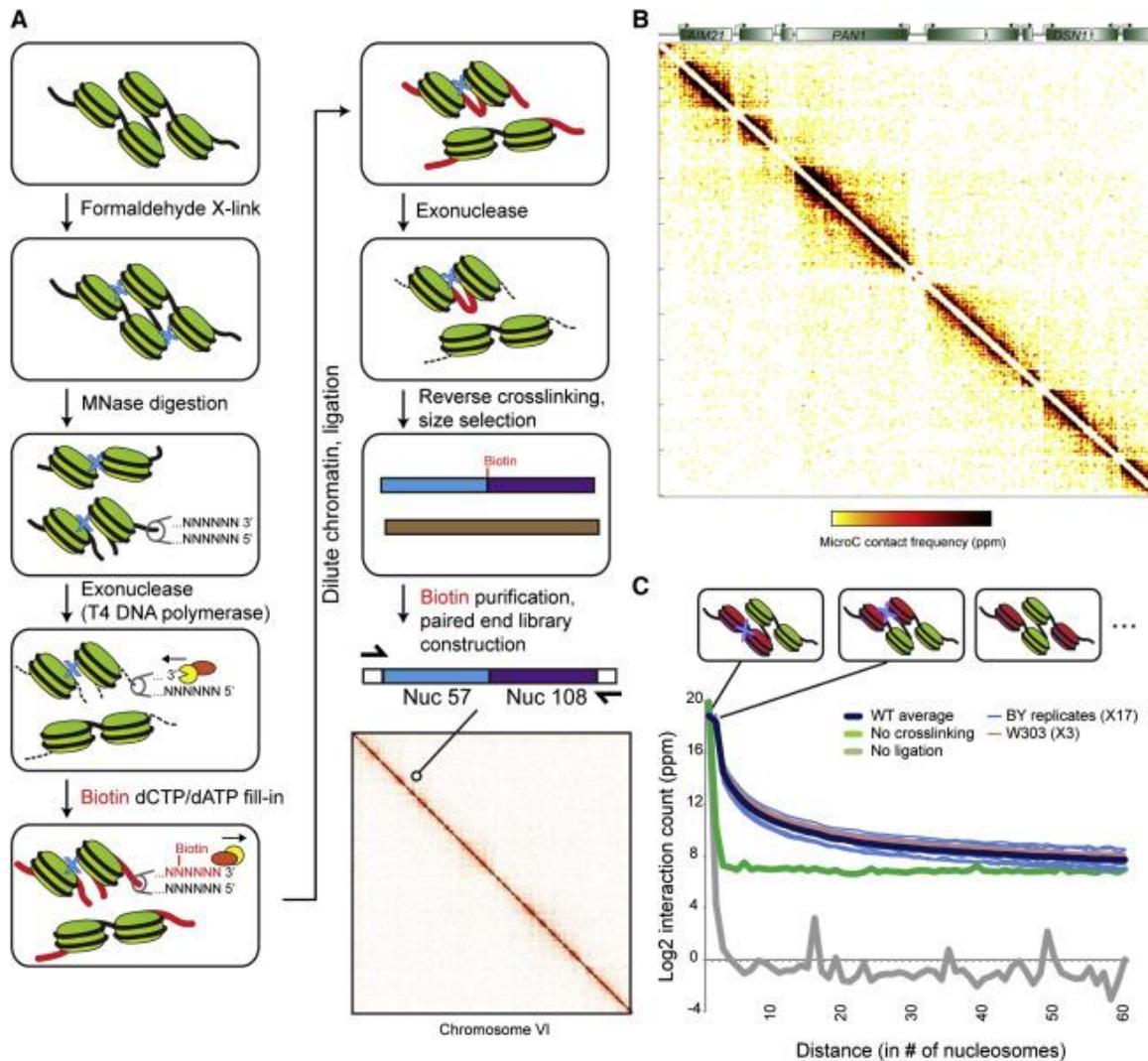
one vs one

one vs all

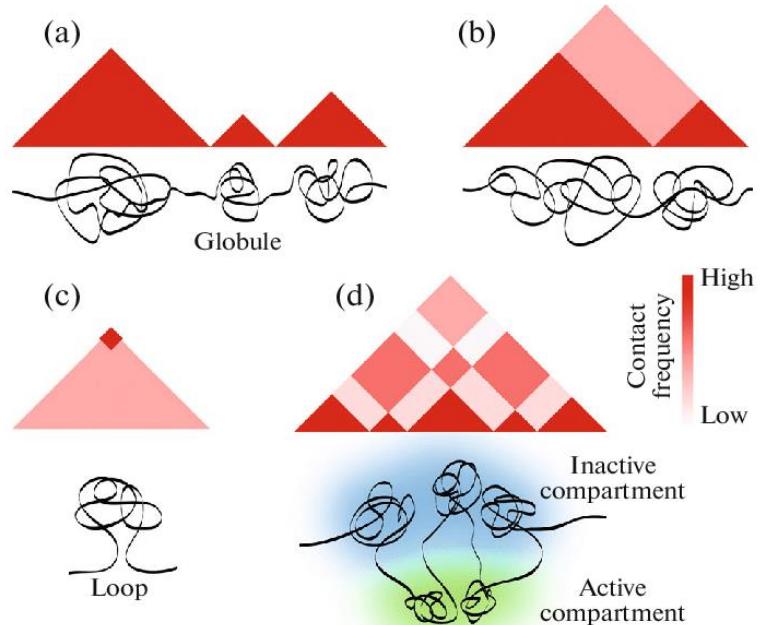
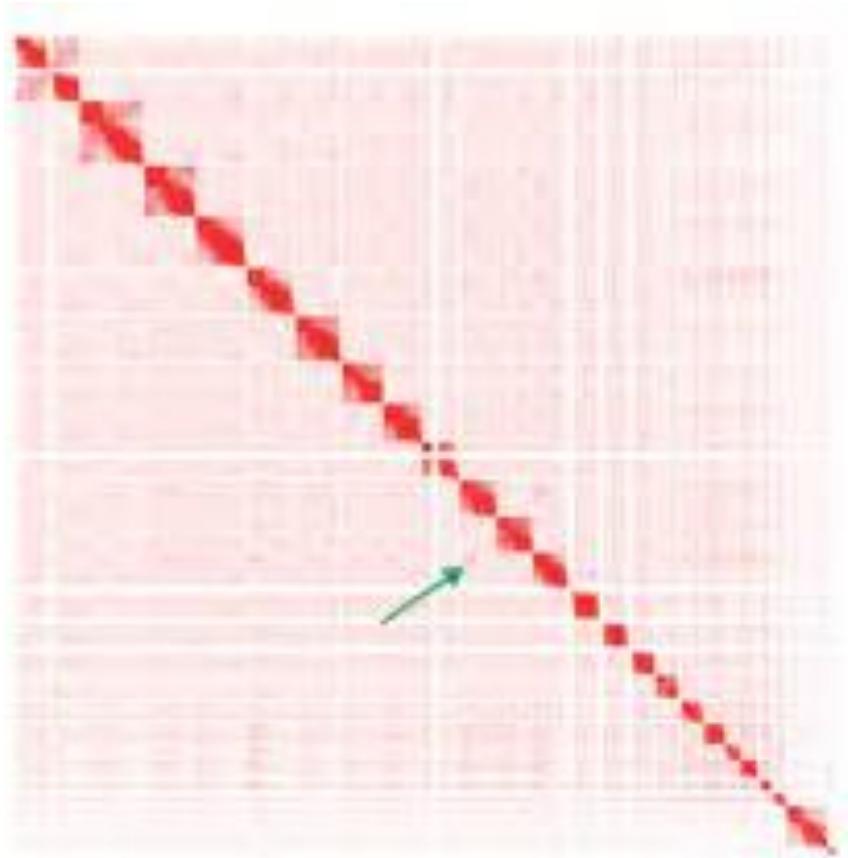
all vs all

all vs all



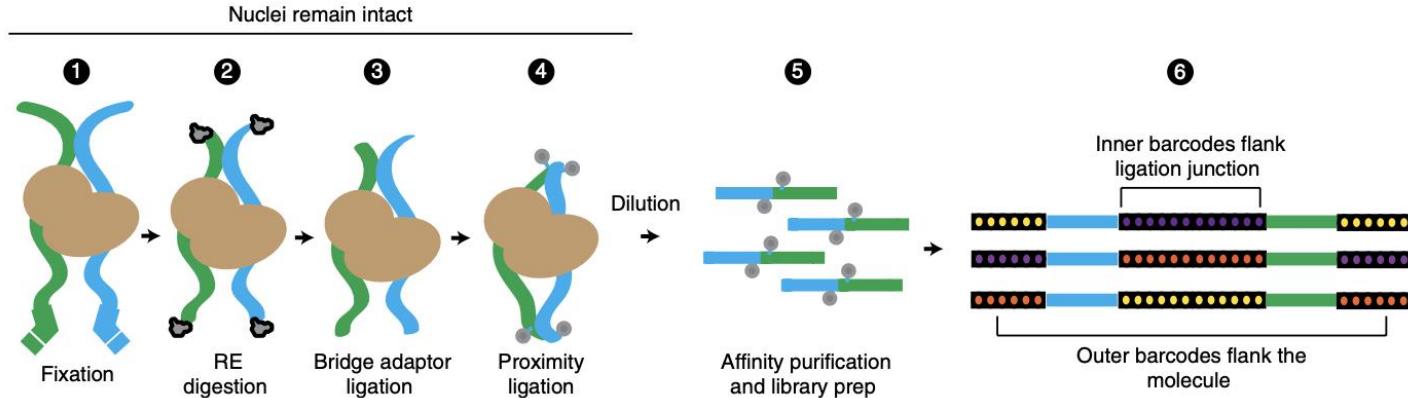


# Micro-C

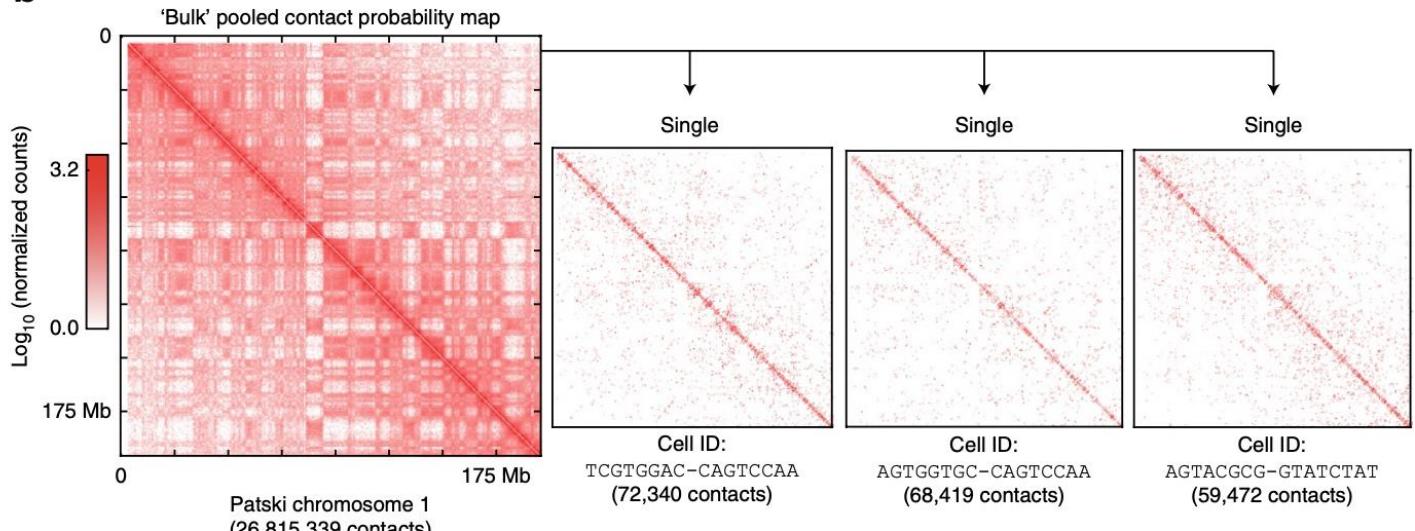


# Single cell Hi-C

a

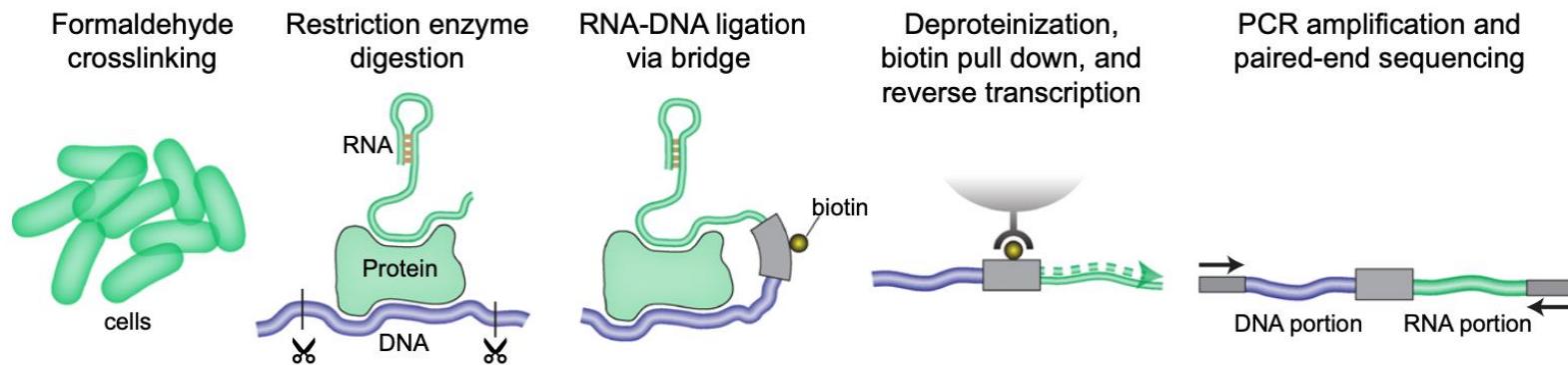


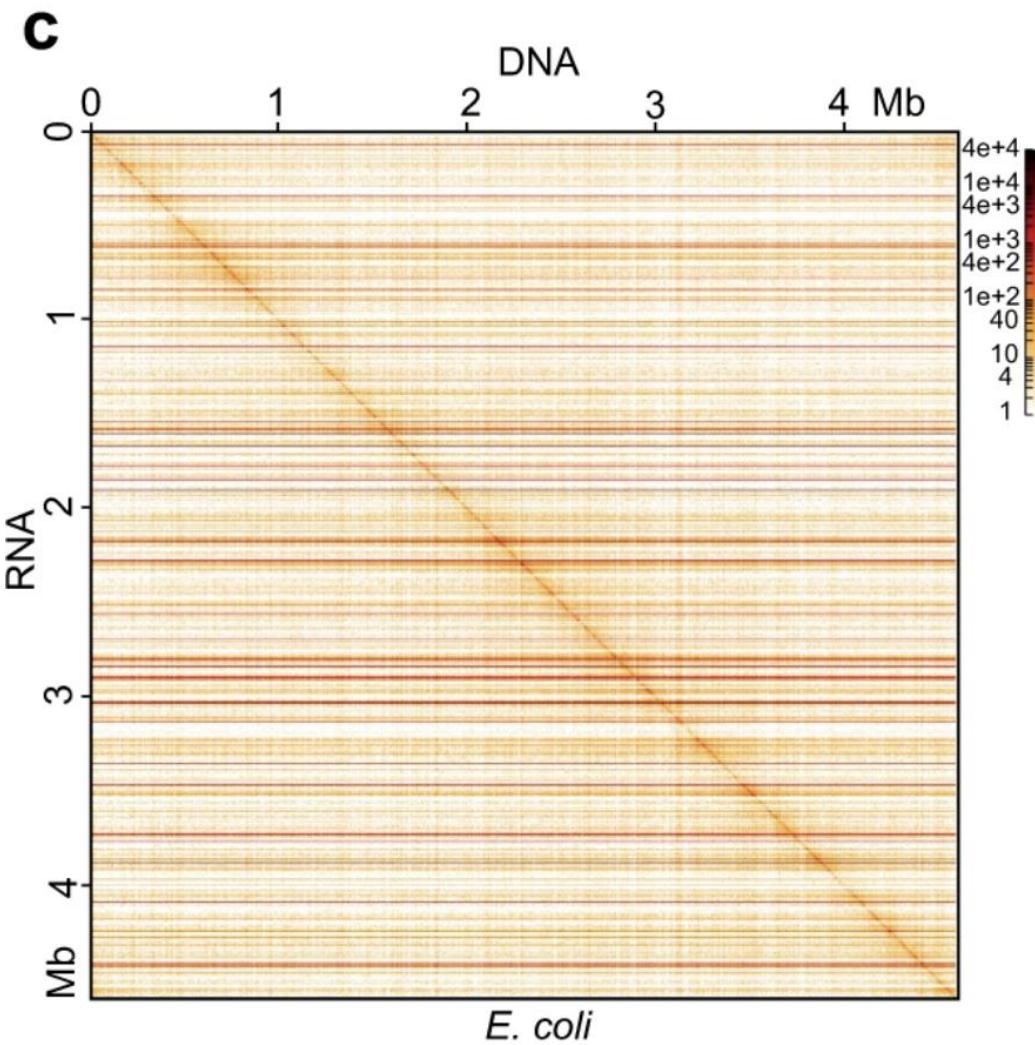
b



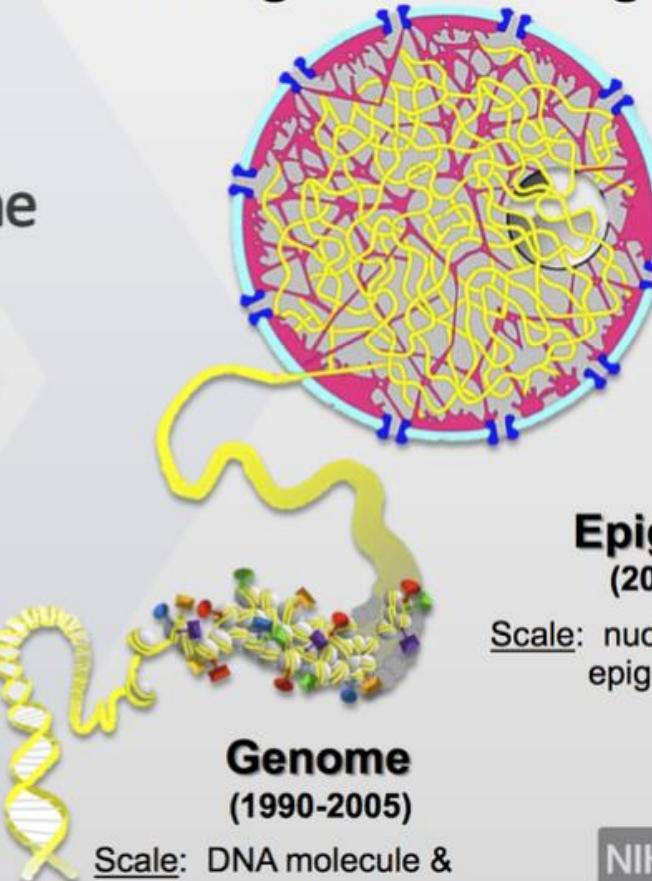
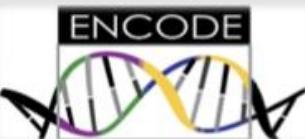
# RNA-DNA interactomes, RedC method

a





# Finishing the Job: Understanding Genome Organization



## Genome (1990-2005)

Scale: DNA molecule & sequence

## 3D Nucleome (2015-2022?)

Scale: cell nucleus & chromatin domains

## Epigenome (2005-2015)

Scale: nucleosome & epigenetic marks



National Institutes of Health  
Office of Strategic Coordination - The Common Fund

# ENCODE Project Writes Eulogy For Junk DNA

SCIENCE VOL 337 7 SEPTEMBER 2012

## ENCODE By the Numbers

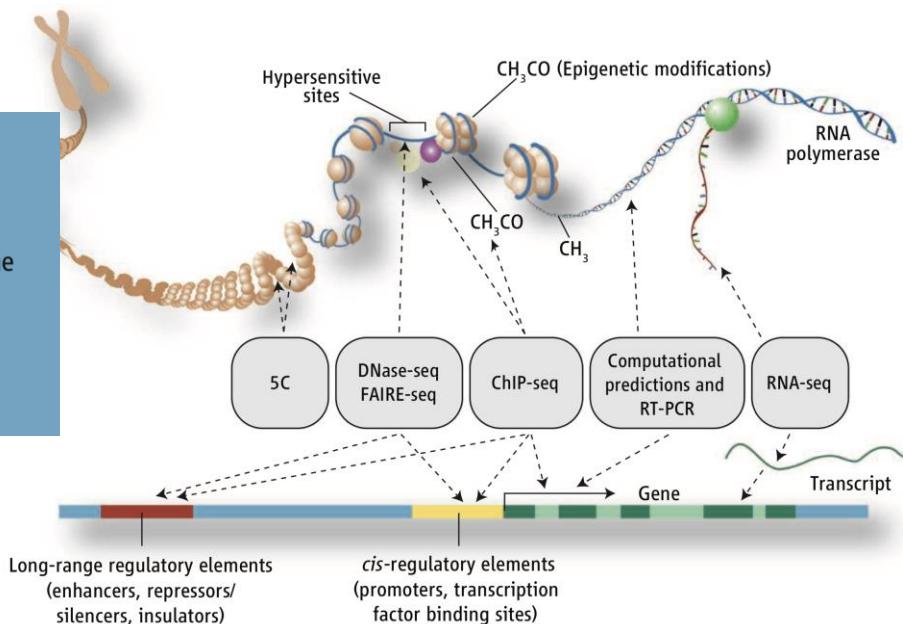
147 cell types studied

80% functional portion of human genome

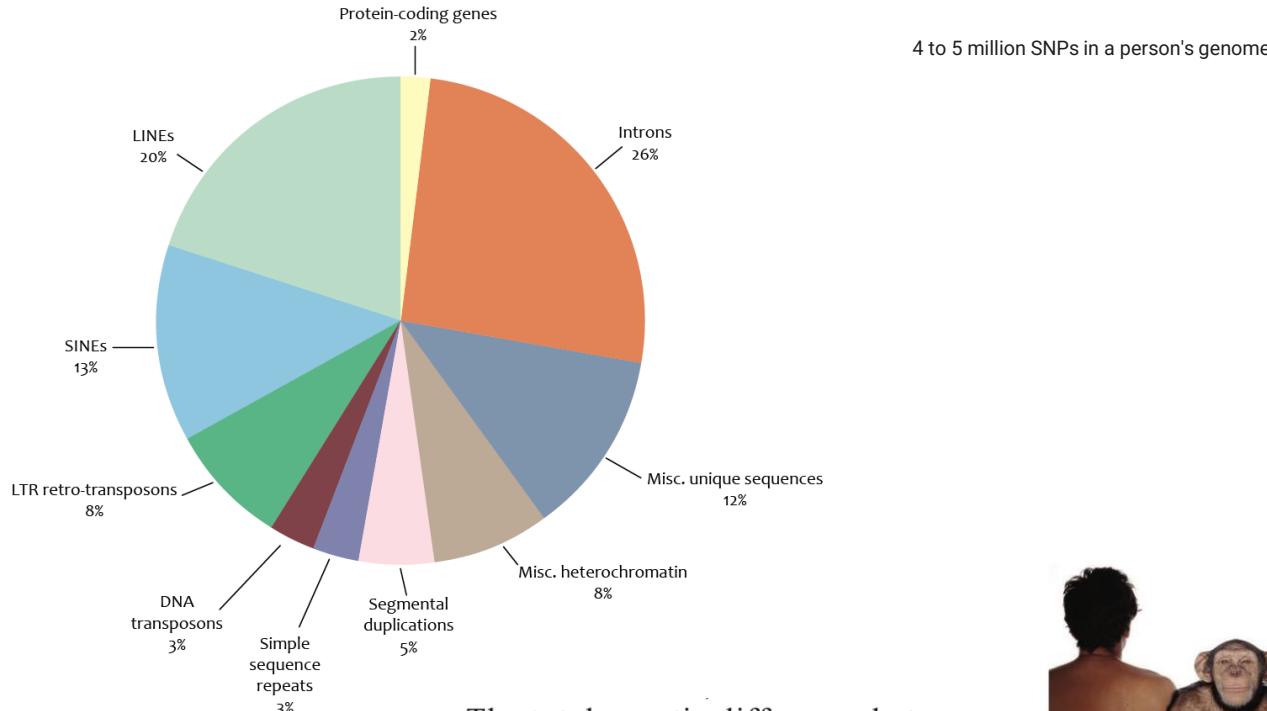
20,687 protein-coding genes

18,400 RNA genes

1640 data sets



# Структура генома человека и вариации



The total genetic difference between humans and chimps, in terms of number of bases, sums to about 4% of the genome. That



2 SEPTEMBER 2005 VOL 309 SCIENCE

99% identity of the aligned sequence  
96% identity between whole genomes

JOURNAL ARTICLE

# On the Immortality of Television Sets: “Function” in the Human Genome According to the Evolution-Free Gospel of ENCODE

Dan Graur , Yichen Zheng, Nicholas Price, Ricardo B.R. Azevedo, Rebecca A. Zufall,  
Eran Elhaik    Author Notes

*Genome Biology and Evolution*, Volume 5, Issue 3, March 2013, Pages 578–590,  
<https://doi.org/10.1093/gbe/evt028>

**Published:** 20 February 2013    **Article history** ▾

Whatever your proposed functions are, ask yourself this question: Why does an onion need a genome that is about five times larger than ours?”

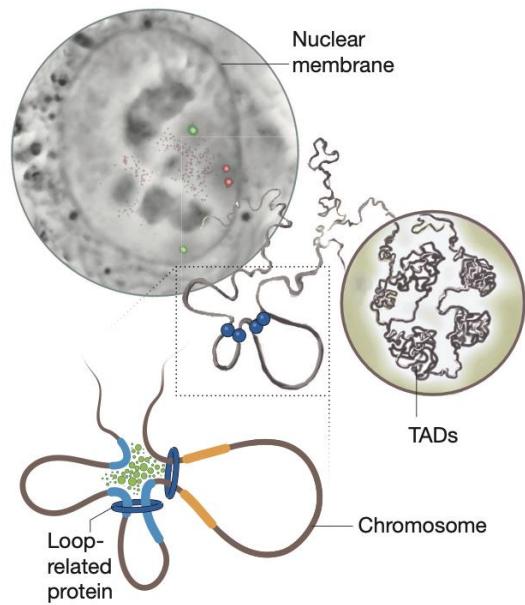
—T. Ryan Gregory (personal communication)

Actually, the ENCODE authors could have chosen any of a number of arbitrary percentages as “functional,” and... they did! In their scientific publications, ENCODE promoted the idea that 80% of the human genome was functional. The scientific commentators followed, and proclaimed that at least 80% of the genome is “active and needed” (Kolata 2012). Subsequently, one of the lead authors of ENCODE admitted that the press conference mislead people by claiming that 80% of our genome was “essential and useful.” He put that number at 40% (Gregory 2012), although another lead author reduced the fraction of the genome that is devoted to function to merely 20% (Hall 2012). Interestingly, even when a lead author of ENCODE reduced the functional genomic fraction to 20%, he continued to insist that the term “junk DNA” needs “to be totally expunged from the lexicon,” inventing a new arithmetic according to which  $20\% > 80\%$ . In its synopsis of the year 2012, the journal *Nature* adopted the more modest estimate, and summarized the findings of ENCODE by stating that “at least 20% of the genome can influence gene expression” (Van Noorden 2012). *Science* stuck to its maximalist guns, and its summary of 2012 repeated the claim that the “functional portion” of the human genome equals 80% (Anonymous 2012). Unfortunately, the ENCODE authors were not the only ones to make such claims.

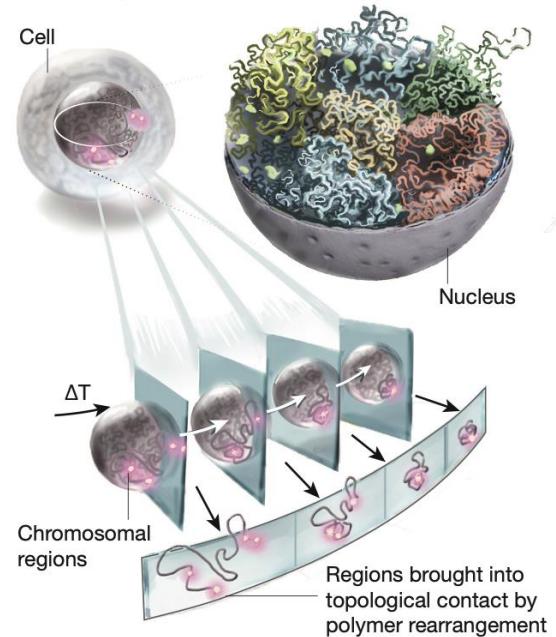
- A recent slew of ENCYclopedia Of DNA Elements (ENCODE) Consortium publications, article signed by all Consortium members, put forward the idea that more than 80% genome is functional. This claim flies in the face of current estimates according to of the genome that is evolutionarily conserved through purifying selection is less according to the ENCODE Consortium, a biological function can be maintained selection, which implies that at least  $80 - 10 = 70\%$  of the genome is perfectly deleterious mutations, either because no mutation can ever occur in these because no mutation in these regions can ever be deleterious. This absurd conclusion through various means, chiefly by employing the seldom used “causal role” function and then applying it inconsistently to different biochemical properties, by logical fallacy known as “affirming the consequent,” by failing to appreciate the between “junk DNA” and “garbage DNA,” by using analytical methods that yield biased errors and inflate estimates of functionality, by favoring statistical sensitivity over specificity, and by emphasizing statistical significance rather than the magnitude of the effect. Here, we detail the many logical and methodological transgressions involved in assigning functionality to almost every nucleotide in the human genome. The ENCODE results were predicted by one of its authors to necessitate the rewriting of textbooks. We agree, many textbooks dealing with marketing, mass-media hype, and public relations may well have to be rewritten.

- Недавнее множество публикаций Консорциума ENCyclopedia Of DNA Elements (ENCODE), особенно статья, подписанная всеми членами Консорциума, выдвинула идею о том, что более 80% генома человека является функциональным. Это заявление идет вразрез с современными оценками, согласно которым доля генома, эволюционно законсервированного посредством очищающего отбора, составляет менее 10%. Таким образом, согласно консорциуму ENCODE, биологическая функция может поддерживаться неопределенном долго время без селекции, что означает, что по крайней мере  $80 - 10 = 70\%$  генома совершенно неуязвимы для вредоносных мутаций либо потому, что никакая мутация никогда не может произойти в этих «функциональных» или потому, что никакая мутация в этих регионах никогда не может быть вредной. Этот абсурдный вывод был сделан различными способами, главным образом путем использования редко используемого определения «каузальной роли» биологической функции, а затем непоследовательного применения его к различным биохимическим свойствам, путем совершения логической ошибки, известной как «подтверждение следствия», из-за неспособности оценить ключевое различие между «junk DNA» и «garbage DNA», используя аналитические методы, которые дают предвзятые ошибки и завышают оценки функциональности, отдавая предпочтение статистической чувствительности, а не специфичности, и подчеркивая статистическую значимость, а не величину эффекта. Здесь мы подробно описываем множество логических и методологических ошибок, связанных с определением функциональности почти каждого нуклеотида в геноме человека. Один из авторов предсказал, что результаты ENCODE потребуют переписывания учебников. Мы согласны с тем, что многие учебники по маркетингу, шумихе в СМИ и связям с общественностью, возможно, придется переписать.

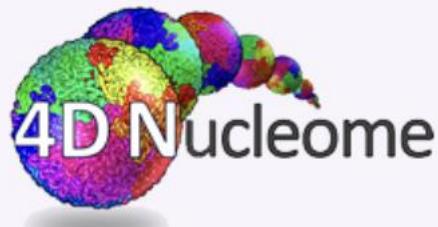
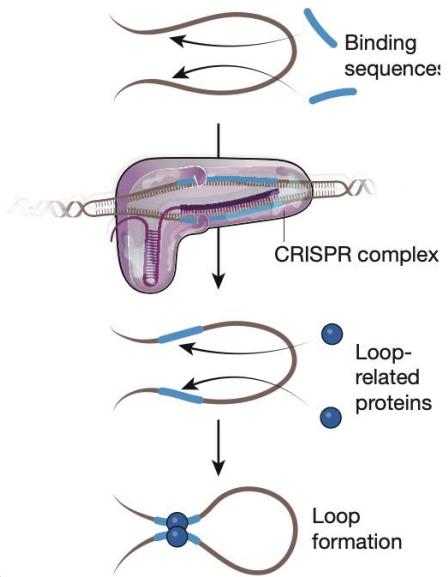
**a Mapping** – molecular genomic mapping of contacts, imaging of contacts and dynamics



**b Model building** – by constraint modelling and polymer simulation



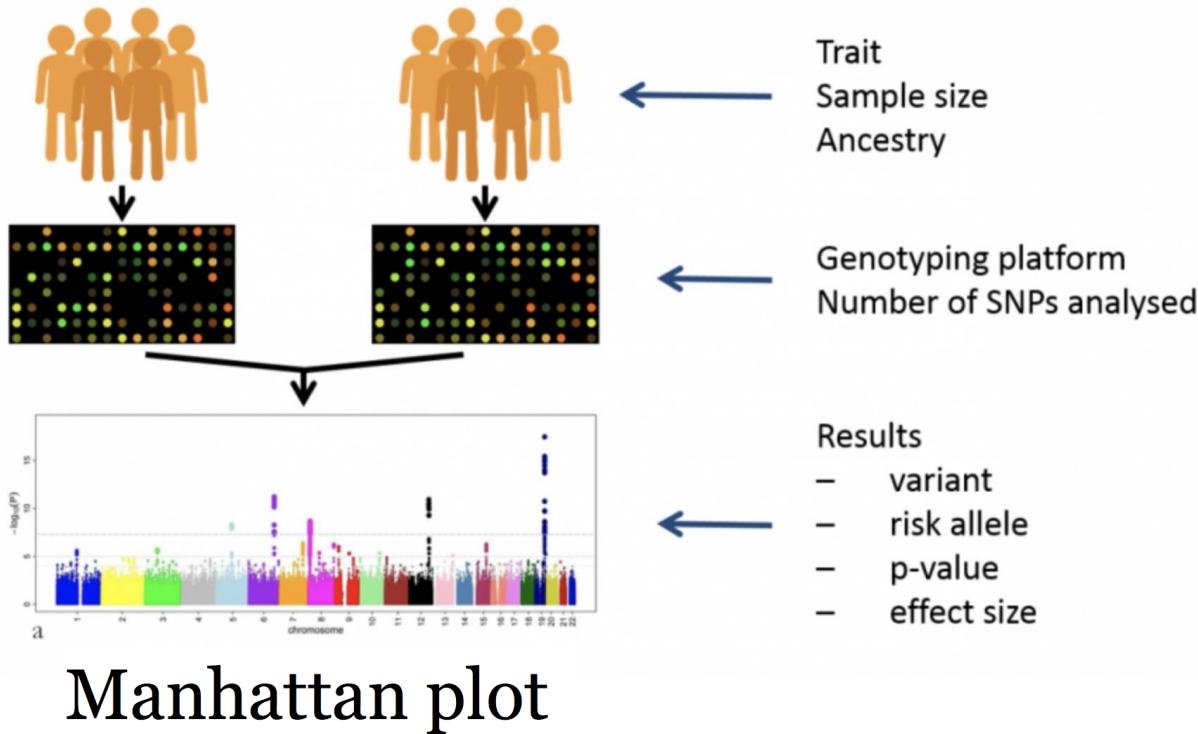
**c Functional validation** – by genetic and biophysical perturbation experiments



<https://www.4dnucleome.org>

# GWAS

## Genome-wide association study





# dpGAP

<https://www.ncbi.nlm.nih.gov/gap>

### CIDR: Collaborative Study on the Genetics of Alcoholism Case Control Study

dbGaP Study Accession: phs000125.v1.p1

[Request Access](#)

[Show BioProject list](#)

[Study](#) [Variables](#) [Documents](#) [Analyses](#) [Datasets](#) [Molecular Data](#)

**Analysis Name and Accession**

Name: GWAS for alcohol dependence in European-Americans

Accession: pha002892.1

[View association results in Genome Browser](#)

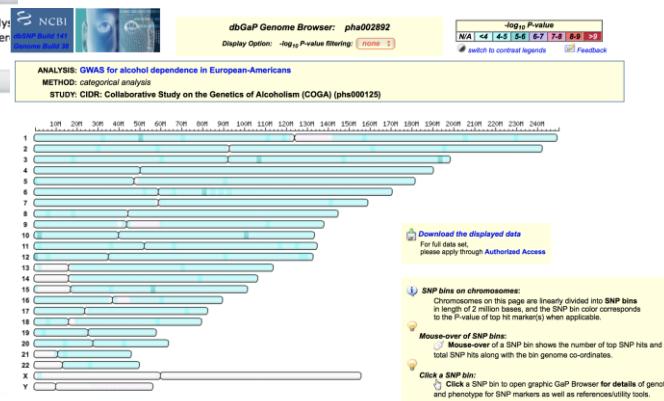
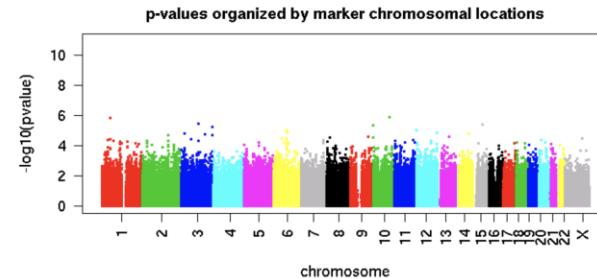
**Analysis Description**

To identify common variants underlying alcohol-dependence, probands were ascertained through alcohol treatment programs and evali and their relatives were administered a validated poly-diagnostic instrument, the Semi- Structured Assessment for the Genetics of Alcc performed by the Center for Inherited Disease Research ([CIDR](#)). DNA sources included blood (n = 1453) and lymphoblastoid cell lines

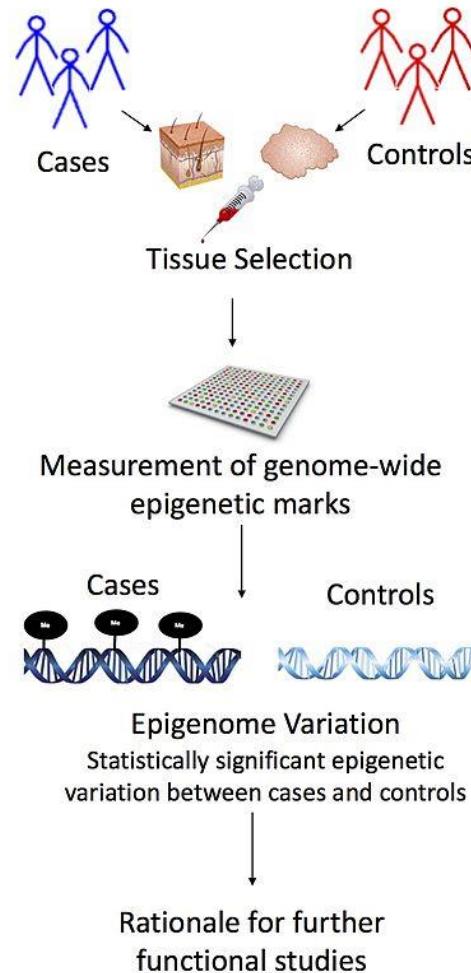
**Analysis Methods**

Sample QC filters consisted of 98% genotyping completeness, unrelatedness of subjects (n = 9). A principal component-based analysis either the European or African-American groups and were excluded. EA and AA SNP QC filters (applied separately in each group) were performed in PLINK adjusting for sex.

**Analysis Plots**

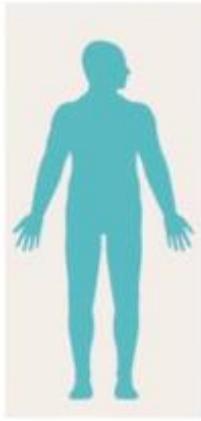


## Epigenome-wide association study (EWAS)

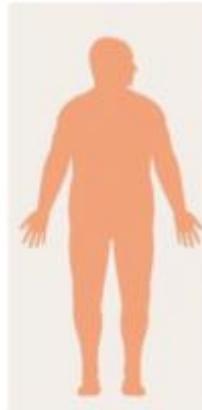


# Meta-EWAS of incident type 2 diabetes

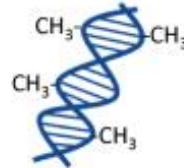
7–10 years prior



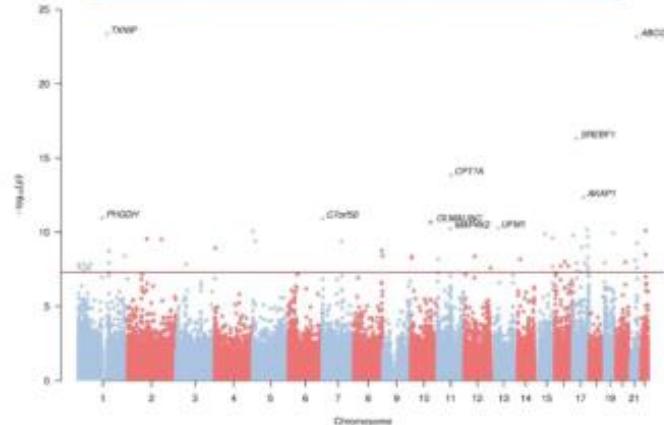
Type 2 diabetes

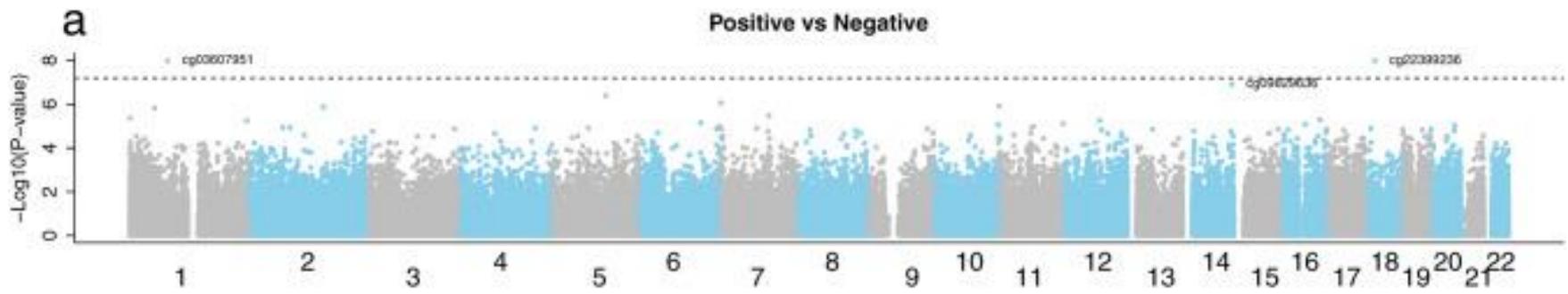


Blood



DNA methylation





*Sci Rep.* 2022; 12: 11478.

PMCID: PMC9261254

Published online 2022 Jul 7. doi: [10.1038/s41598-022-15467-1](https://doi.org/10.1038/s41598-022-15467-1)

PMID: 35798818

EWAS of post-COVID-19 patients shows methylation differences in the immune-response associated gene, *IFI44L*, three months after COVID-19 infection

[Yunsung Lee](#),<sup>#1</sup> [Espen Riskedal](#),<sup>#2</sup> [Karl Trygve Kalleberg](#),<sup>2</sup> [Mette Istrø](#),<sup>3</sup> [Andreas Lind](#),<sup>4</sup> [Fridtjof Lund-Johansen](#),<sup>5</sup> [Olaug Reiakvam](#),<sup>3</sup> [Arne V. L. Søraas](#),<sup>3</sup> [Jennifer R. Harris](#),<sup>1</sup> [John Arne Dahl](#),<sup>3</sup> [Cathrine L. Hadley](#),<sup>1,2</sup> and [Astanand Jugessur](#)<sup>1,6</sup>

Спасибо за внимание!