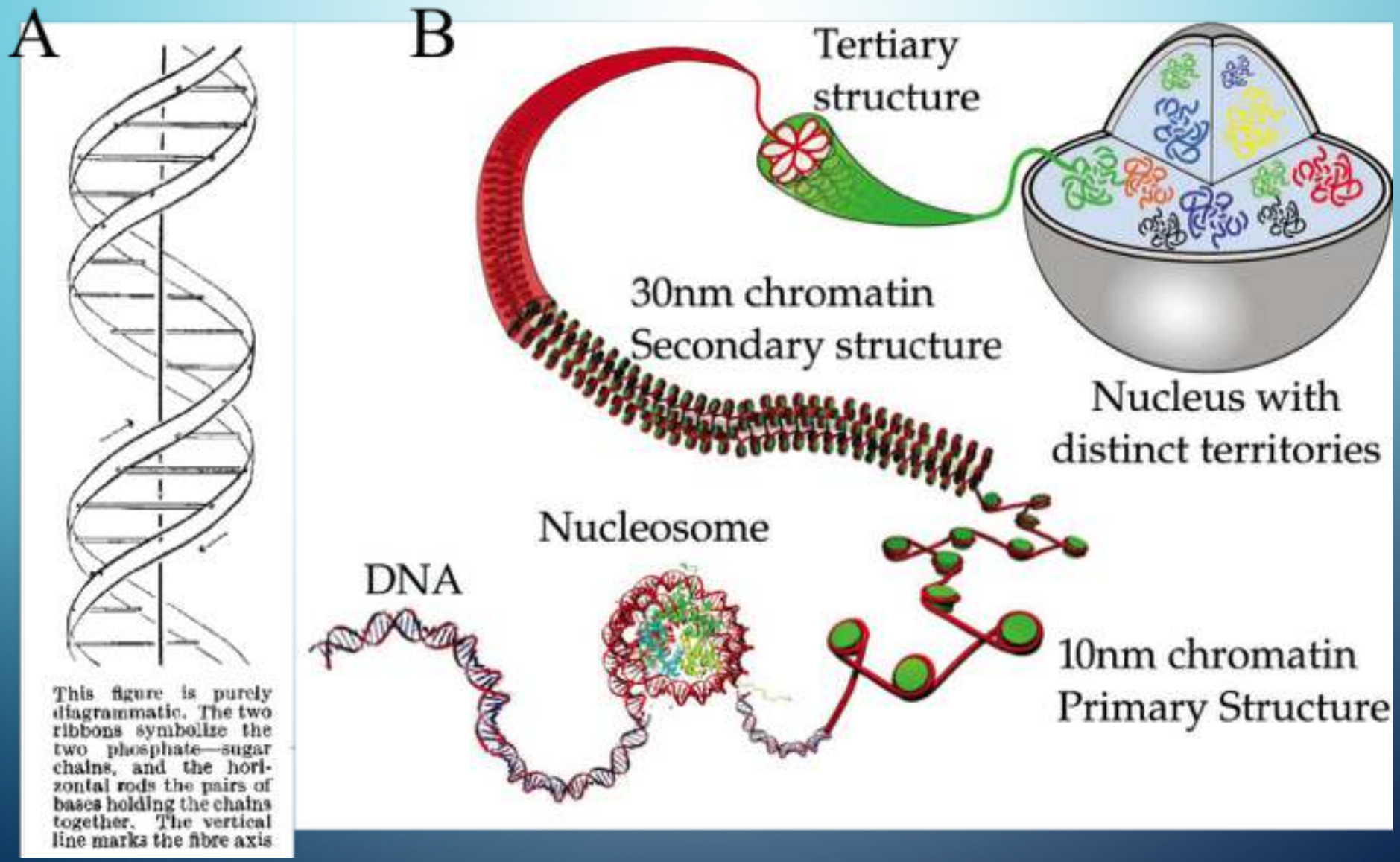
Технологии Секвенирования

Павел Добрынин

Что такое геном?

Структура ДНК и генома



Как его можно изучать?

Friedrich Miescher publishes his paper identifying the presence of ‘nuclein’ (now known as DNA) and associated proteins, in the cell nucleus.

1871

Как его можно изучать?

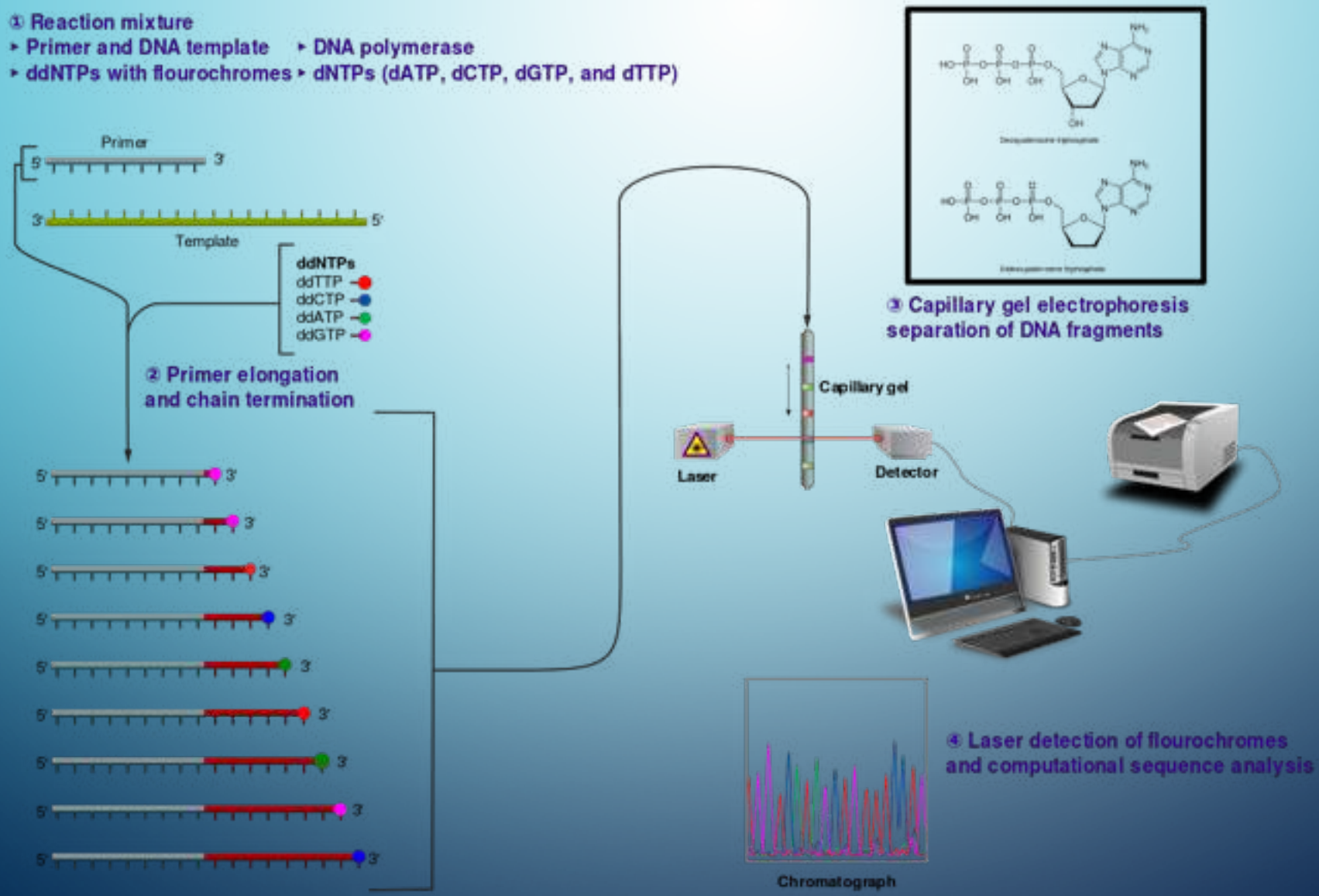
Friedrich Miescher publishes his paper identifying the presence of ‘nuclein’ (now known as DNA) and associated proteins, in the cell nucleus.

1871

Frederick Sanger develops a DNA sequencing technique which he and his team use to sequence the first full genome – that of a virus called phiX174.

1977

Секвенирование по Sanger



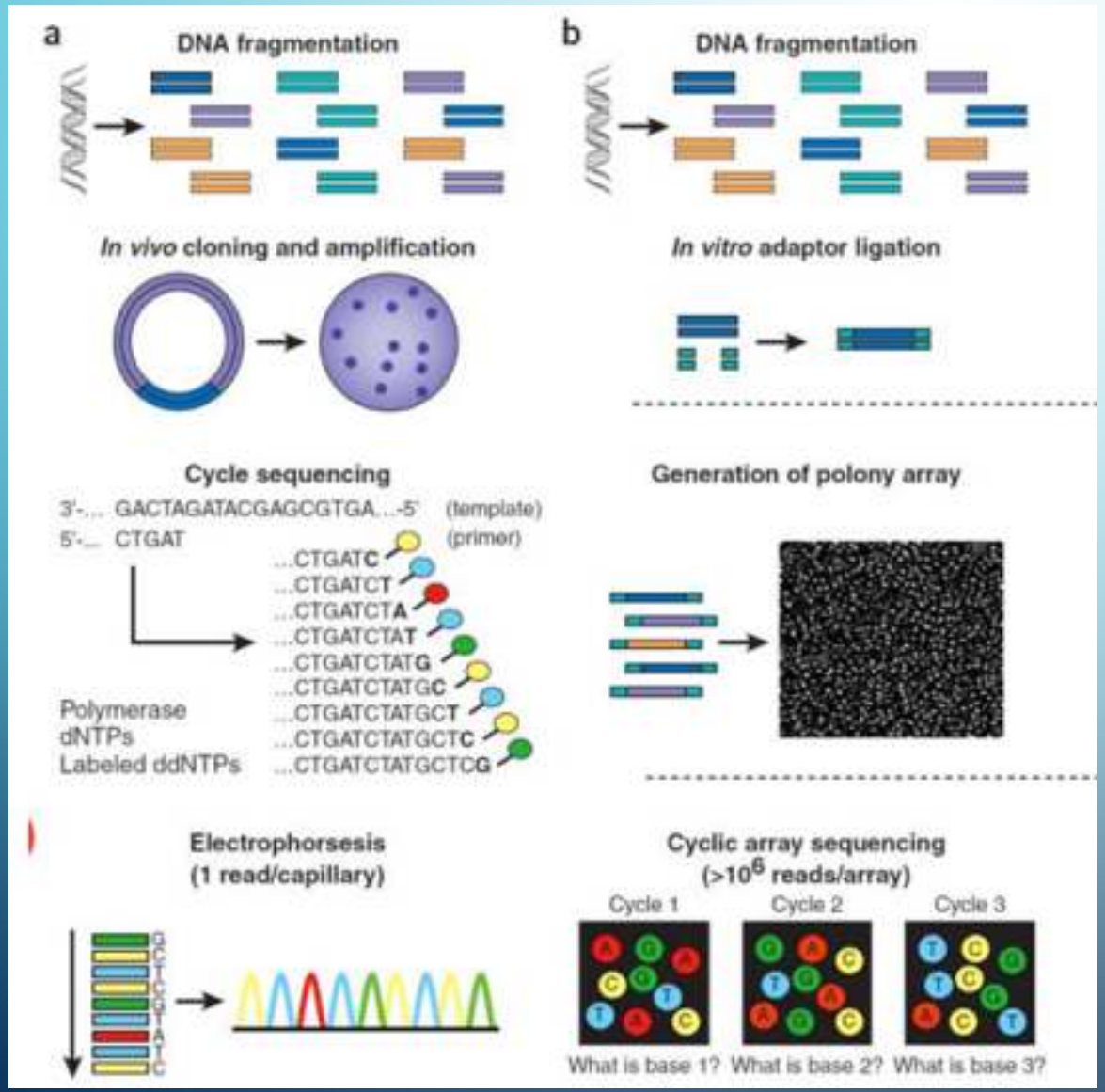
NGS Synonym is : **High-throughput Sequencing (HTS)**

First Generation :
SANGER Sequencing

Third Generation :
NGS = HTS, Single Molecule Sequencing

Second Generation :
NGS = Massively Parallel Sequencing

NGS vs Sanger

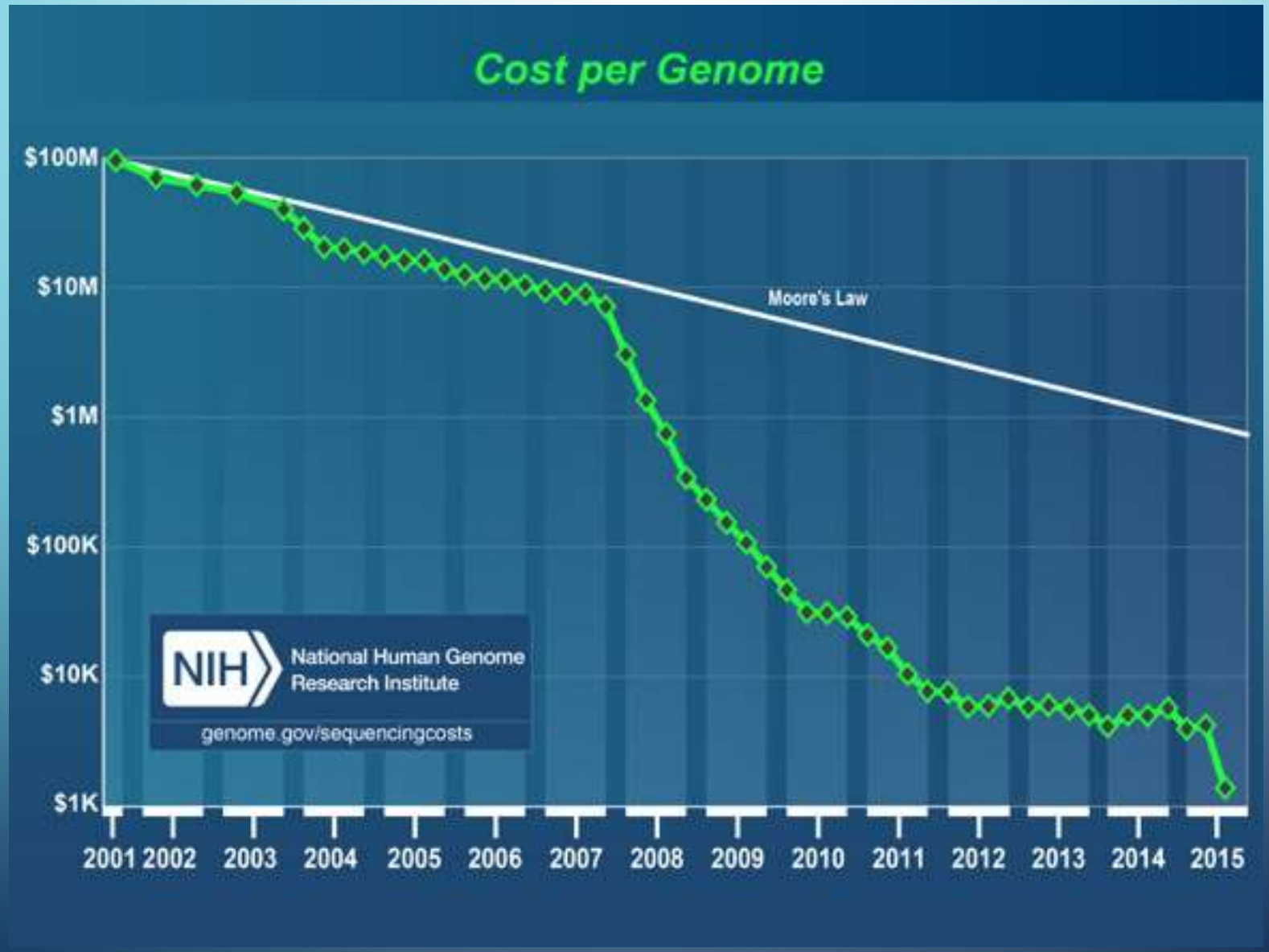


Library preparation

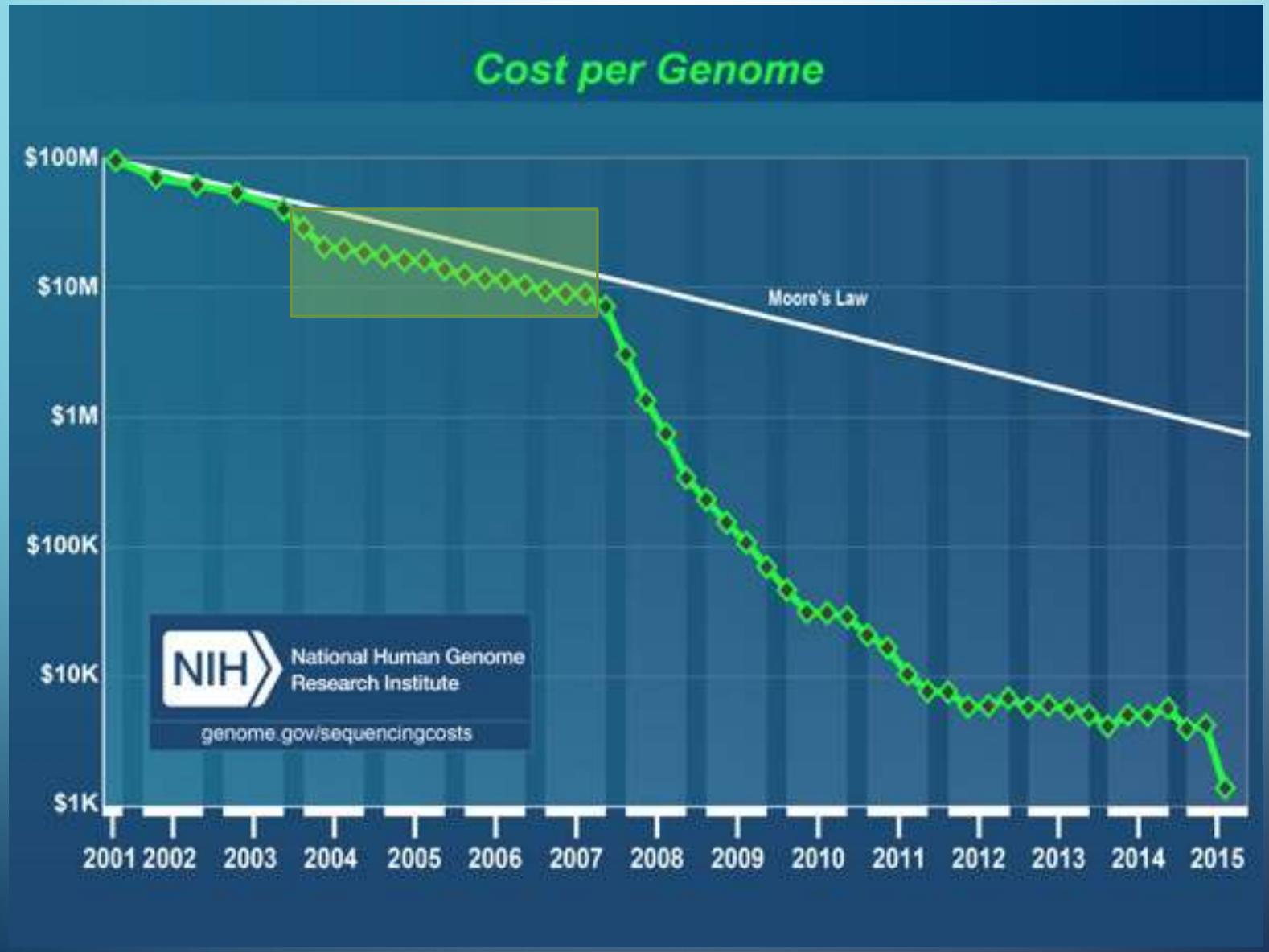
Template preparation

Sequencing

Динамика цены за один геном

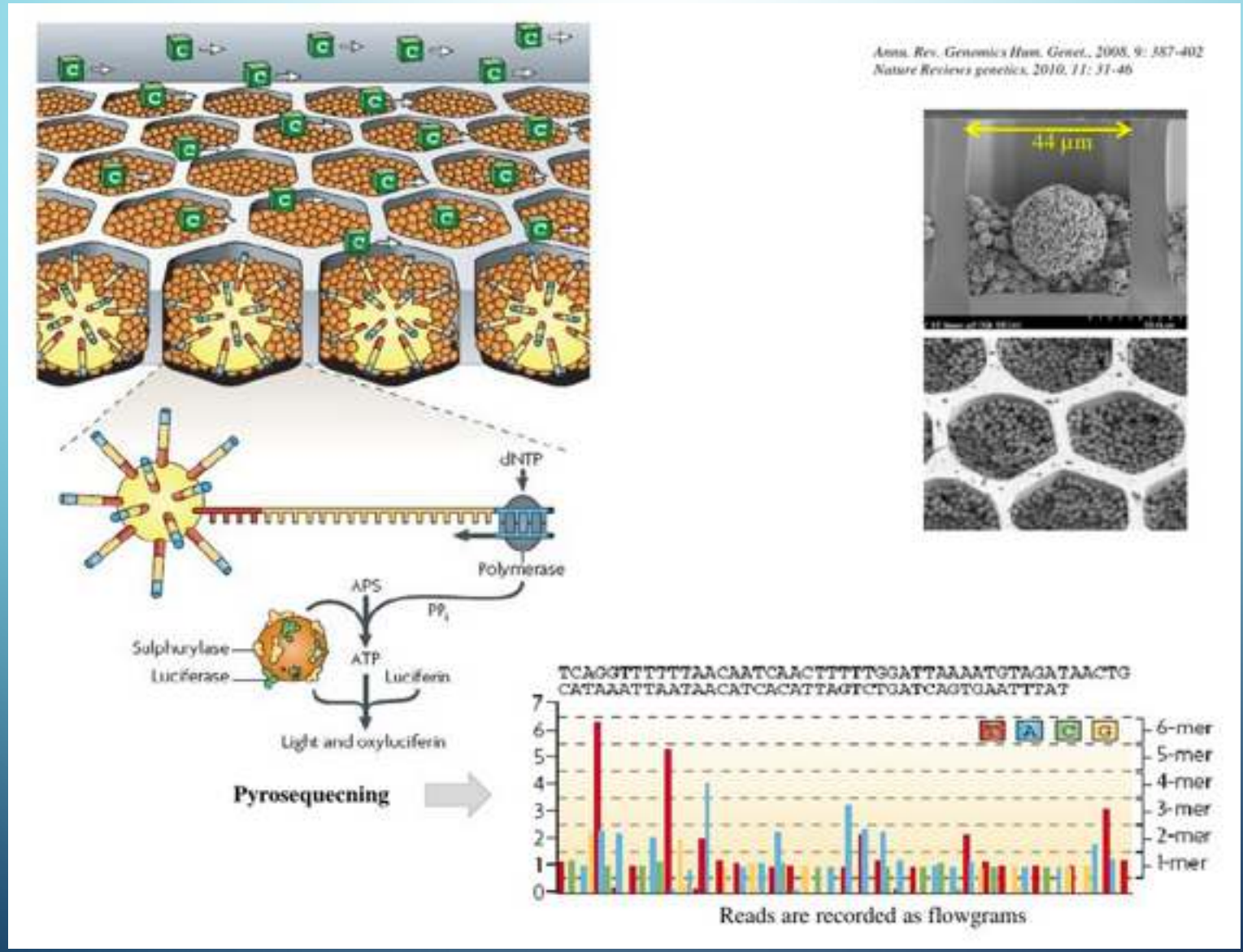


Динамика цены за один геном



Pyrosequencing (454)

Pyrosequencing (454)



454 GS FLX titanium

No more Cloning step
From purified DNA to Sequencing
Fit the laboratory bench top / small
LONG Sequences (400 nt)
GS Junior system not so expensive

Capabilities : Multiplexing & paired-ends

Well fitted to :
- proK. Genome sequencing
- RNA-seq

- Seq. Accuracy not so high (especially in case of homopolymers
→ Main error type is indel

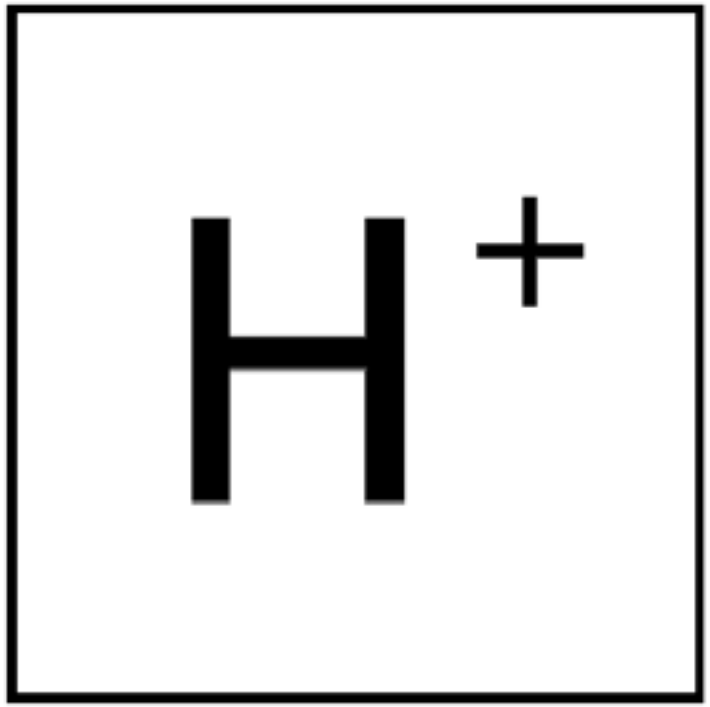
- Cost : approx. 20K€ / Gb
Cost per base is cheaper (regarding Sanger) but still High regarding others NexGen Machines

Ion Torrent

Simplicity

Simple Post-Light™ Sequencing System

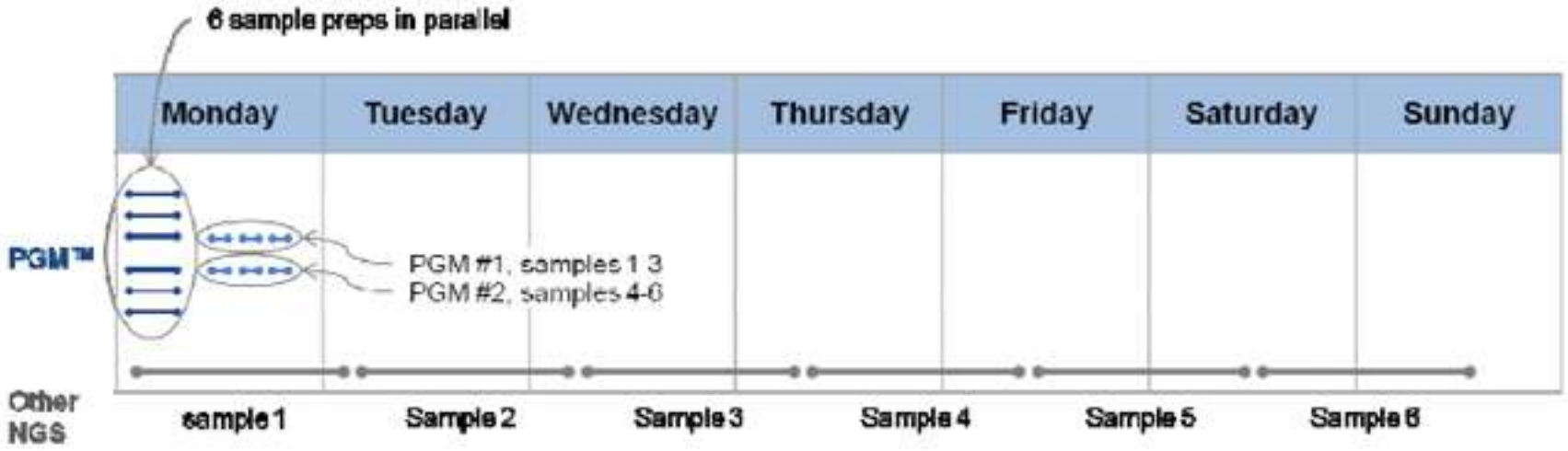
- Highly uniform genome coverage
- Rapidly improving per base accuracy
  Measured, not predicted and includes homopolymers
- Low-cost reagents and detection
- Quarterly improvements simplifying template prep for Post-Light™ sequencing



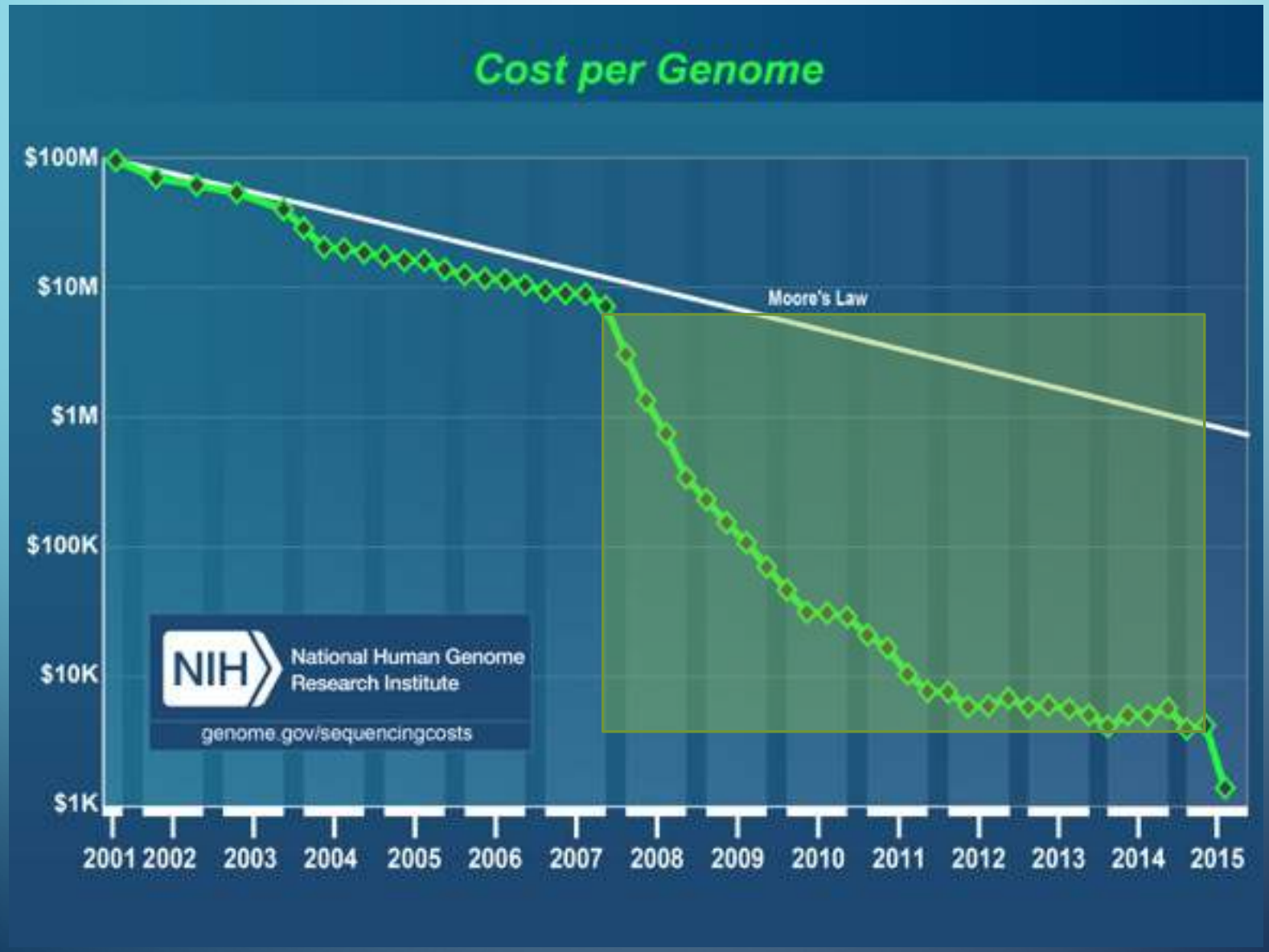
Speed

10X Faster Workflow for Any Project

- ~2 hour sequencing runs
- Batch sample preparation
  6 samples in 6 hours
- Capable of 6 samples/day on 2 PGM Systems

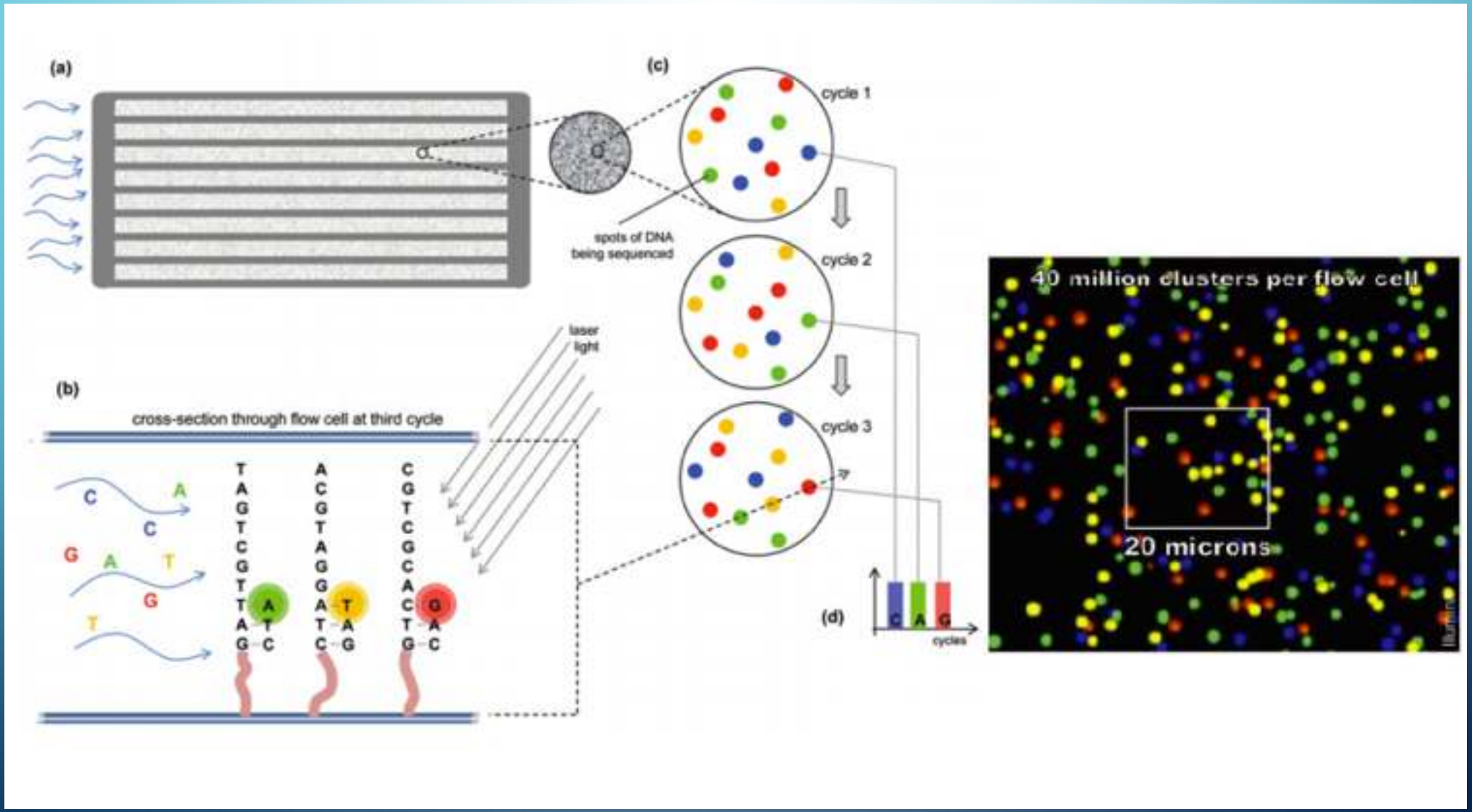


Динамика цены за один геном

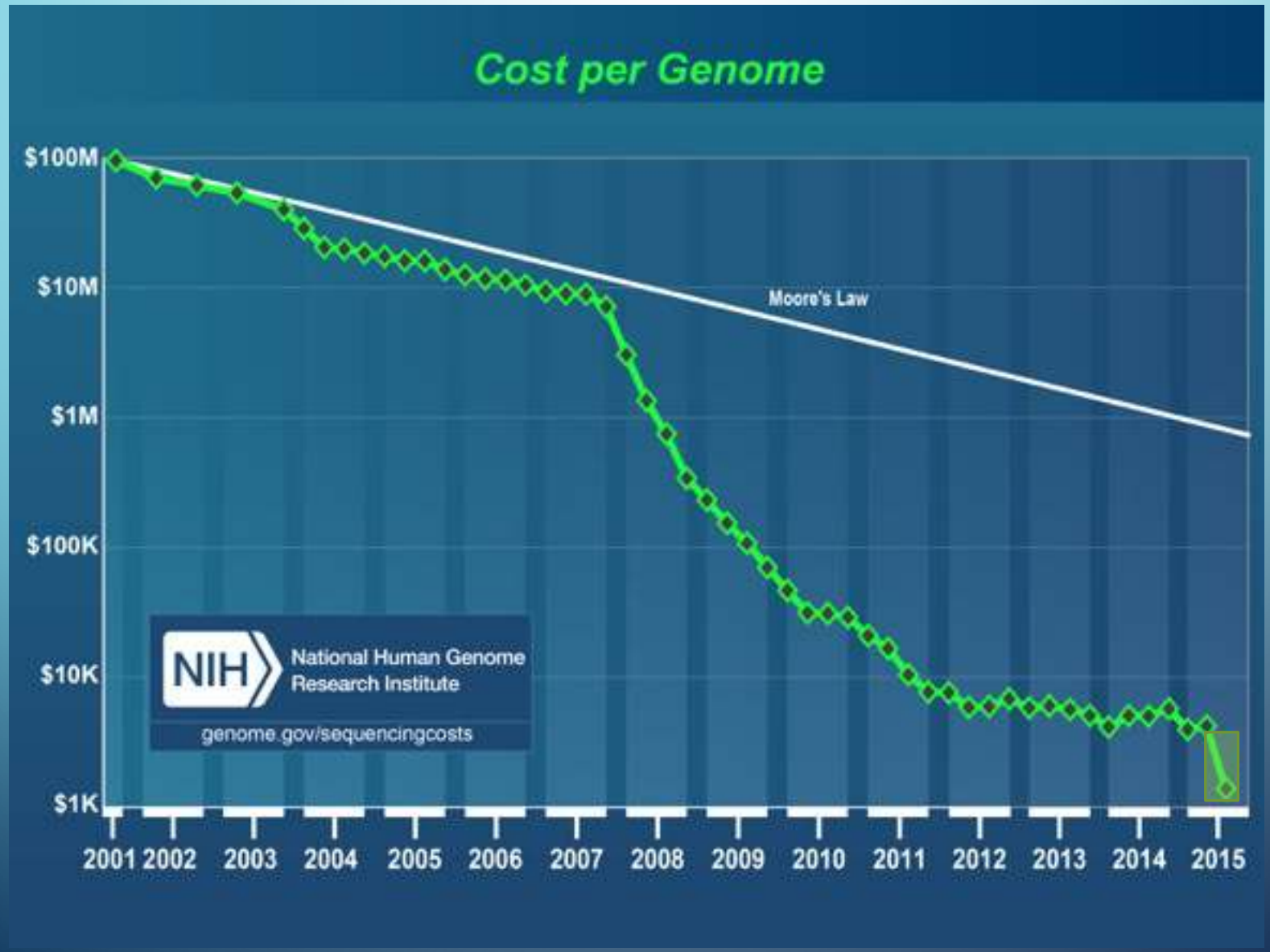


Illumina

Illumina


(a)
(b)
cross-section through flow cell at third cycle
laser light
(c)
cycle 1
cycle 2
cycle 3
spots of DNA being sequenced
(d)
cycles
40 million clusters per flow cell
20 microns

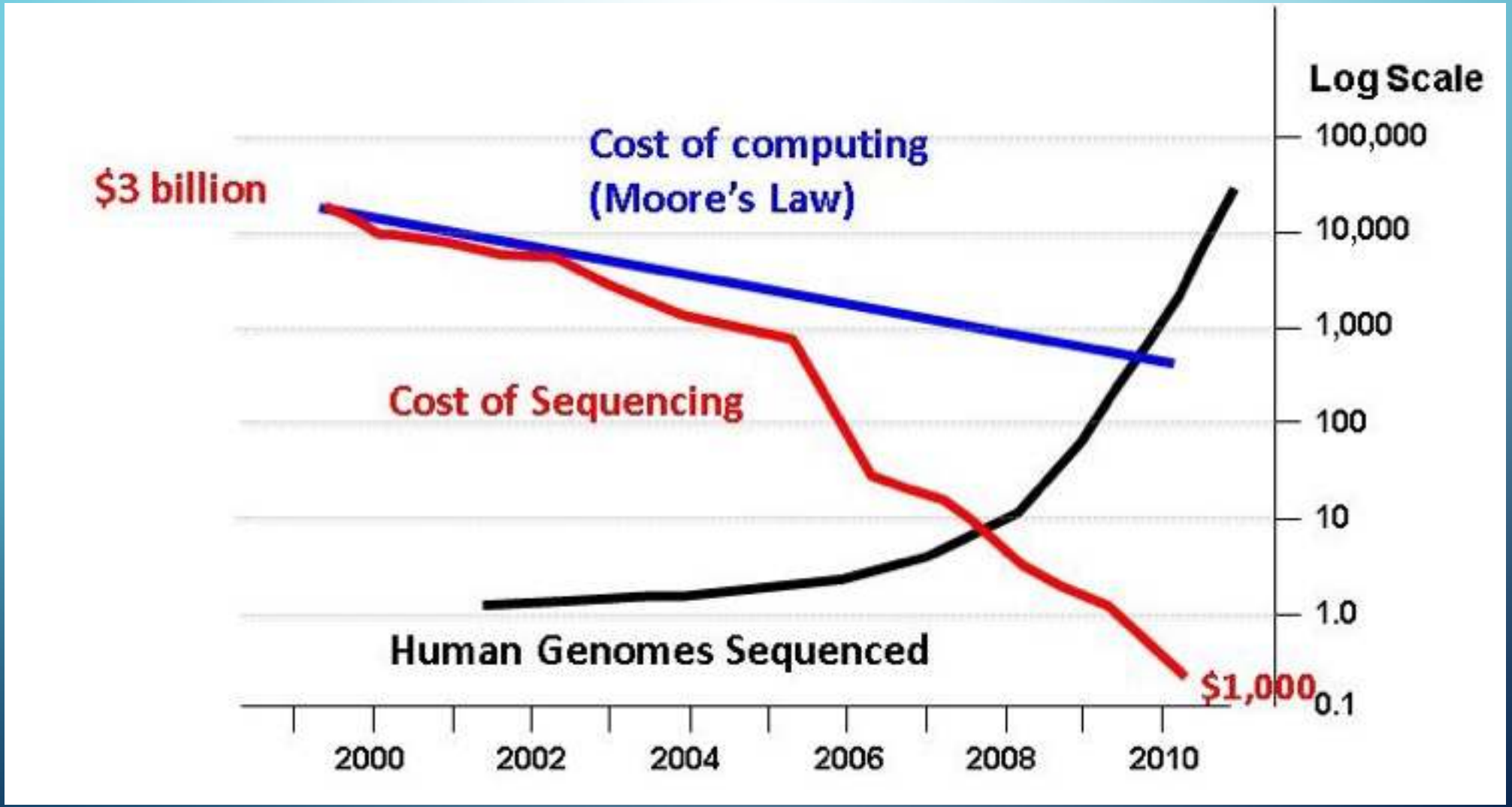
Динамика цены за один геном



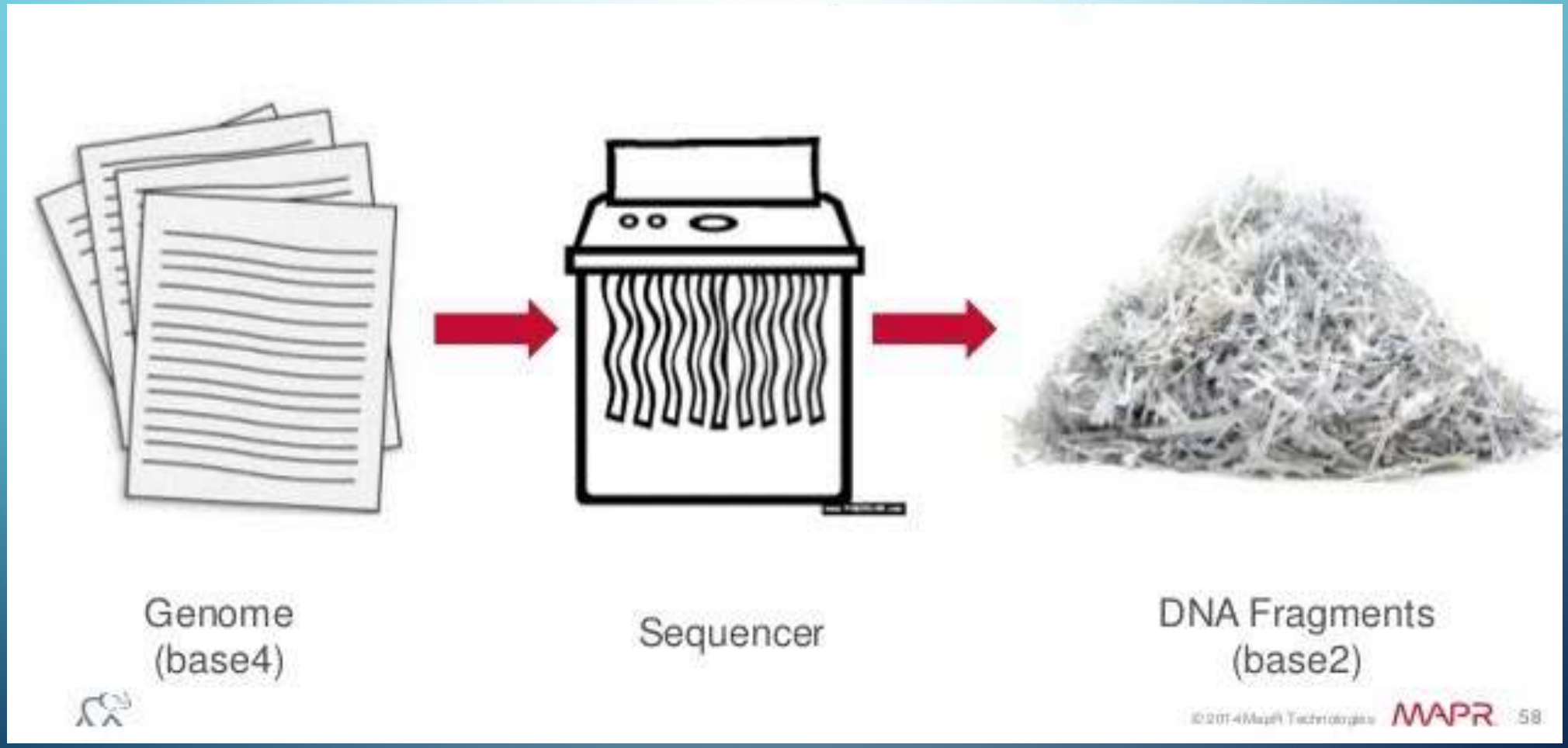
PLATFORM PORTFOLIO | Sequencing Power For Every Scale

	MiSeq	NextSeq 500	HiSeq 2500	HiSeq X Ten
Output:	15 Gb	120 Gb	1,000 Gb	1,800 Gb
Price per Gb:	$93	$33	1T: $29 v3: $43	$7

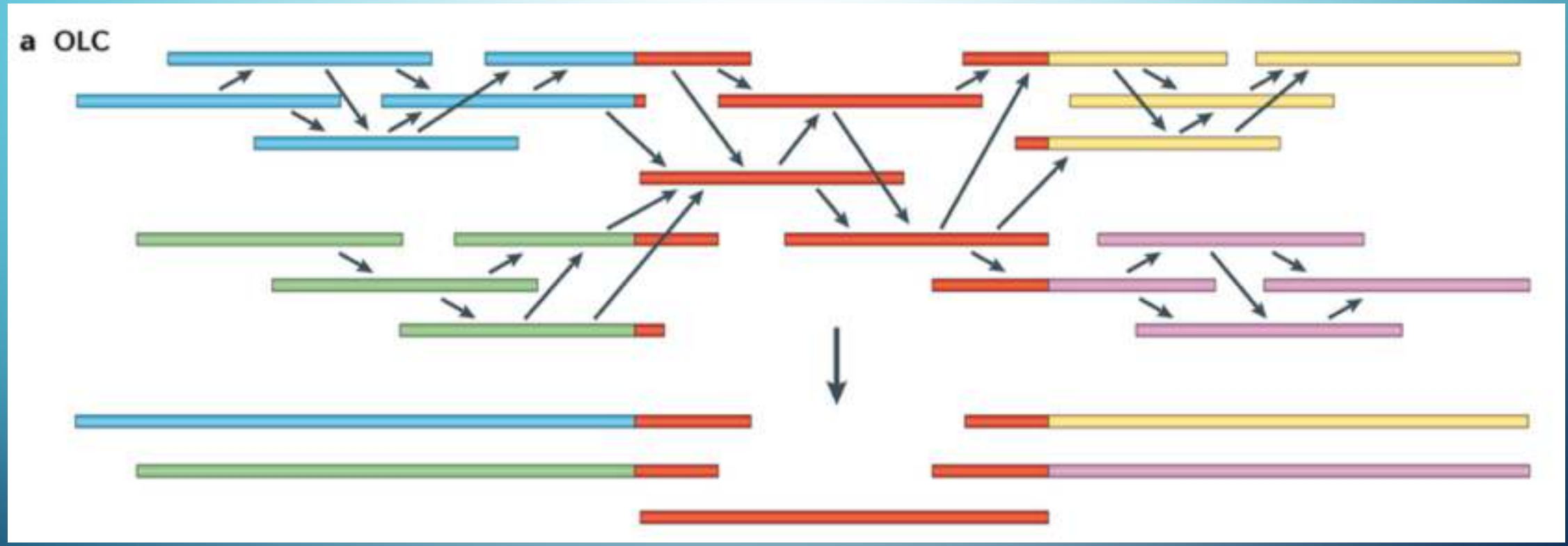
Динамика цены за один геном



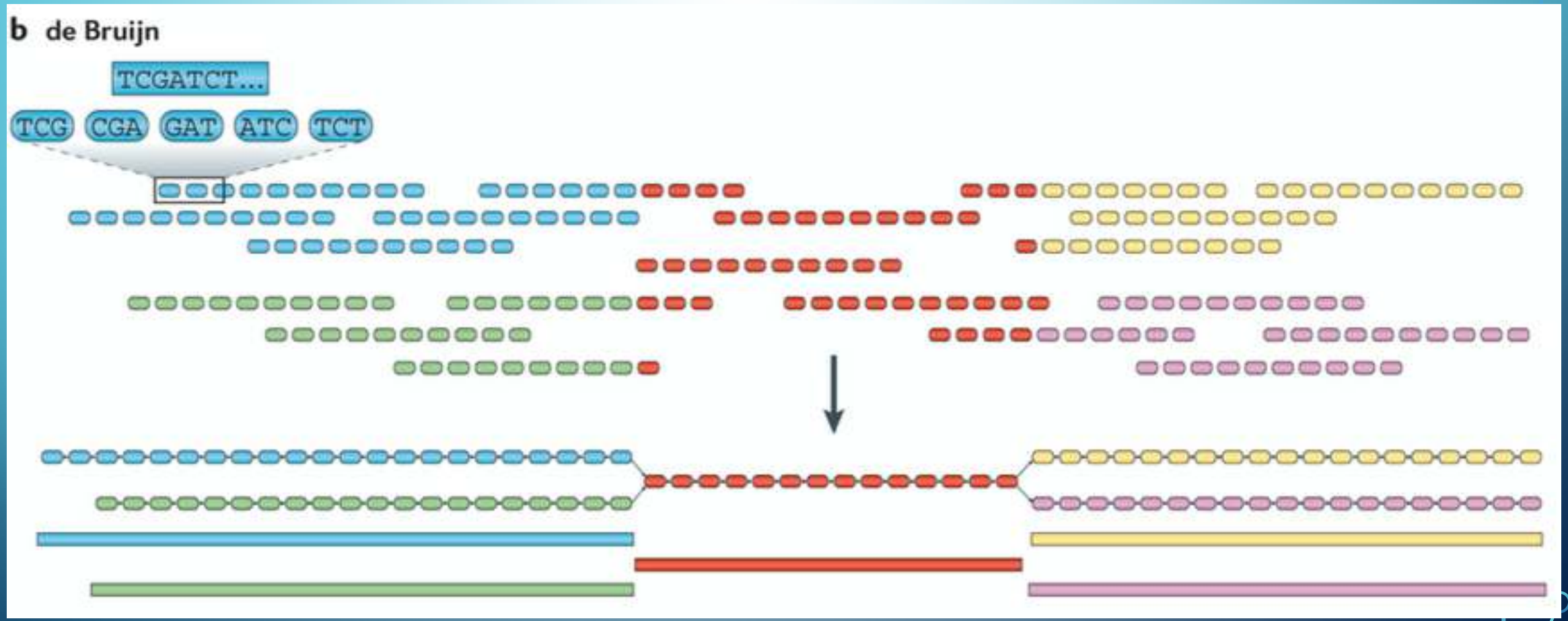
Шаг 1: Фрагментация генома



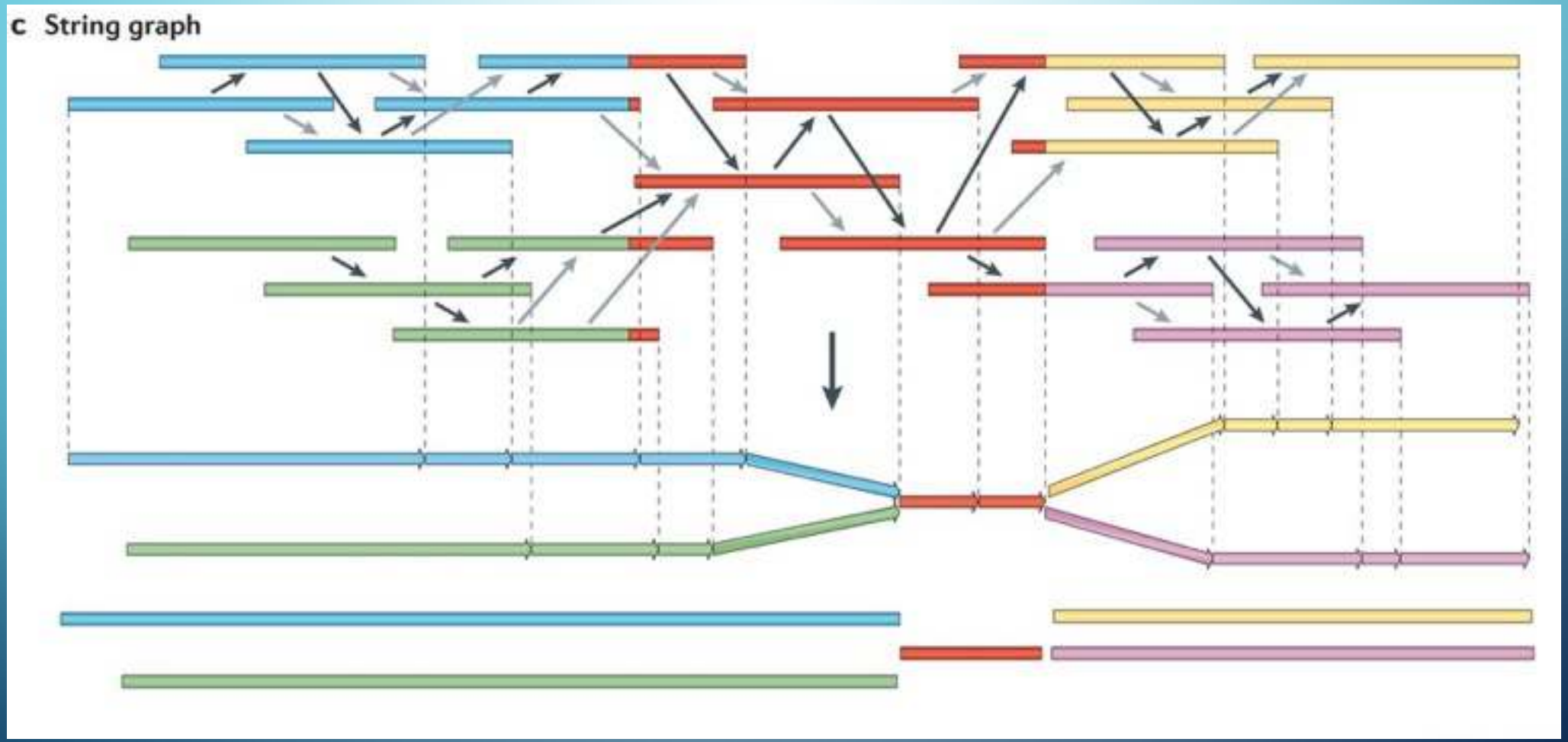
Шаг 2: Сборка генома



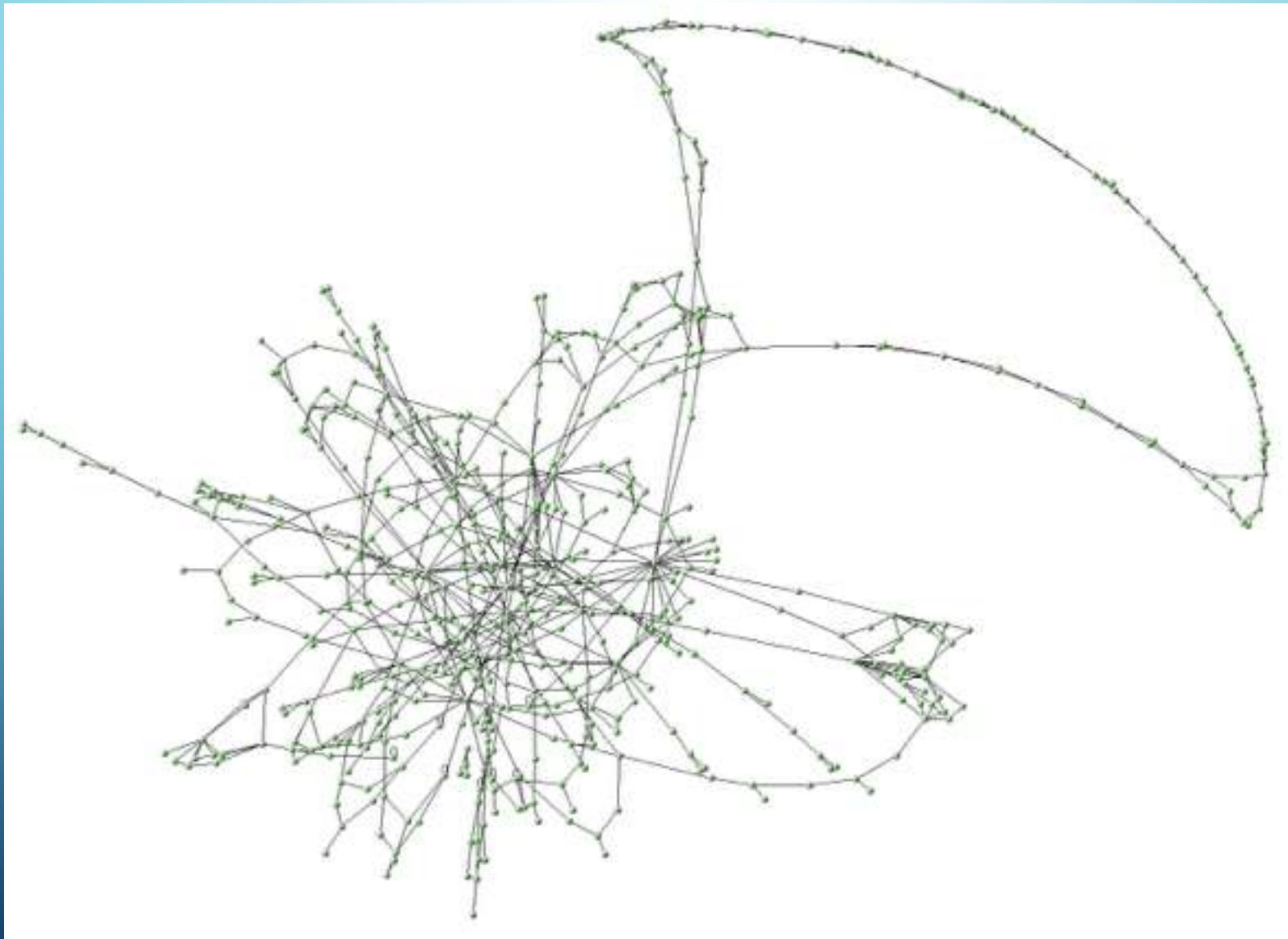
Шаг 2: Сборка генома



Шаг 2: Сборка генома

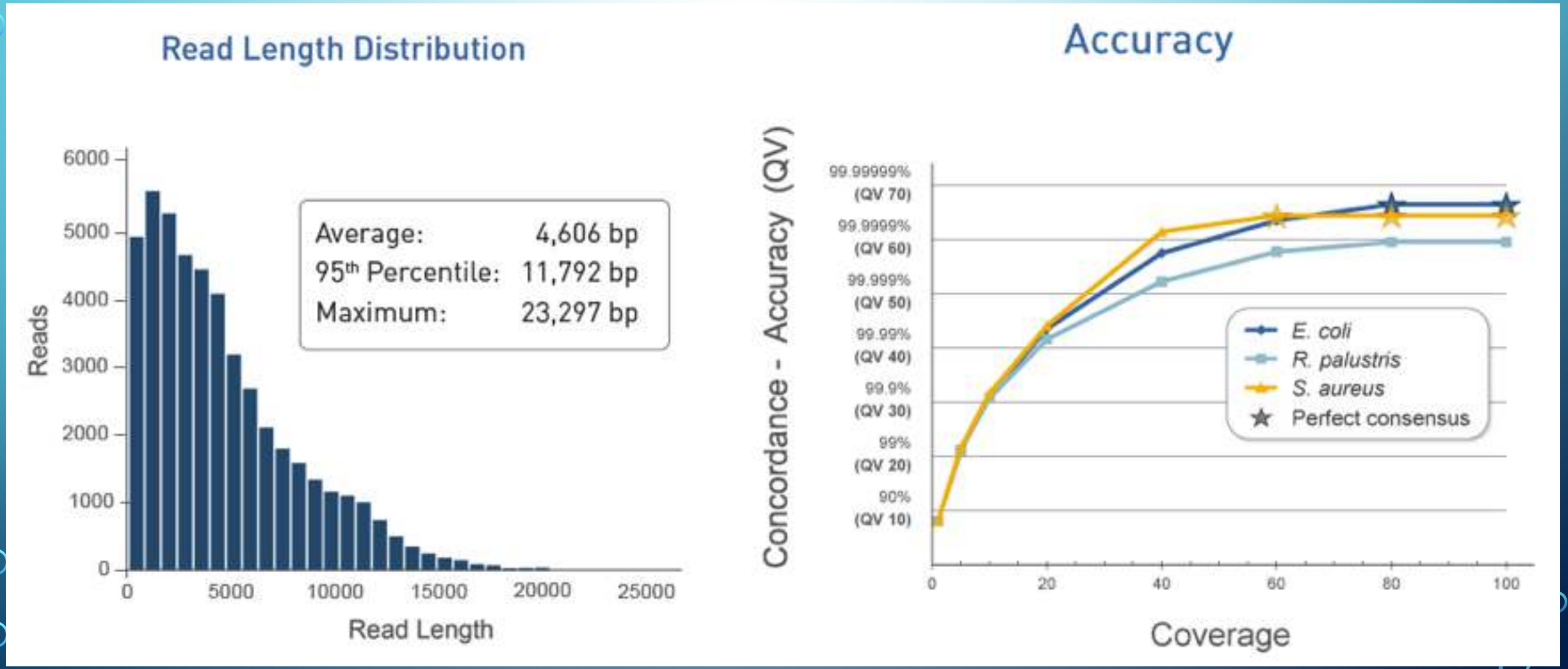


Шаг 2: Сборка генома (Реальнос



Новое Новое Поколение

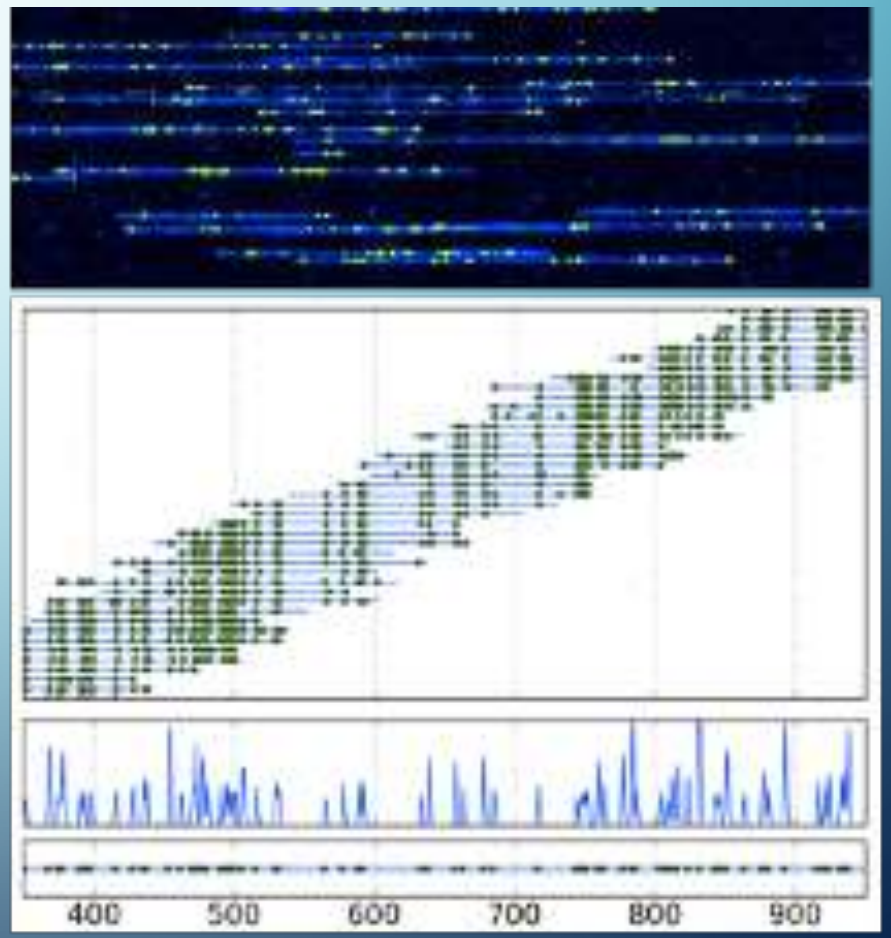
Качество


Read Length Distribution
Reads
6000
5000
4000
3000
2000
1000
0
0
5000
10000
15000
20000
25000
Read Length
Average: 4,606 bp
95th Percentile: 11,792 bp
Maximum: 23,297 bp
Accuracy
Concordance - Accuracy (QV)
99.99999% (QV 70)
99.9999% (QV 60)
99.999% (QV 50)
99.99% (QV 40)
99.9% (QV 30)
99% (QV 20)
90% (QV 10)
0
20
40
60
80
100
Coverage
E. coli
R. palustris
S. aureus
Perfect consensus

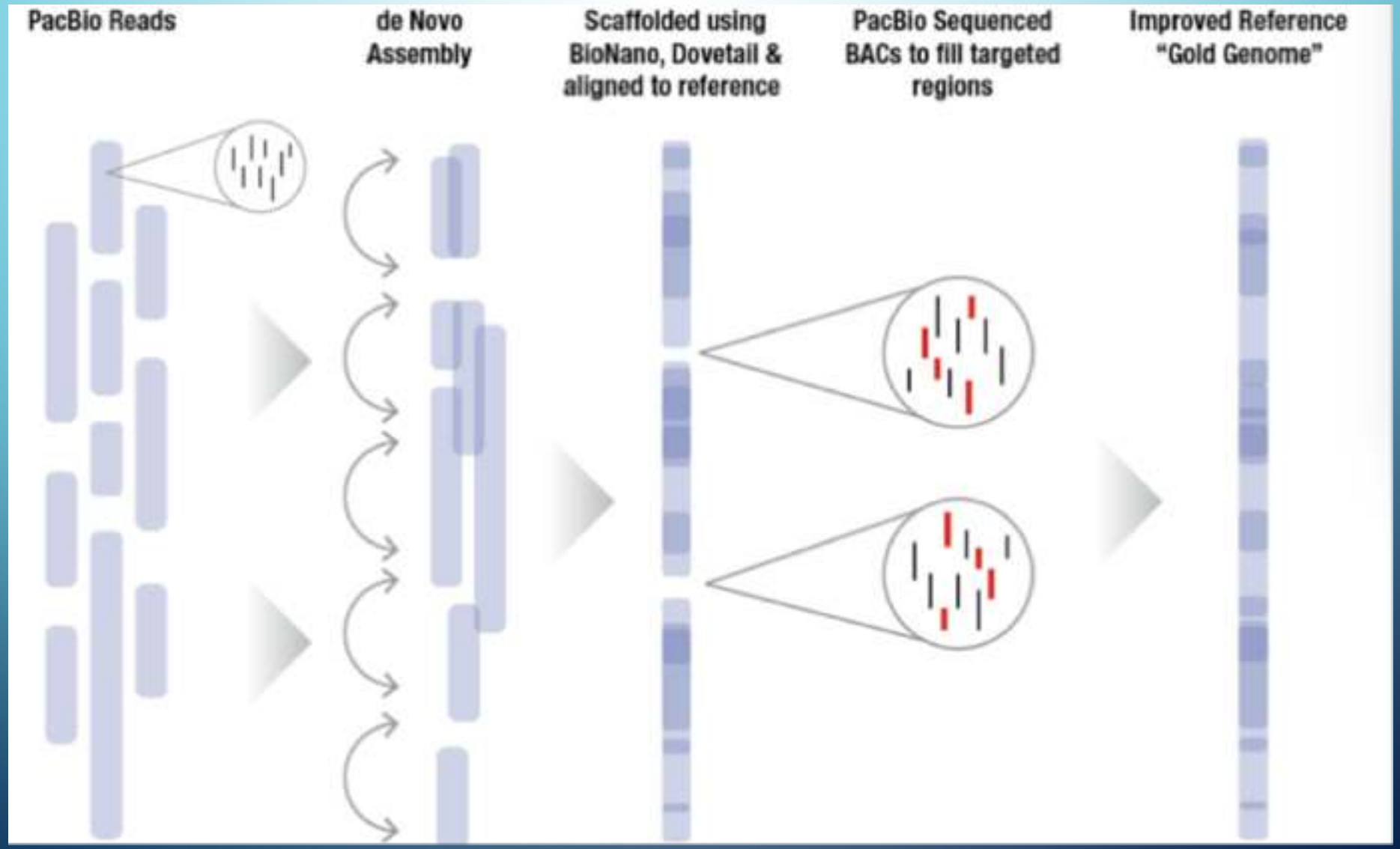
Organism	Assembled with	Assembly bp	Contigs	N50	LAP	Discordant Bases	QV
E. coli K12	MiSeq 100X 2x150bp 300bp (MaSuRCA iCORN)	4682345	139	113852	9.68E + 07	28	52.23
	454 50X	4569757	93	117490	−9.73E + 07	17	54.29
	PBcR 200X	**4653486**	**1**	**4653486**	**−9.64E + 07**	**3**	**>60**
E .coli O157:H7	MiSeq 100X 2x150bp 500bp (SPAdes iCORN)	5433737	413	133641	−3.67E + 07	62	49.43
	454 22X + 8X 5Kbp + 10X 10Kbp	5347050	409	133665	−3.73E + 07	66	49.09
	PBcR 200X	**5611389**	**9**	**4324437**	**−3.66E + 07**	**0**	**>60**
B. trehalosi	MiSeq 100X 2x150bp 500bp (SPAdes iCORN)	2377594	83	222446	−3.31E + 07	10	53.76
	454 50X	2364704	66	117742	−3.32E + 07	9	54.20
	PBcR 200X	**2411068**	**1**	**2411068**	**−3.27E + 07**	**0**	**>60**
M. haemolytica	MiSeq 100X 2x150bp 500bp (MaSuRCA iCORN)	2721965	89	84094	−3.33E + 07	47	47.63
	PBcR 200X	**2736037**	**1**	**2736037**	**−3.31E + 07**	**0**	**>60**
F. tularensis	MiSeq 100X 2x250bp 500bp (SPAdes iCORN)	1825374	130	24065	−1.33E + 07	0	>60
	454 50X	1655657	326	7316	−1.33E + 07	28	47.72
	PBcR 300X	**1877407**	**3**	**573021**	**−1.33E + 07**	**0**	**>60**
S. enterica Newport	MiSeq 56X 2x150bp 500bp (MaSuRCA iCORN)	5187269	114	195780	−2.24E + 07	360	41.59
	454 23X + 2X 10Kbp	5005089	172	372513	−2.25E + 07	39	51.08
	PBcR 200X	**5029197**	**2**	**4919684**	**−2.24E + 07**	**2**	**>60**

Table 1. Accuracy comparison for different sequencing technologies (last column, PacBio® assembly accuracy in bold, and compared to Illumina® and 454® data, respectively). From Koren et al. (2013) *Genome Biology* 14:R101[5].

BioNano Genomics



BioNano Genomics

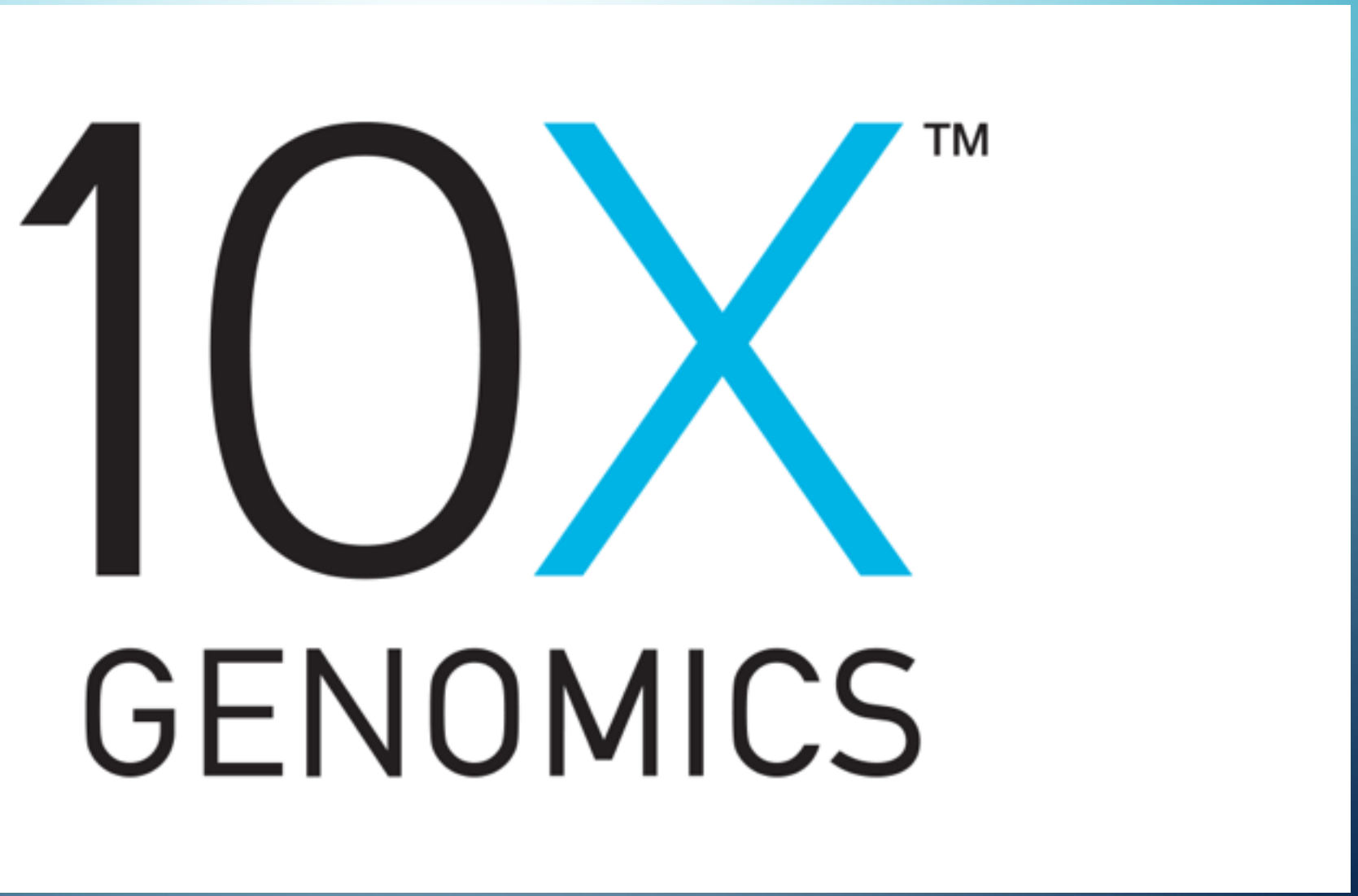


Dovetail

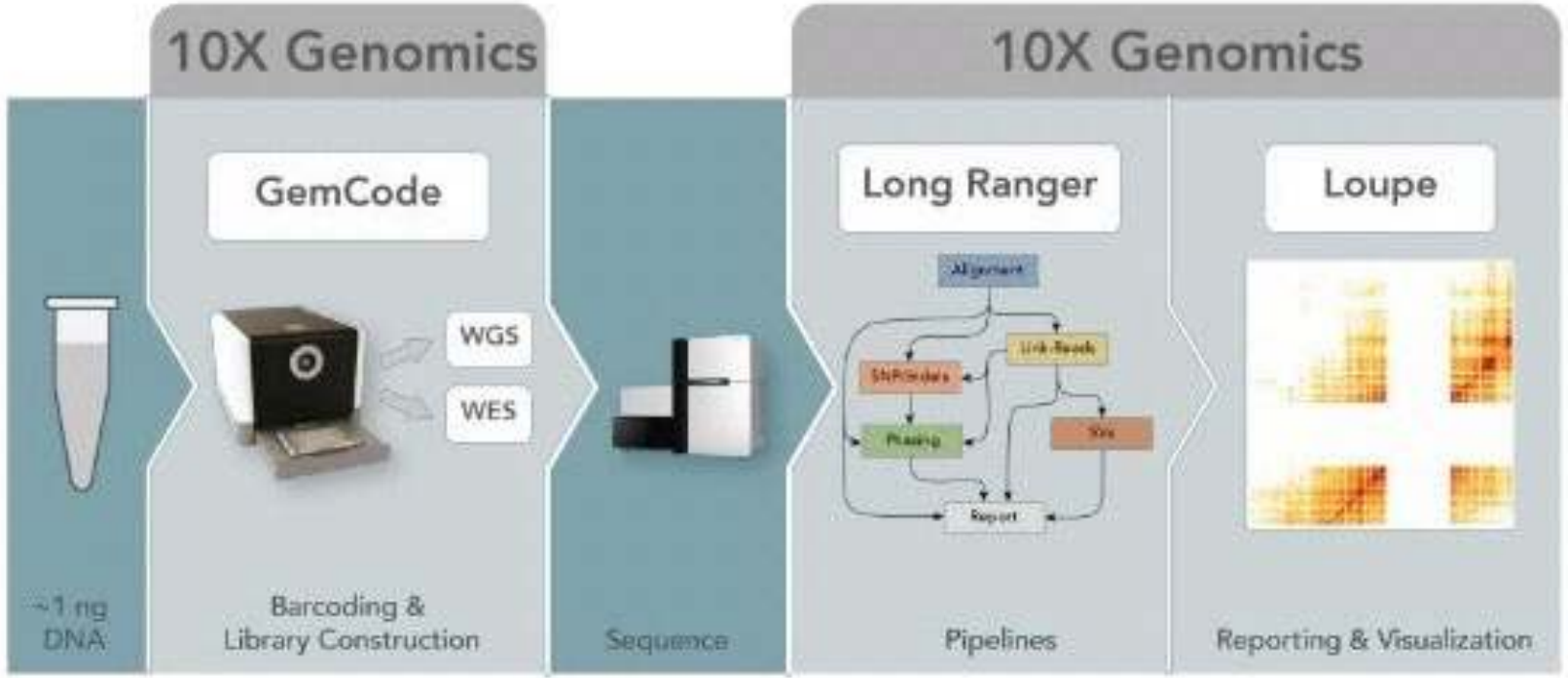
GENOMICS

Dovetail Genomics

Assembly	Data	N50	N90
Original	Shotgun + mate-pair	5.5 Mbp	0.9 Mbp
HiRise #1	Original assembly + 83X Chicago data	10 Mbp	2.2 Mbp
HiRise #2	Original assembly + 220X Chicago data	27 Mbp	9.5 Mbp

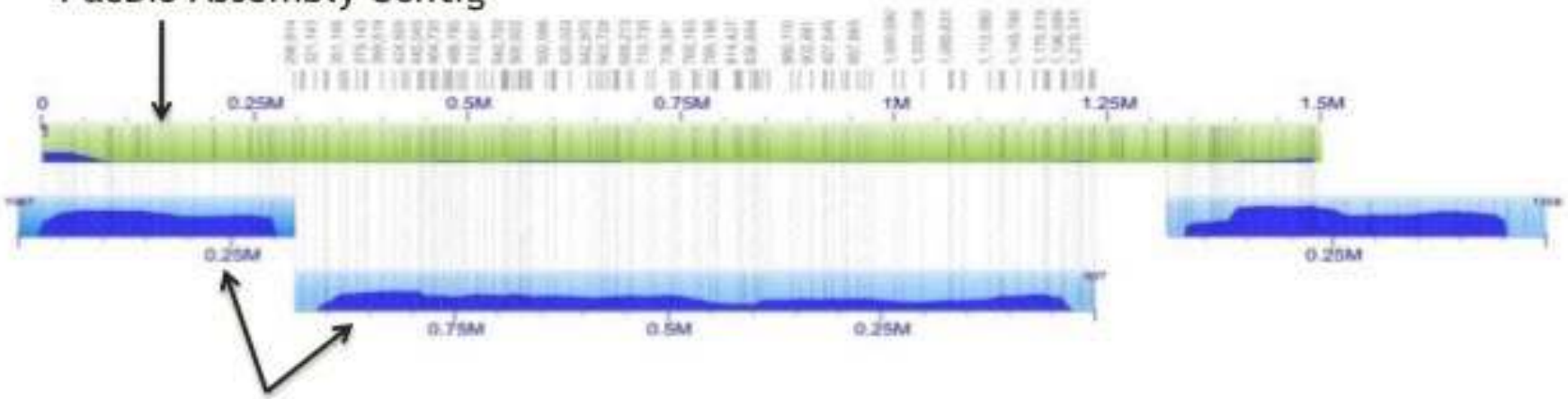

10X™
GENOMICS

GemCode Platform Workflow



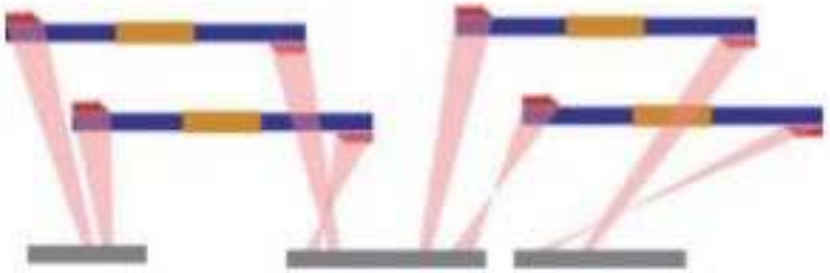
Sequencing Plan

PacBio Assembly Contig



BioNano Genome Map Contigs

Add 10X Linked Read information



Add Dovetail Hi-C/chiCago Data

Спасибо за внимание