



# HTS sequencing technologies

Sten Anslan <[sten.anslan@ut.ee](mailto:sten.anslan@ut.ee)>

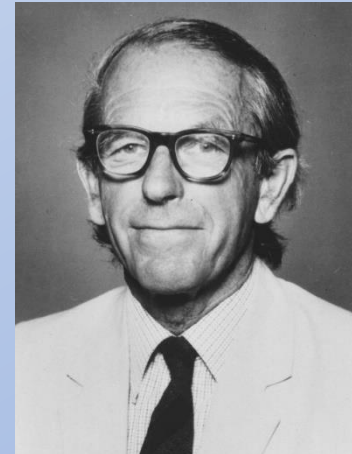
# First-generation DNA sequencing

- **1965** - first nucleic acid sequence (Robert W. Holley et al)  
(*Saccharomyces cerevisiae*)



Robert W. Holley  
(photo from the Nobel Foundation archive)

- **1977** - Sanger's chain-termination technique  
**Sanger sequencing**



Frederick Sanger  
(photo from Wikimedia Commons)

# First-generation DNA sequencing

- 1977 - Sanger's chain-termination technique (Sanger sequencing)

*Proc. Natl. Acad. Sci. USA*  
Vol. 74, No. 12, pp. 5463-5467, December 1977  
Biochemistry

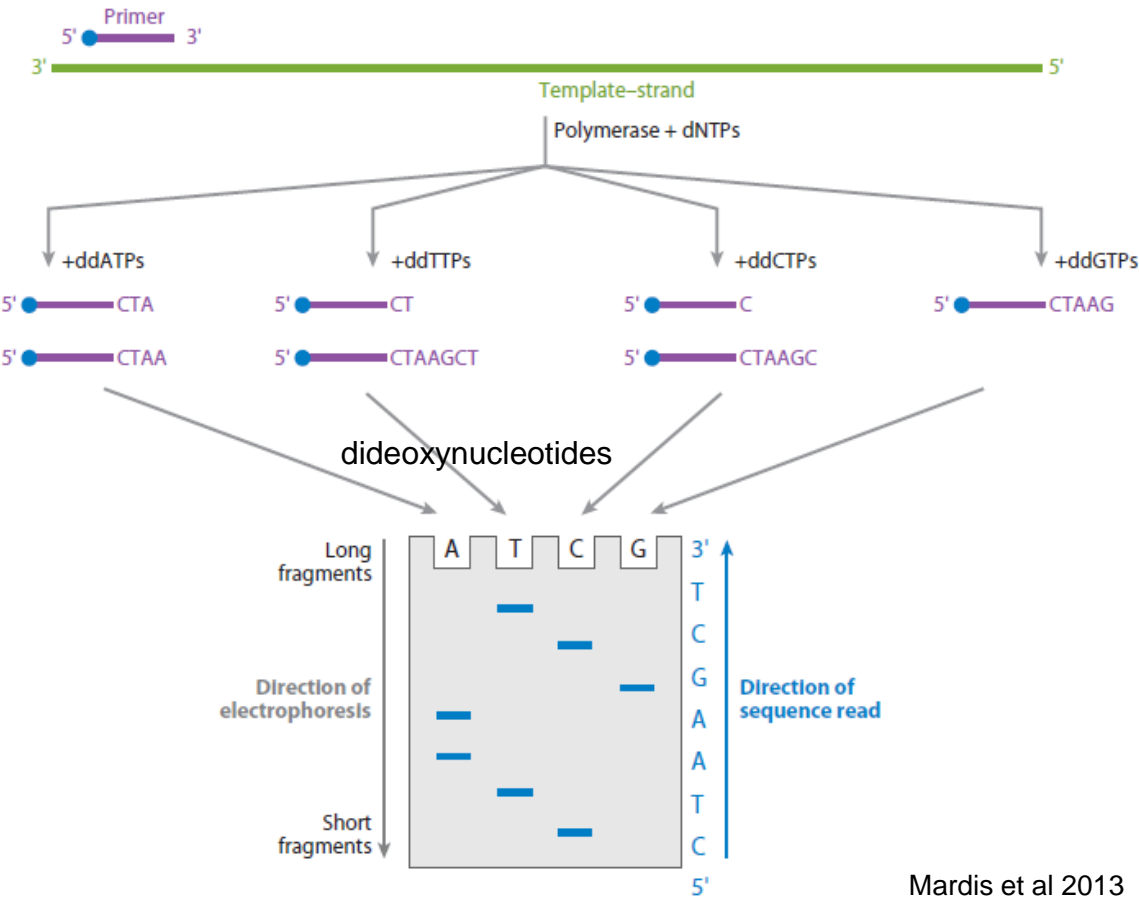
## DNA sequencing with chain-terminating inhibitors

(DNA polymerase/nucleotide sequences/bacteriophage  $\phi$ X174)

F. SANGER, S. NICKLEN, AND A. R. COULSON

Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 2QH, England

Contributed by F. Sanger, October 3, 1977



Mardis et al 2013

# First-generation DNA sequencing

- 1977 - Sanger's chain-termination technique  
(Sanger sequencing)

*Proc. Natl. Acad. Sci. USA*  
Vol. 74, No. 12, pp. 5463-5467, December 1977  
Biochemistry

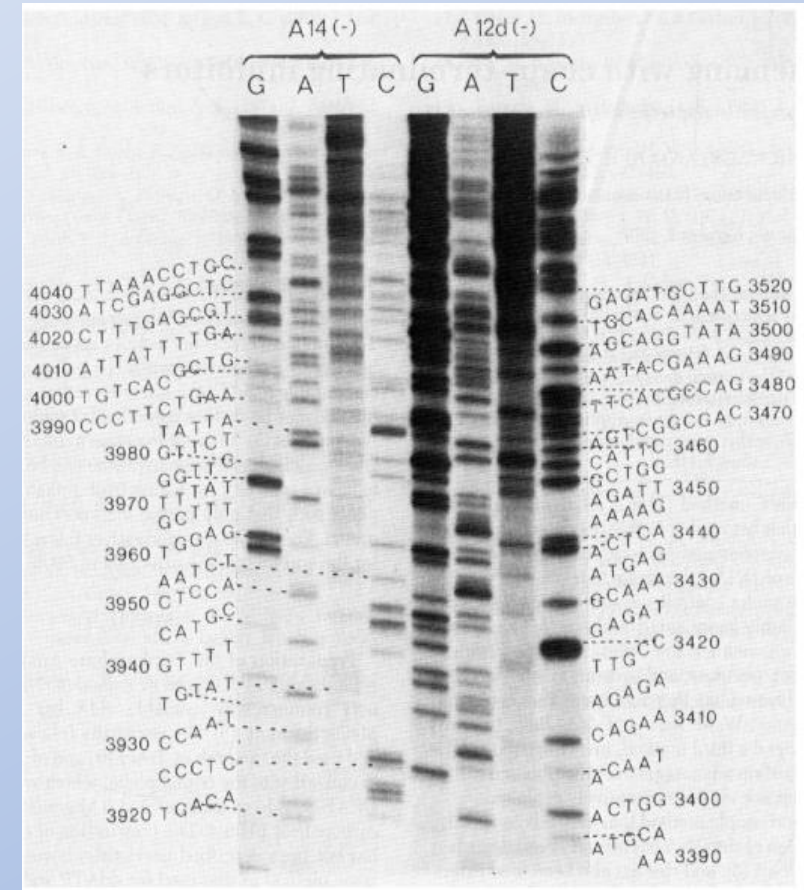
## DNA sequencing with chain-terminating inhibitors

(DNA polymerase/nucleotide sequences/bacteriophage  $\phi$ X174)

F. SANGER, S. NICKLEN, AND A. R. COULSON

Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 2QH, England

*Contributed by F. Sanger, October 3, 1977*



ca .14h gel (polyacrylamide) running

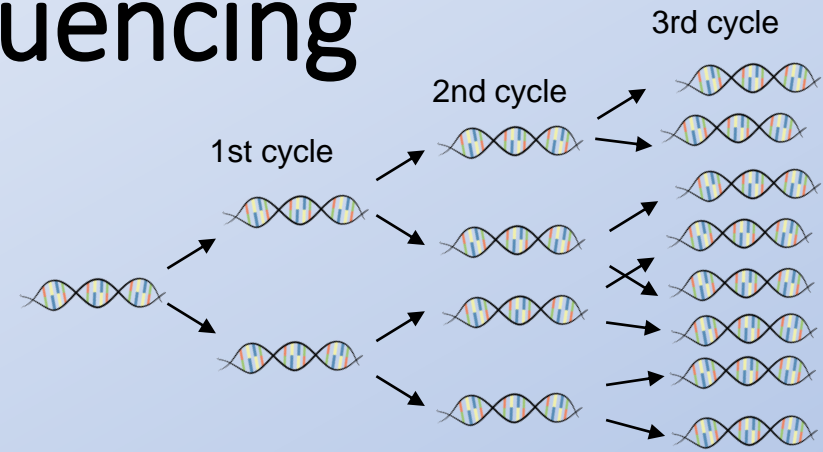
Sanger et al 1977

# First-generation DNA sequencing

- **1983** (-1986) – *Polymerase chain reaction (PCR)*,  
*Kary Mullis*
- **1986** – First automated DNA sequencing machine

*AB370A, introduced by Applied Biosystems*

*96 samples simultaneously*

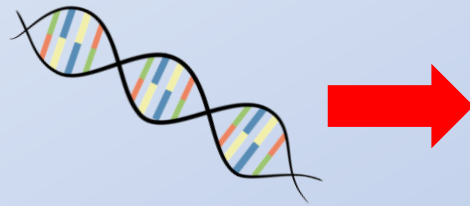


**Kary Banks Mullis**  
(1944 – 2019)

# First-generation DNA sequencing

## Sanger sequencing

sequencing by chain termination



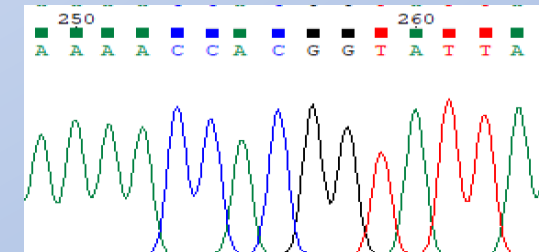
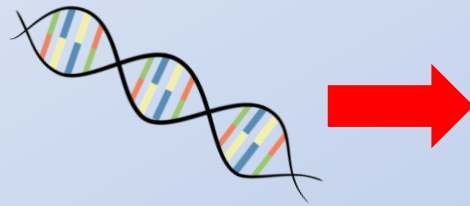
AAA**A**  
AAAAC**C**  
AAAACC**C**  
AAAACCA**A**  
AAAACCAC**C**  
AAAACCACG**G**  
AAAACCACGG**G**  
AAAACCACGGT**T**  
AAAACCACGGTA**A**  
AAAACCACGGTAT**T**  
AAAACCACGGTATT**T**  
AAAACCACGGTATT**A**



# First-generation DNA sequencing

## Sanger sequencing

sequencing by chain termination

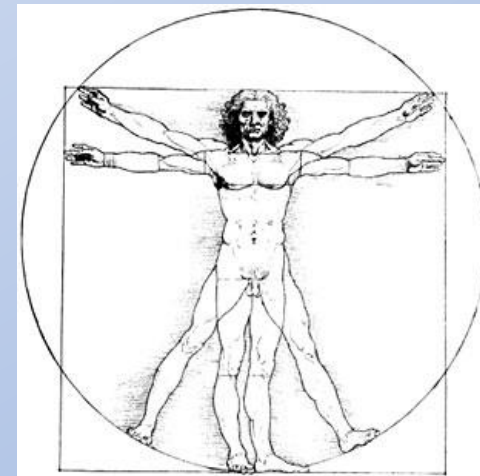


GGGAAGTACTCCACCTGGAGCCTCCGTAGACCTA  
ACCATCTTCTCCTTACACCTAGCAGGTGTCTCCTCT  
ATCTTAGGGGCCATCAATTTTCATCACAACAATTATC  
AATATAAAACCCCTGCCATAACCCAATACCAAACG  
CCCCTCTTCGTCTGATCCGTCCTAATCAGCAGTC  
CTACTTCTCCTATCTCTCCAGTCCTAGCTGCTGGC  
ATCACTATACTACTAACAGACCGCAACCTCAACACC  
ACCTTCTTCGACCCCGCCGAGGAGGAGACCCATT  
CTATACCAACACCTATTCTGATTTT

*Homo sapiens*  
cytochrome oxidase subunit I (COI) gene,  
partial cds; mitochondrial

# First-generation DNA sequencing

- **1983** – *Polymerase chain reaction (PCR)*
- **1986** – First automated DNA sequencing machine
- **1990** - The Human Genome Project





# First-generation DNA sequencing

- **April 2003**


The Human Genome Project completed

***\$3-billion project***



[https://en.wikipedia.org/wiki/DNA\\_sequencer#/media/File:DNA-Sequencers\\_from\\_Flickr\\_57080968.jpg](https://en.wikipedia.org/wiki/DNA_sequencer#/media/File:DNA-Sequencers_from_Flickr_57080968.jpg)

# High-throughput sequencing (HTS)

- High-throughput sequencing (**HTS**)
  - Next-generation sequencing (**NGS**)
- 
- 2<sup>nd</sup> and 3<sup>rd</sup> generation sequencing

Techniques that allow the simultaneous sequencing of millions of DNA fragments

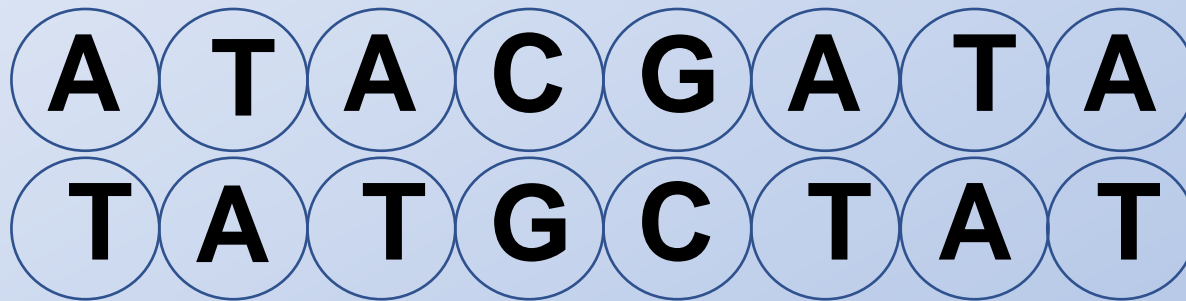
	Sanger	HTS
Sequences per run	960	~5 000 000 000

# Second-generation sequencing

sequencing by synthesis

(Sanger - sequencing by chain termination)

## Sequencing by synthesis



→ TATGCTAT  
JIHEJJB<

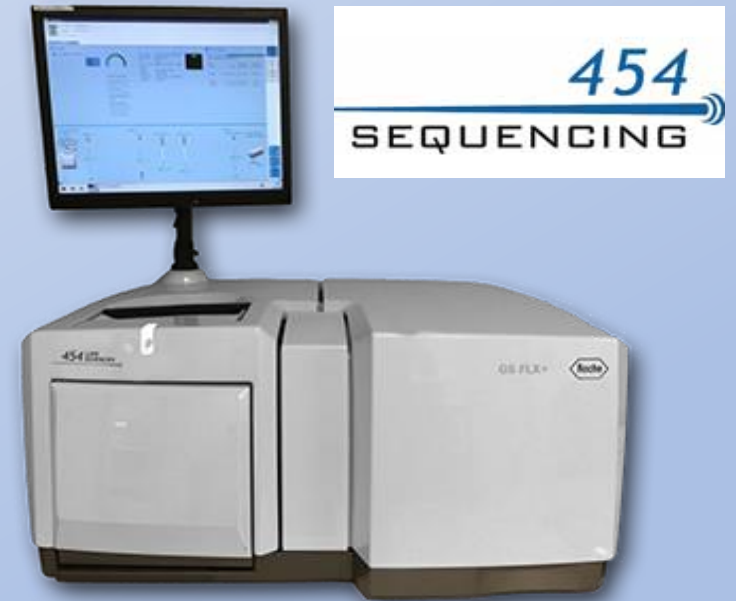
# Second-generation sequencing

sequencing by synthesis

- **454 Life Sciences** (released in 2005) ---- > **Roche** (2007 - 2013)  
*(pyrosequencing)*

Max read length: **700 bp**

Data: ~ **700 Mb**



# Second-generation sequencing

sequencing by synthesis

- 454 Life Sciences (released in 2005) ---- > Roche 454 (2007-2013)  
(*pyrosequencing*)
- Solexa Genome Analyzer (released in 2006) - > **illumina**<sup>®</sup> (2007 - )

Read length: **2x75 bp**

Data: up to **1 GB**



illumina.com

# Second-generation sequencing

sequencing by synthesis

Max read length: MiSeq, **2x300 bp** (release: 2011); NovaSeq **2x250 bp** (release: 2019)  
Data: MiSeq, up to **15 Gb**; NovaSeq **6000 Gb**





# Second-generation sequencing

sequencing by synthesis

- 454 Life Sciences (released in 2005) ---- > Roche 454 (2007-2013)
- Solexa Genome Analyzer (released in 2006) ---- > Illumina (2007 - )

- **Ion Torrent**  
commercially available since **2010**

Max read length: **400 bp**

Data: up to **50 Gb**

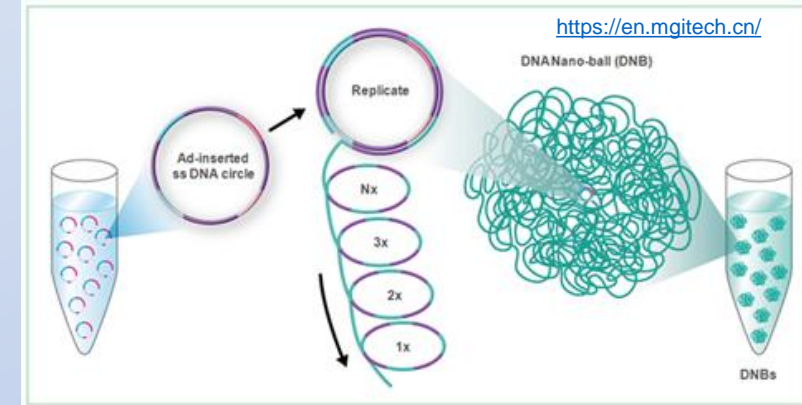
## ion torrent



# Second-generation sequencing

combinatorial probe-anchor synthesis (cPAS)

DNA nanoball sequencing, 2009



DNBSEQ platform was released by BGI in **2015** (BGISEQ-500)



In **2017**, **DNBSEQ-G400** was released by MGI  
(MGISEQ-2000)



Max read length: **400 bp**  
Data: up to **1440 Gb**



# Second-generation sequencing

combinatorial probe-anchor synthesis (cPAS)

September 2019. **DNBSEQ-T7**

Max read length: **2x150 bp**  
Data: up to **6000 Gb**



<https://en.mgitech.cn/>

# Highly comparable metabarcoding results from MGI-Tech and Illumina sequencing platforms

Sten Anslan<sup>1,2</sup>, Vladimir Mikryukov<sup>1,2</sup>, Kęstutis Armolaitis<sup>3</sup>, Jelena Ankuda<sup>3</sup>, Dagnija Lazdina<sup>4</sup>, Kristaps Makovskis<sup>4</sup>, Lars Vesterdal<sup>5</sup>, Inger Kappel Schmidt<sup>5</sup> and Leho Tedersoo<sup>1,2</sup>

**Table 1** Cost calculations for Illumina NovaSeq 6000 and MGI-Tech DNBSEQ-G400RS based on the best offering service providers and data retrieved (euros).

	NovaSeq 6000 (2 × 250 bp)	DNBSEQ-G400RS* (2 × 200 bp)
Library preparation for sequencing	100	170
Offer for sequencing 50 million reads	1,000	170
Actual cost per million raw reads	30.07	7.21
Actual cost per million filtered reads (matrix#1)	53.23	11.92
<u>Actual cost per raw gigabit (Gb)</u>	<u>26.44</u>	<u>8.25</u>
Actual cost per filtered Gb (merged and quality filtered)	104.76	33.66

# Second-generation sequencing

sequencing by synthesis

December 2021.  
Singular Genomics **G4** platform

- Novel high-speed chemistry

Max read length: **2x150 bp**  
Data: up to **100 Gb**



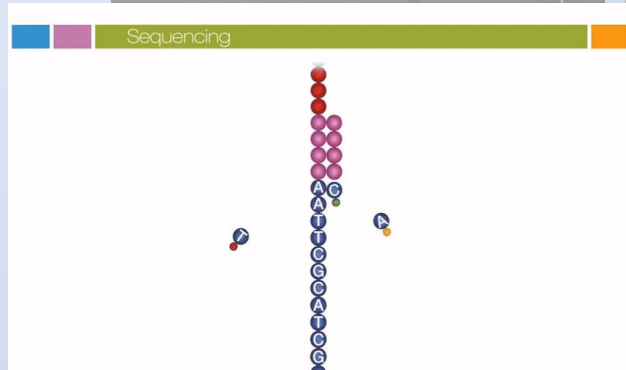
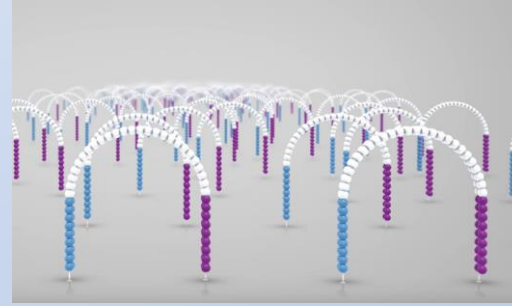
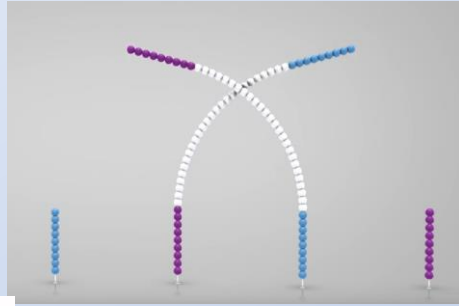
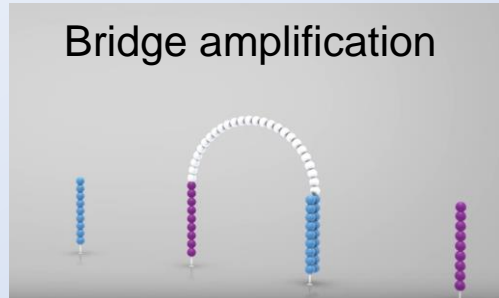
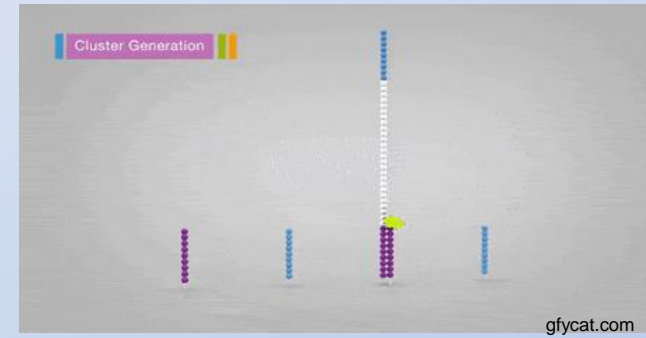
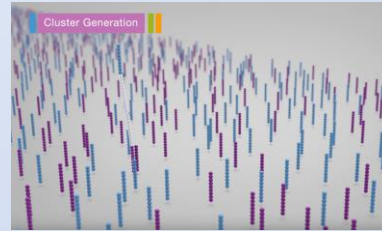
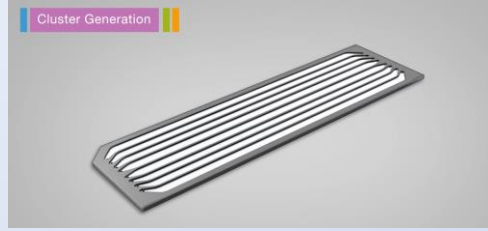
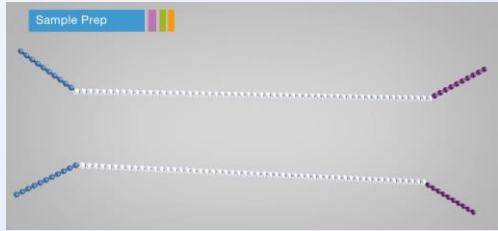
[singulargenomics.com](https://singulargenomics.com)

# Second-generation sequencing



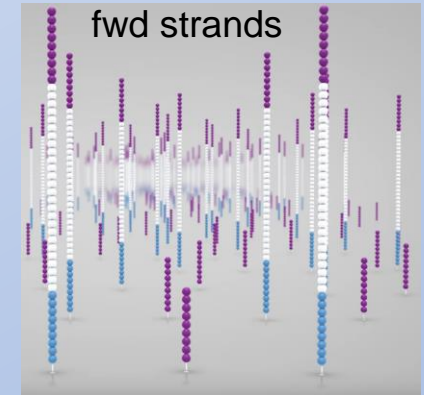
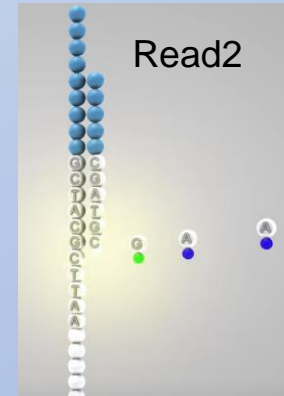
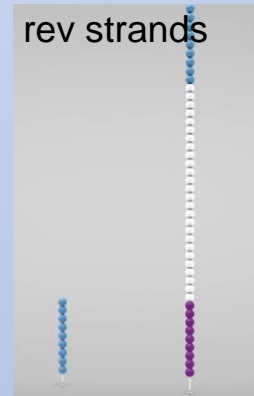
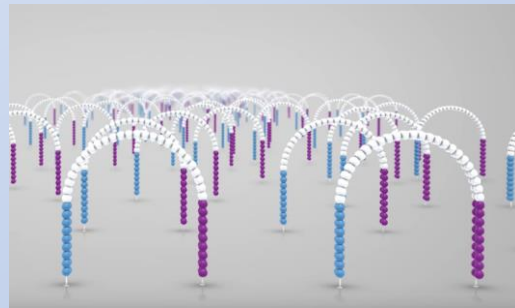


# illumina® sequencing



Read1

[gfycat.com/fairhonestamazontreeboa](http://gfycat.com/fairhonestamazontreeboa)



Read1

300 bp

300 bp  
Read2

[illegible]

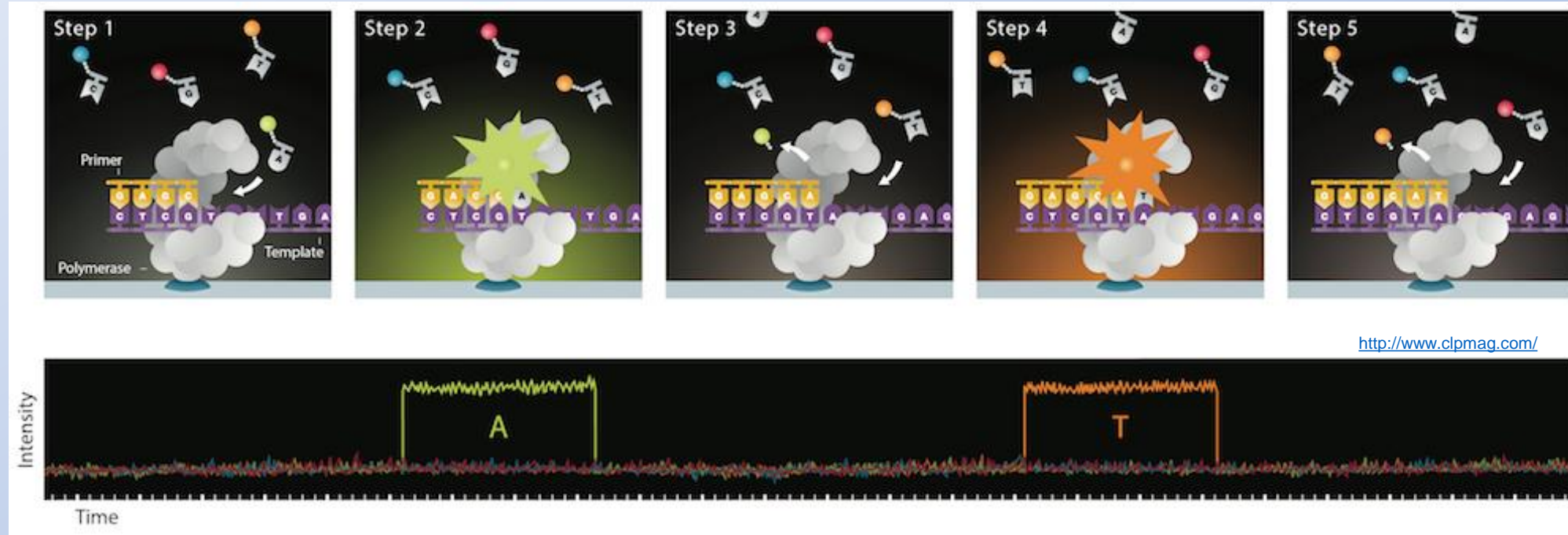
# Third-generation sequencing

Bigger pieces of puzzle  
makes assembly easier



## Long-read sequencing

Single molecule real-time sequencing



sequence directly RNA

# Third-generation sequencing

- **Helicos Genetic Analysis System**, released in 2009 (-2012)

First to use the principle of single molecule sequencing.

Read length: **35 bp**

Data: up to **35 Gb**



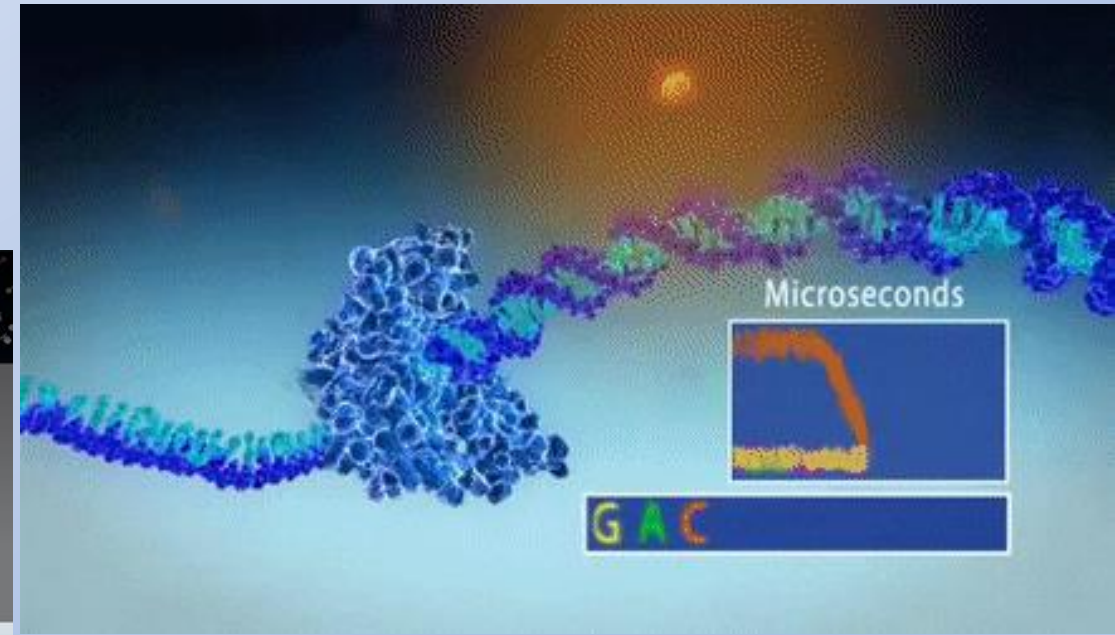
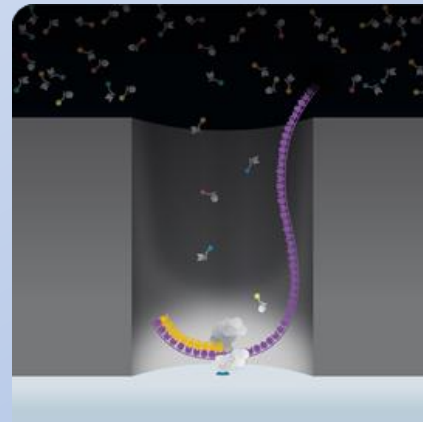
<https://www.technologyreview.com/>



# Third-generation sequencing

*single molecule sequencing*

- Pacific Biosciences (PacBio) (commercially available since **2011**)



[glycat.com/impolitethisamericanblackvulture](http://glycat.com/impolitethisamericanblackvulture)

# Third-generation sequencing

*single molecule sequencing*

- Pacific Biosciences (PacBio) (commercially available since **2011**)

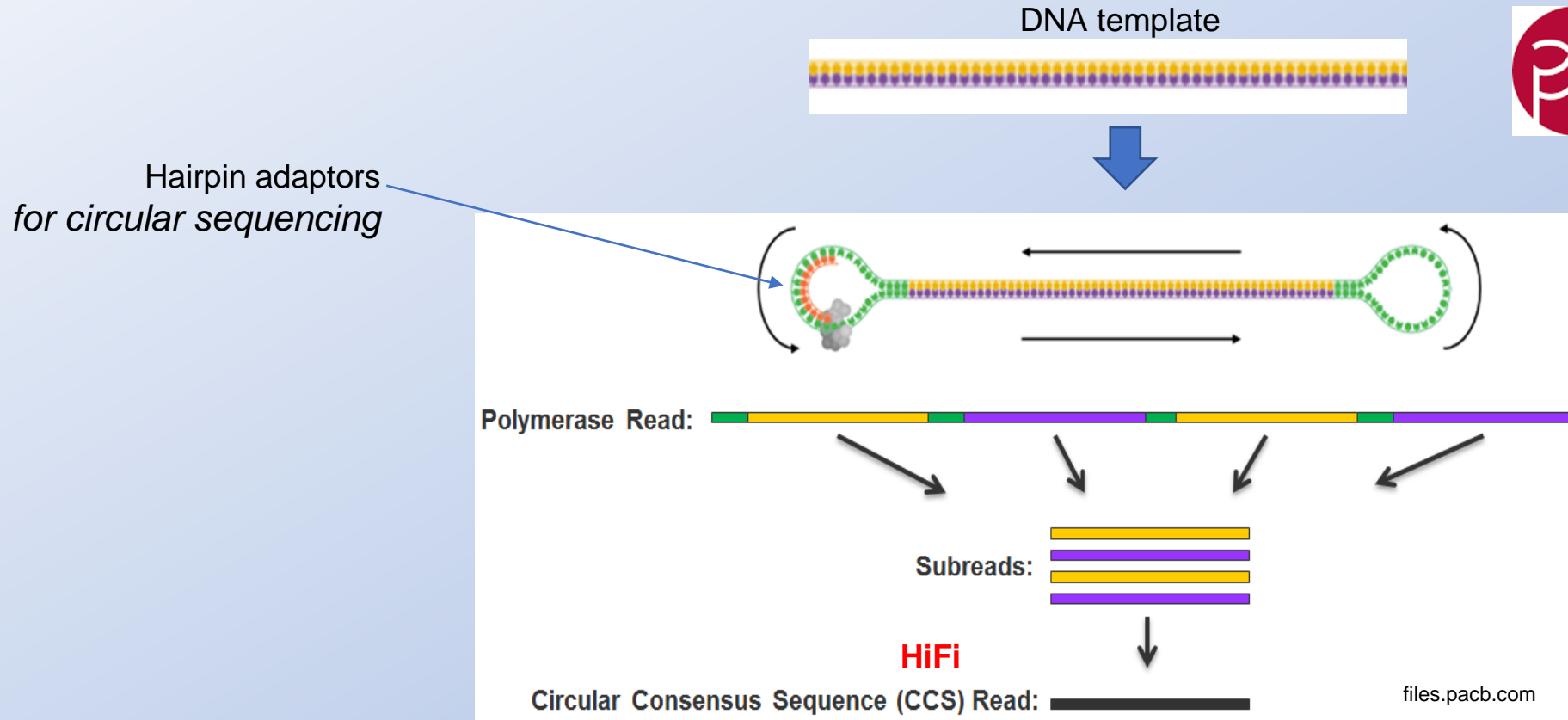


HiFi read length: ~**20 000 bp**  
Data: up to **50 Gb**



# Third-generation sequencing

*single molecule sequencing*



files.pacb.com

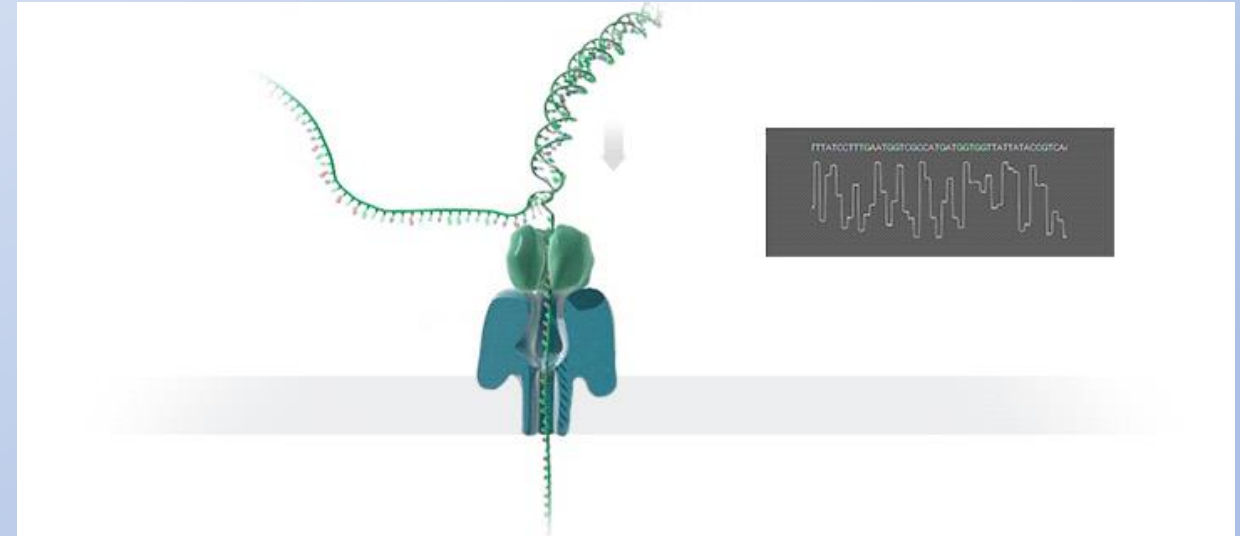
# Third-generation sequencing

*single molecule sequencing*

- **Oxford Nanopore Technologies (ONT)**  
(commercially available since **2015**)



<https://nanoporetech.com/>



Read length: **>100 000 bp (2mil)**  
Data: up to **8000 Gb**

# 4th-generation sequencing

S T R A T O S  
genomics inc.

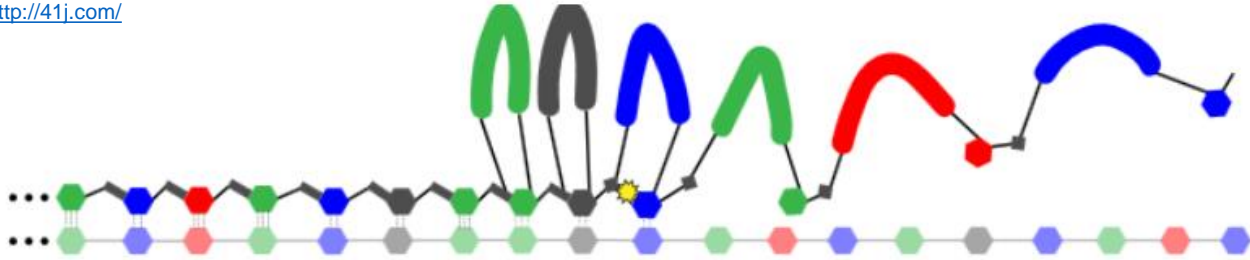


2020



## Sequencing by Expansion (SBX)

<http://41j.com/>



Xpandomer – 50x larger

Product in development, not commercially available.



# 4th-generation sequencing

S T R A T O S  
genomics inc.

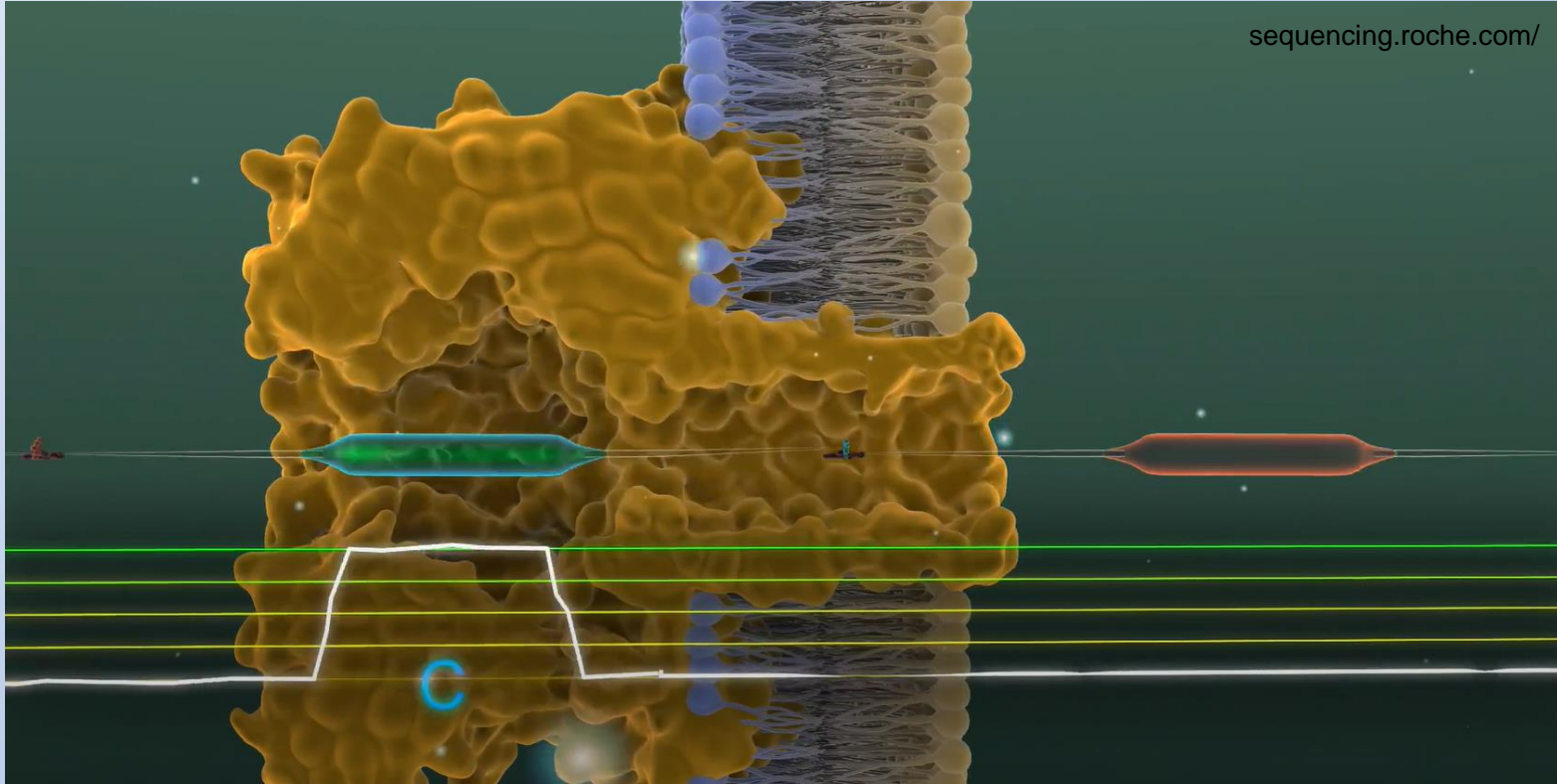
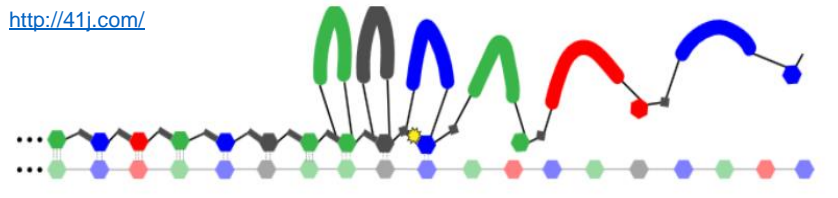


2020

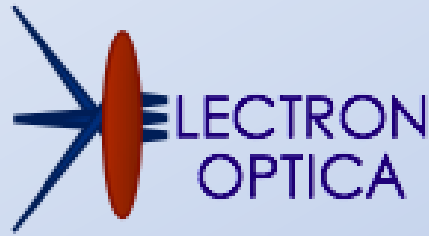


Sequencing by Expansion (SBX)

<http://41j.com/>



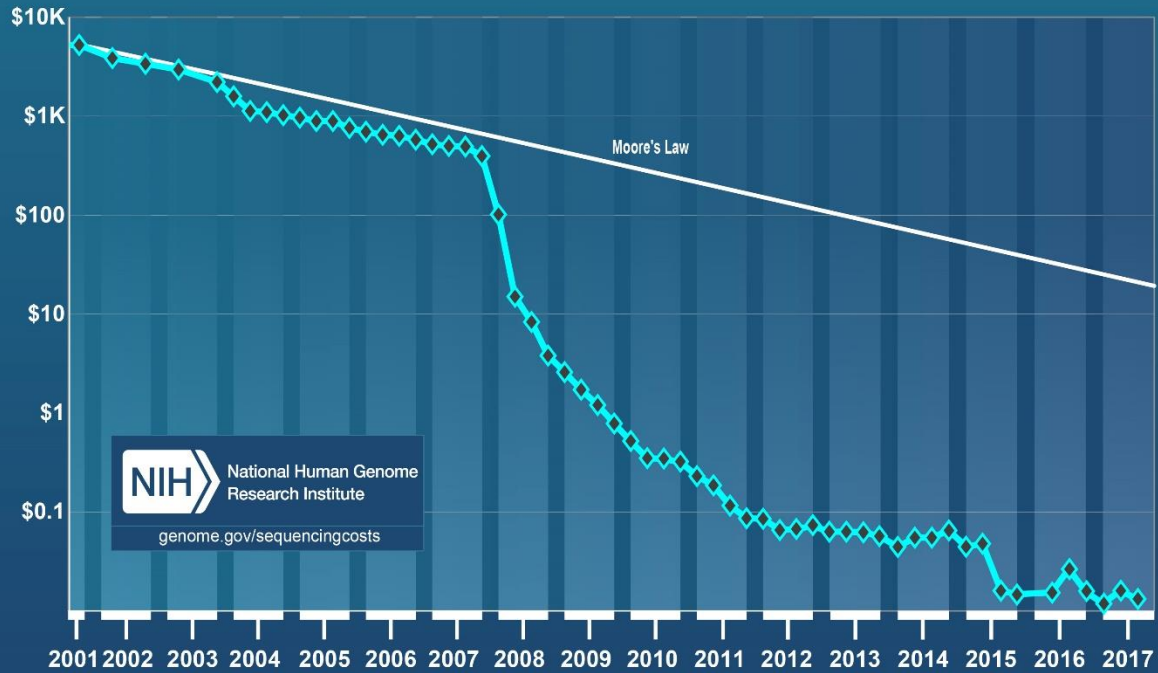
# Emerging Technologies



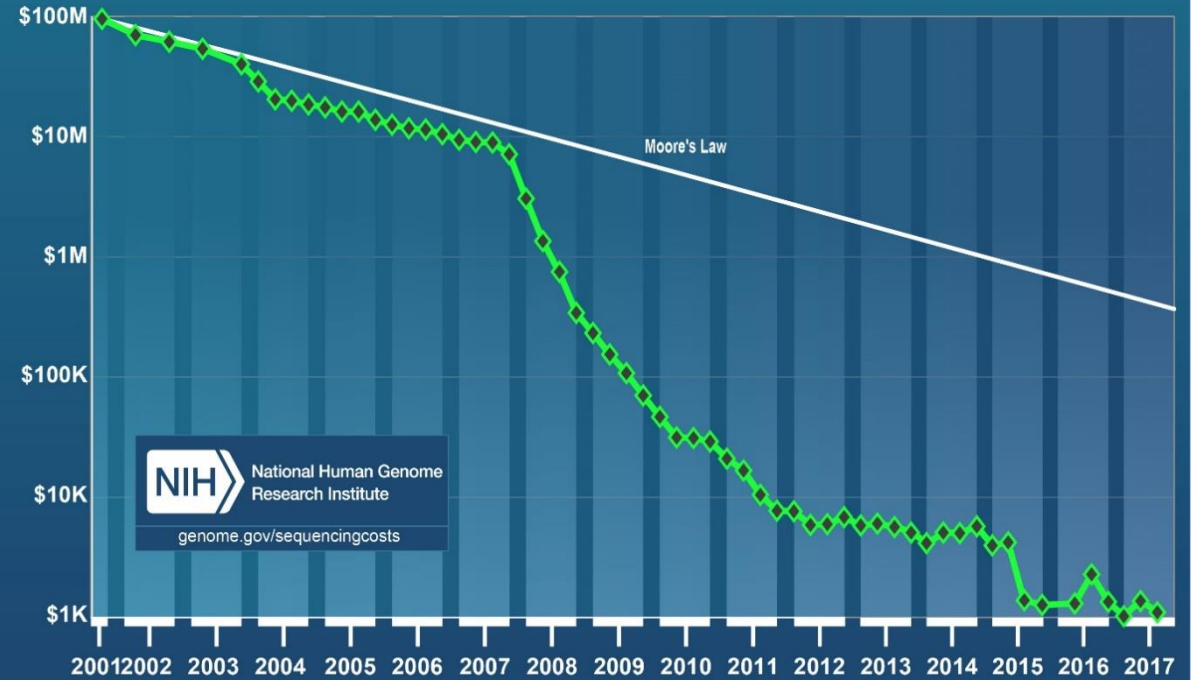


# DNA sequencing costs

*Cost per Raw Megabase of DNA Sequence*



*Cost per Genome*





Gen.		Read length	Output up to	Error rate
<i>1st</i>	Sanger	1000 bp	few Mb	0.001%
<i>2nd</i>	Roche 454	700 bp	0.7 Gb	< 2%
<i>2nd</i>	Ion Torrent	400 bp	50 Gb	< 2%
<i>2nd</i>	Illumina MiSeq	2x300 bp	15 Gb	0.1%
<i>2nd</i>	Illumina NovaSeq	2x250 bp	6000 Gb	0.1%
<i>2nd</i>	DNBSeq-G400	400 bp	1440 Gb	0.1%
<i>2nd</i>	DNBSEQ-T7	2x150 bp	6000 Gb	0.1%
<i>3rd</i>	MinION	> 100 000 bp	50 Gb	~1%
<i>3rd</i>	PromethION	> 100 000 bp	14000 Gb	~1%
<i>3rd</i>	Sequel II	~20 000 bp	500 Gb	<0.1% HiFi (ccs) reads



# HTS sequencing technologies

Sten Anslan <[sten.anslan@ut.ee](mailto:sten.anslan@ut.ee)>