



测序技术回顾与展望

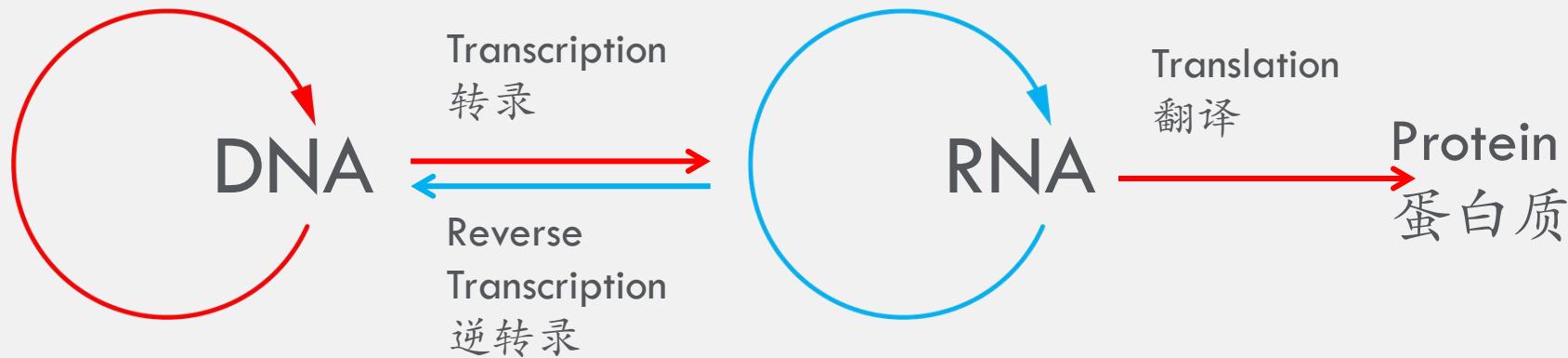
2021.10.09

Ming Ni

niming@genomics.cn

The difference in DNA sequence makes the difference in life

Central Dogma

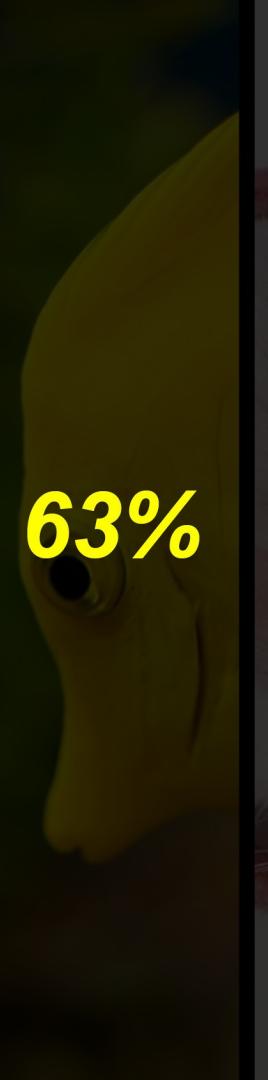




17%



39%



63%



80%



93%

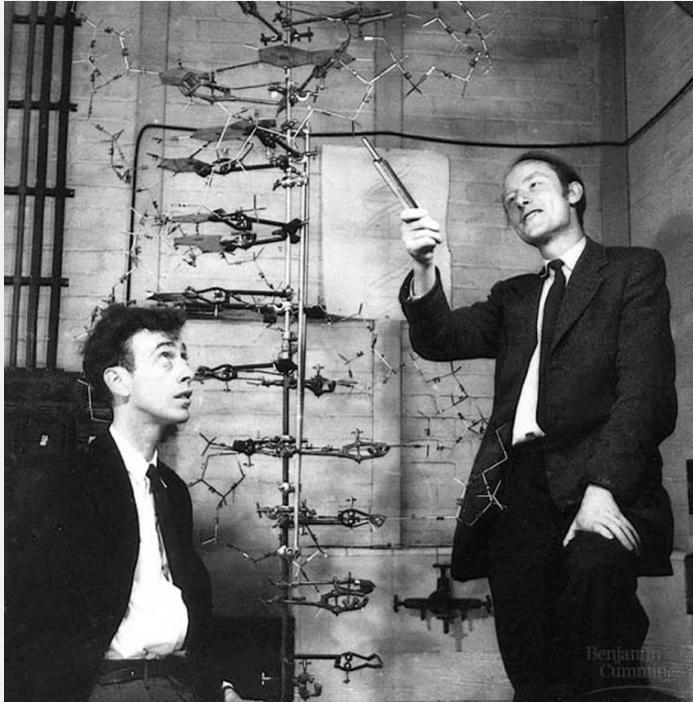


96%



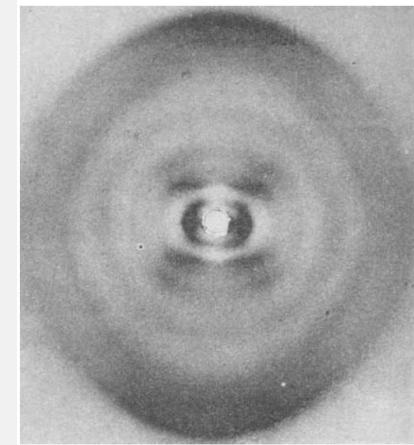
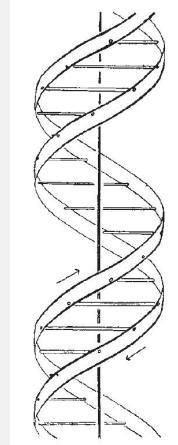
99.5%

DNA replication mechanism is embedded in it's double helix structure



James Watson

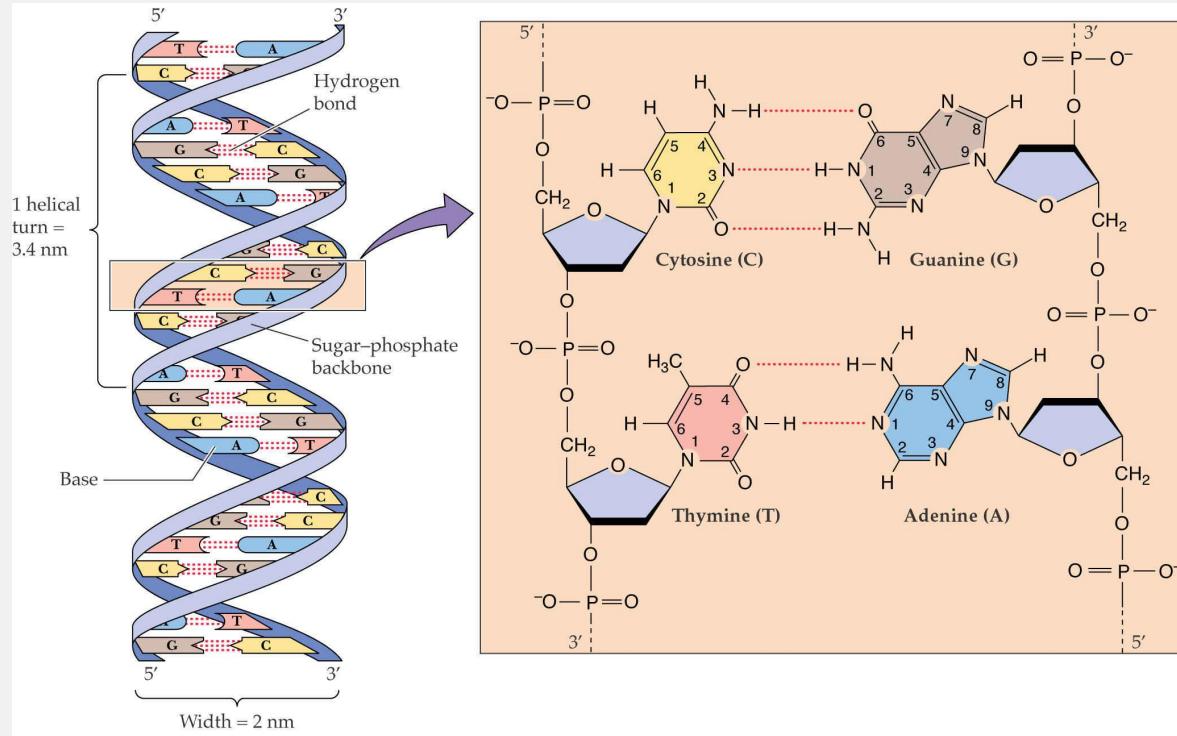
Francis Crick



It has not escaped our notice that the specific pairing that we have postulated immediately suggests a possible **copying mechanism** for the genetic material.

Nature **171**, 737–738 (1953)

Numbers about DNA and Genome



Length per base pare: 0.3nm
Width of base pare: 2nm
Molecular Weight: 650DA

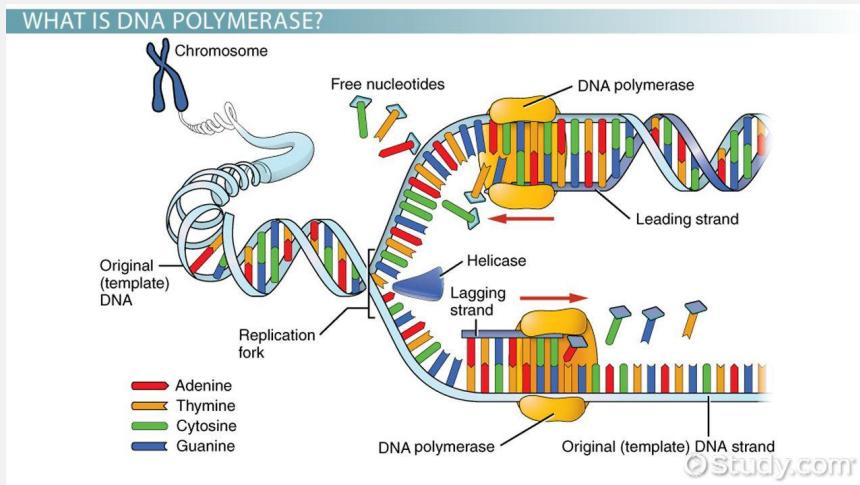
Genome length

Human:	3 Billion
Mouse:	3 Billion
Corn:	3.9 Billion
Fruit Fly:	0.14 Billion
Budding Yeast:	0.015 Billion
Escherichia Coli:	0.0047 Billion

Difficulty in Sequencing:
High throughput
High precision
Long continue reads

How to recognize such nano-scale particles, quickly and precisely?

DNA polymerase: The nano machine that can recognize nuclear base



Signal amplification: PCR/MDA/RCA

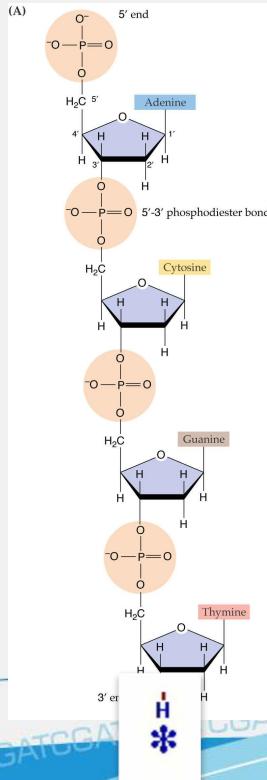
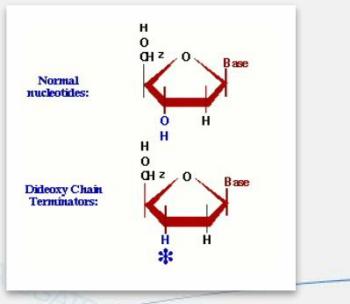


Signal read out: Optical/electrical



Sanger method

In 1977, Fred Sanger achieved signal synchronization by stopping the polymerase.

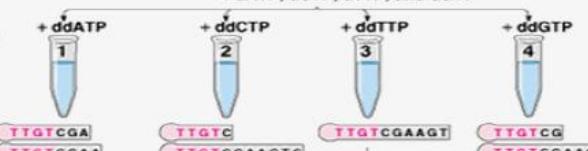


- ① Incubation of single-stranded DNA with unknown sequence in DNA synthesis reaction mixtures containing dideoxyribonucleotides

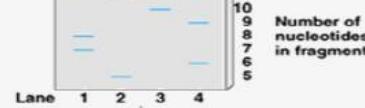
Unknown sequence
3' AACA**GCTTCAGT** 5'

+ 5' TTGT 3' Labeled primer
+ DNA polymerase
+ dATP, dCTP, dTTP, and dGTP

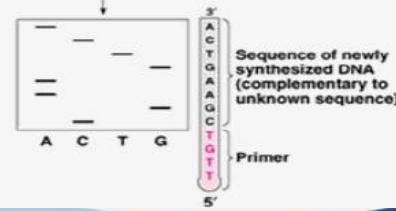
- ② Products of the reactions



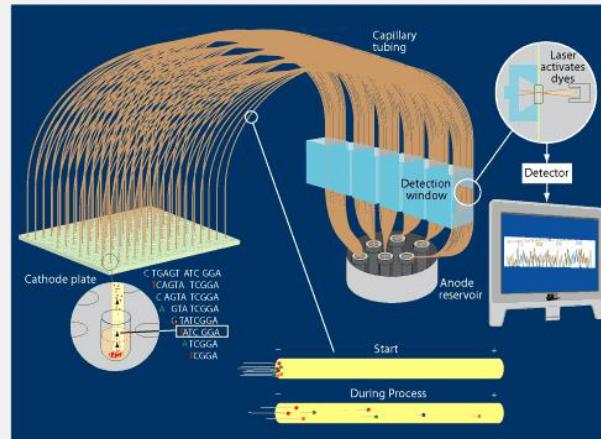
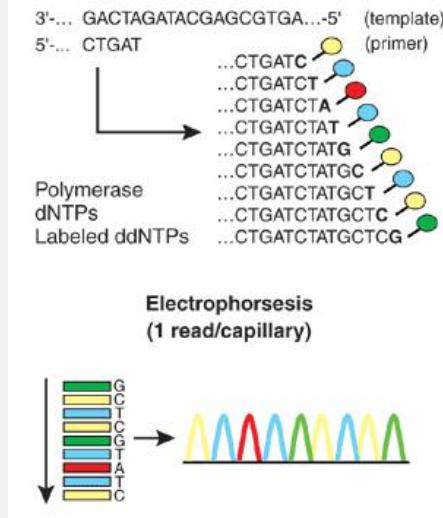
- ③ Electrophoresis of reaction mixtures



- ④ Autoradiography to visualize bands and deduction of 5' → 3' sequence of newly synthesized DNA strand by reading order of bands from bottom to top



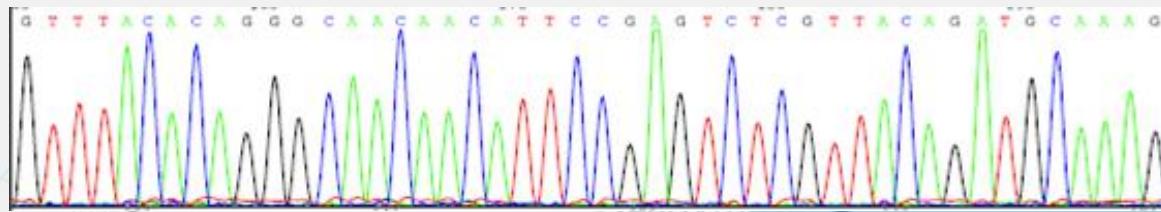
High throughput 1s generation sequencing



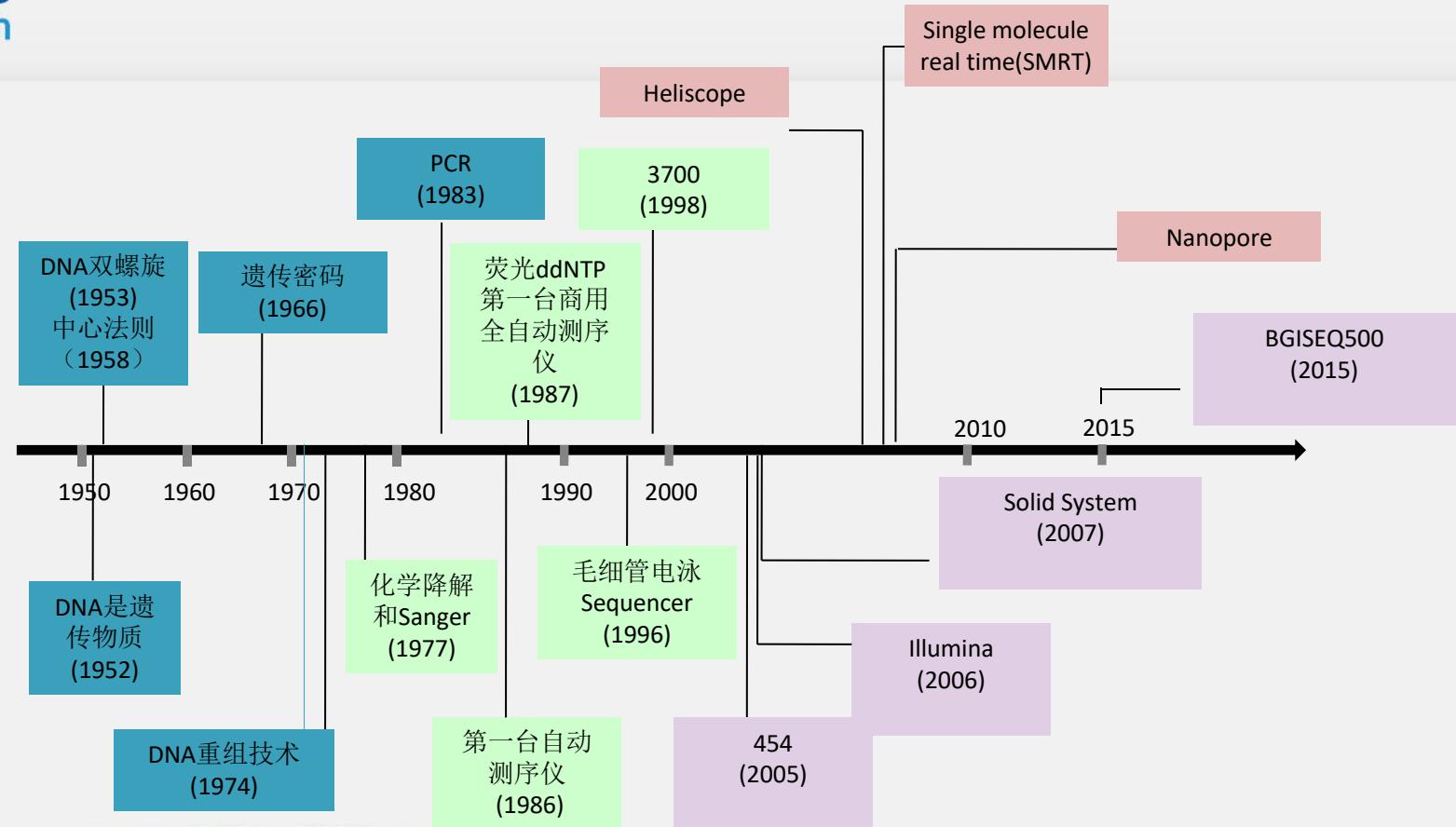
Read length: 1000bp

Accuracy: 99.999%

Through put: 600000
 per run



Cost: 500 Million per
 100G (30X Human
 Genome)



Current NGS sequencing platforms

Throughput
(G/run)

1G

6T



iSeq 100系统



MiniSeq系统



MiSeq系列⊕



NextSeq系列⊕



HiSeq系列⊕



HiSeq X系列‡



NovaSeq系列⊕



BGISEQ-50



MGISEQ-200



BGISEQ-500



MGISEQ-2000



DNBSEQ-T7



Ion PGM

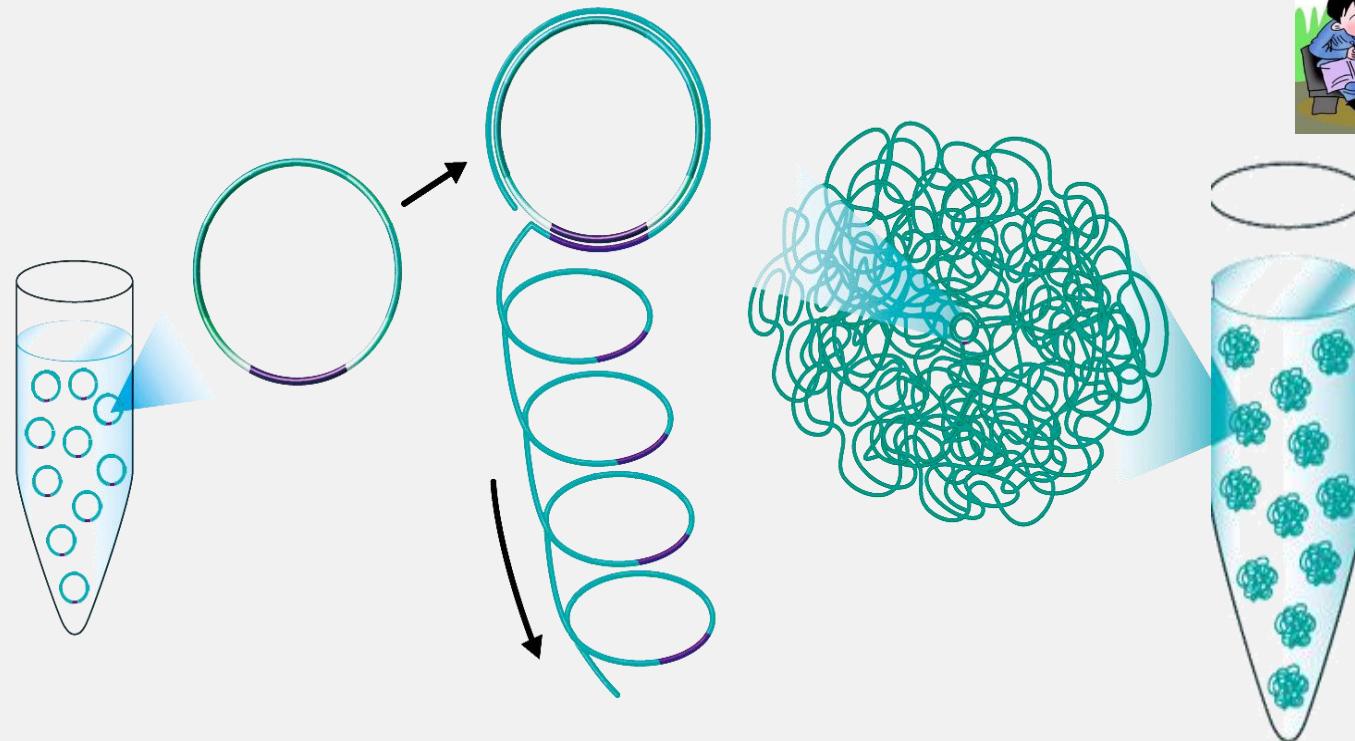


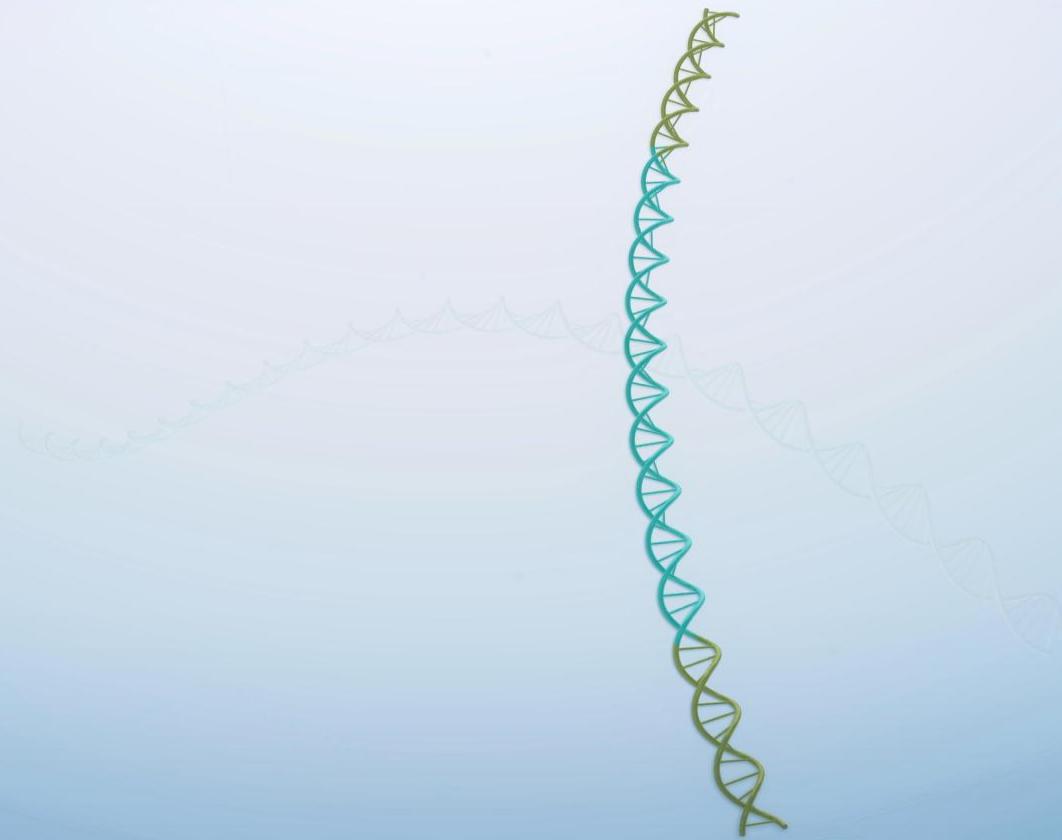
Ion Proton



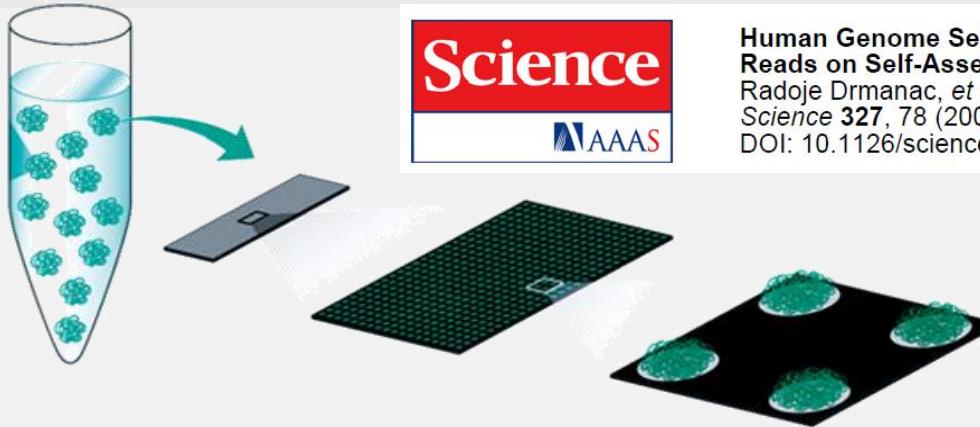
Ion S5

基于DNA 纳米球核心技术进行高通量测序

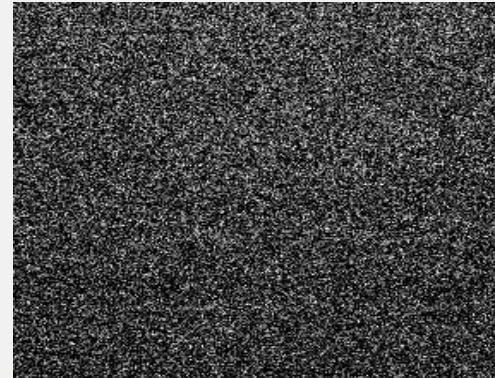
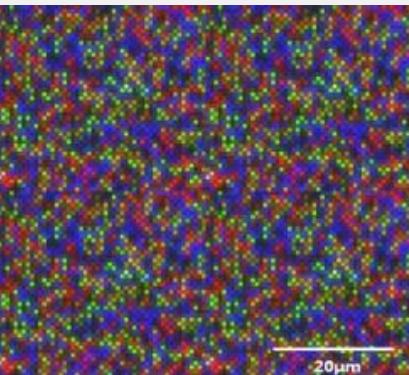
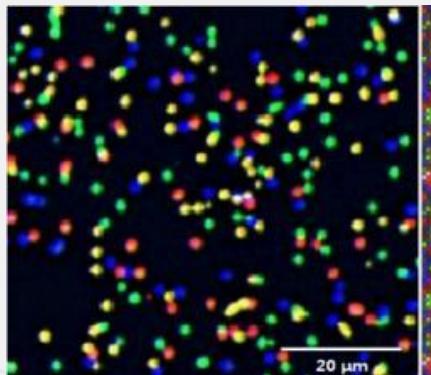


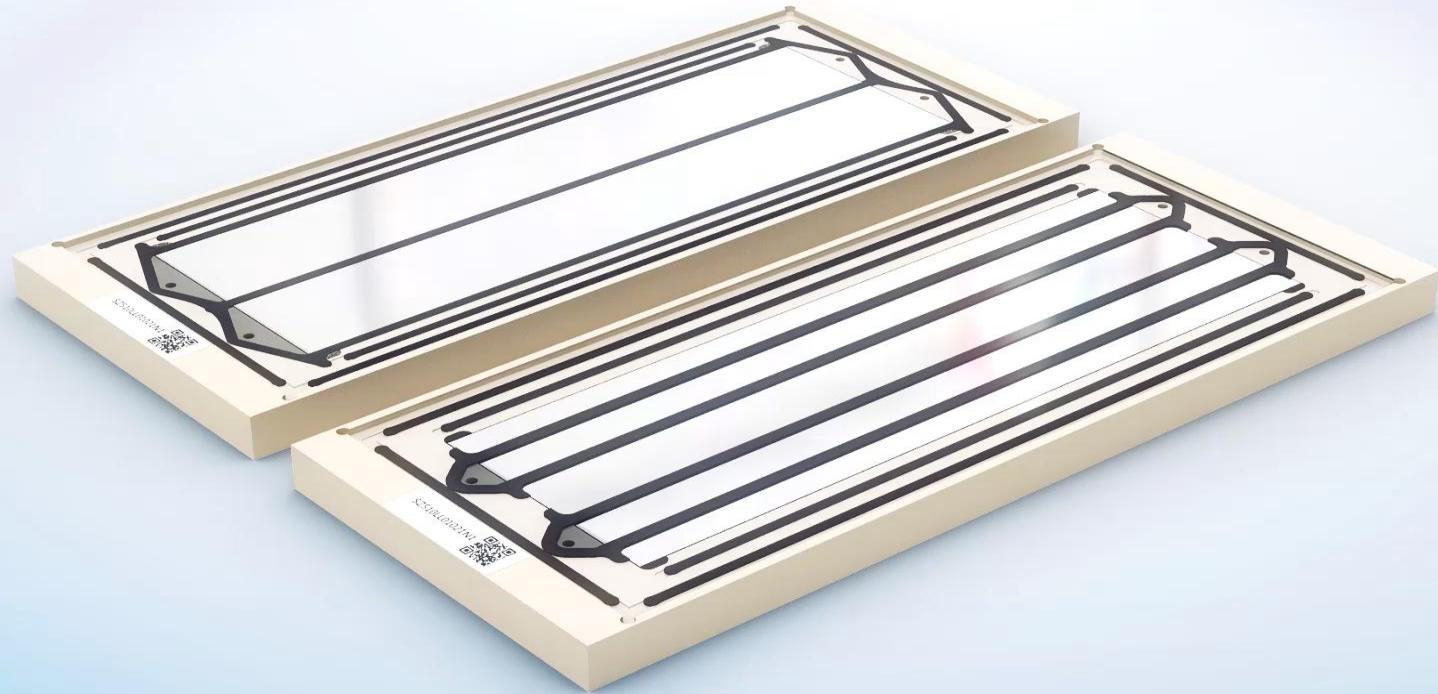


纳米阵列式测序芯片用于承载DNA纳米球

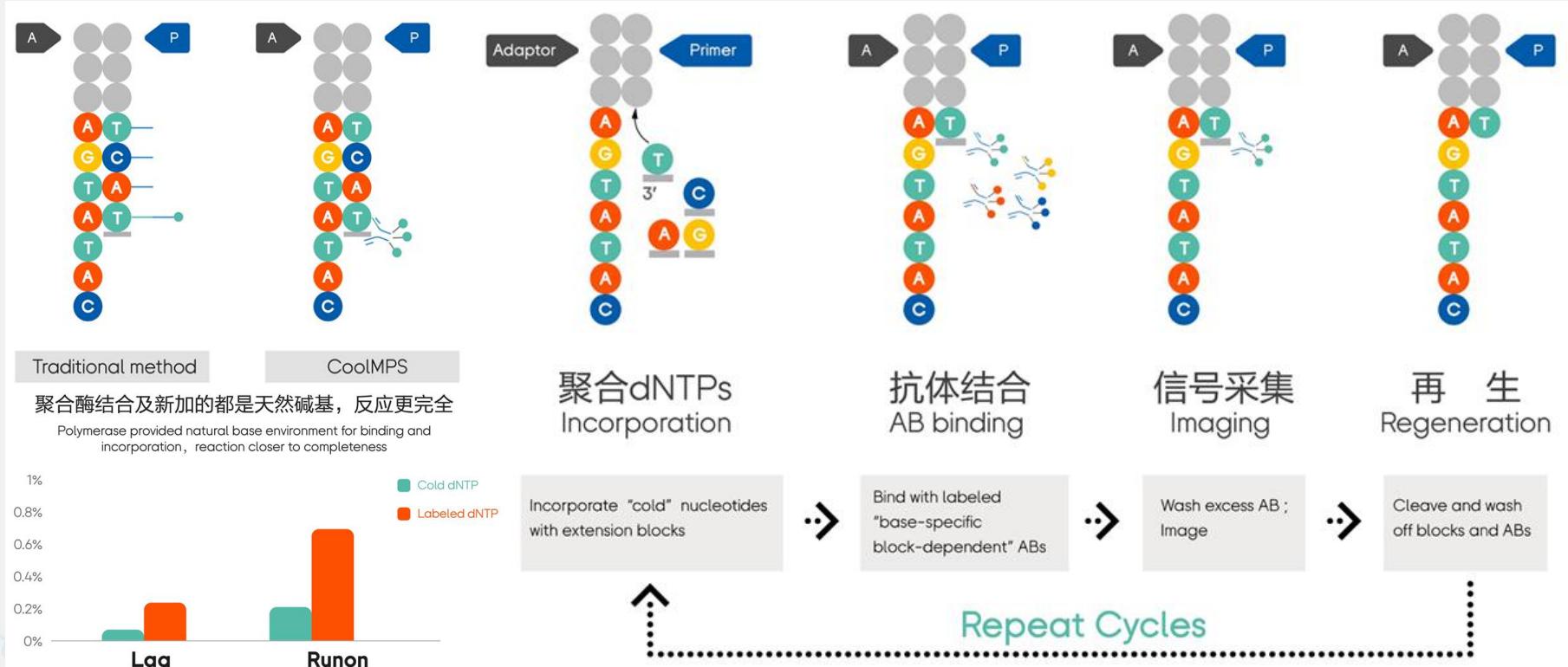


Human Genome Sequencing Using Unchained Base
Reads on Self-Assembling DNA Nanoarrays
Radoje Drmanac, et al.
Science 327, 78 (2009);
DOI: 10.1126/science.1181498

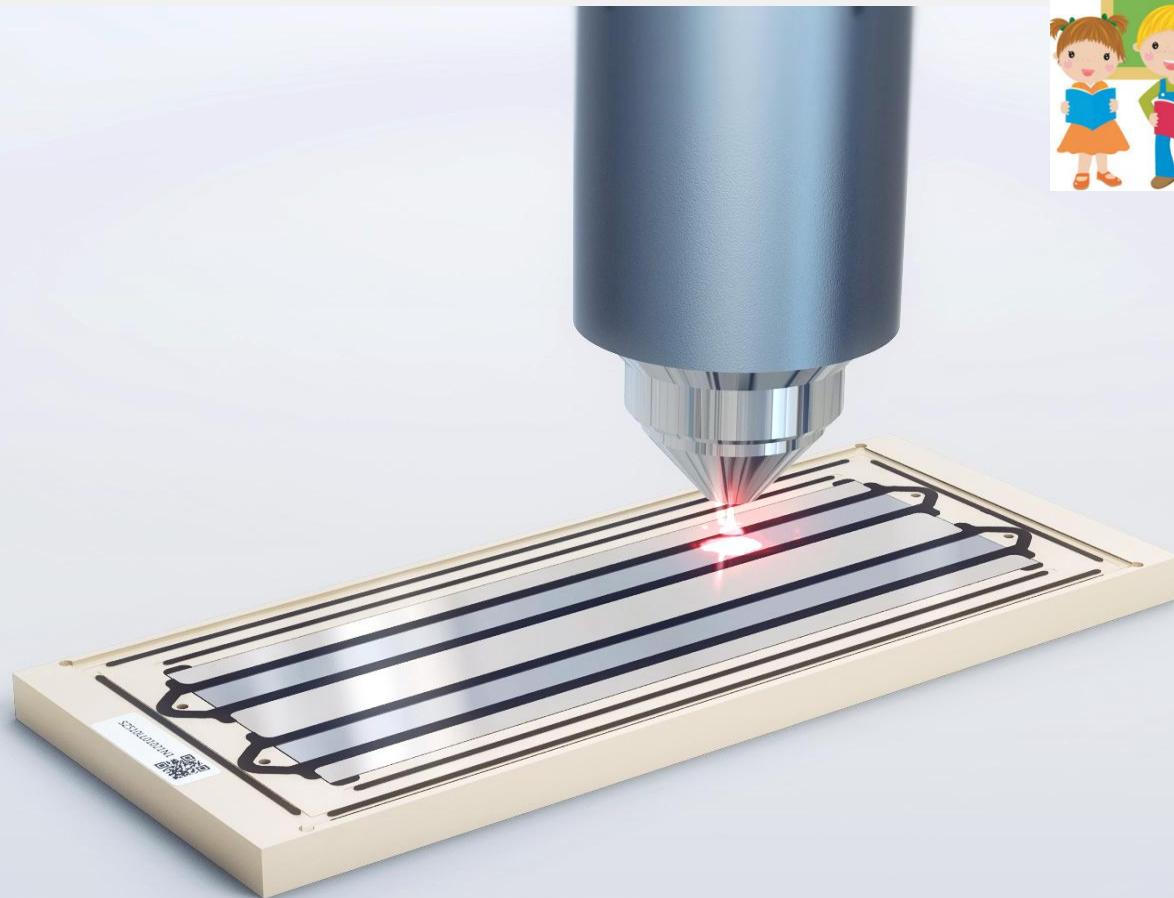




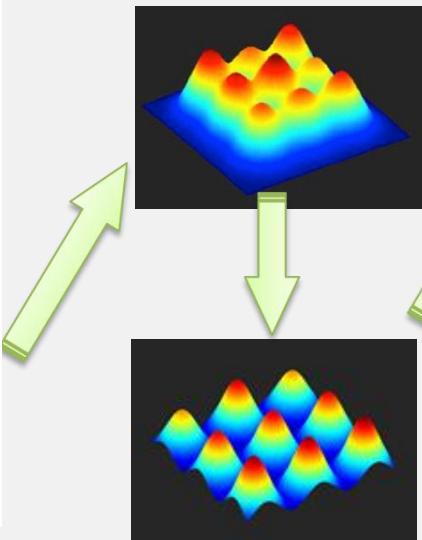
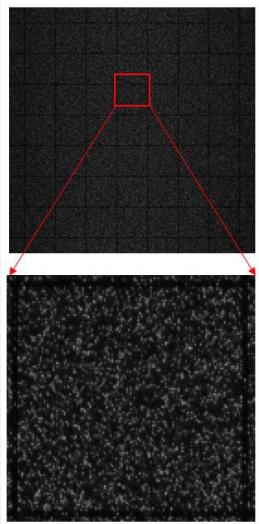
CoolIMPS: A novel antibody-based sequencing chemistry



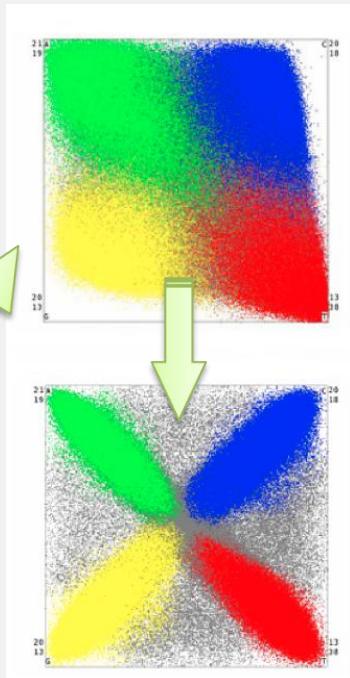
<https://www.biorxiv.org/content/10.1101/2020.02.19.953307v1>



测序芯片装载信号分析



荧光显微镜
拍摄图片



碱基及其测序质
量值识别结果

```

ATTTAAATGAGGGTATGGCTGGGCCAAACTCAGCTTAACTTCAGGCTTGAGA = CCAATTCCA
TCGAACCMGGTAGATTTTGCTCGAACCGTACCGTAGATATGAAAAGNACTCAAGTA = CAATAGTC
GAGATAGGGGGCTCTTGCTNCTGTCGCCCTTGTCCCACCCCGAAATNGCTGAGATA = ATGGCAGTT
GGGTGATACAACAAATGACTNCAGGGATGATGGCCAGTGGGCTGCTACNTGCCCCC = AAATATGAGC
CAAGAGAAGTAGTTACANGAAAAAGANTCACCAAAGTAGTGTAAATNGAAGGGAAA = ATGGATCAT
GCTGGGACTACTACGGGCACNGCTGCAACCCAGAGATCACCACTGGGNAGAGGAGG = ATGGATCAT
CTCTCTCACTCACATGAANNTTTAACATAGAACCTCTCTGTATGGNCCTGTGTTA = CTAGACTTC

0.435, 0.084, 0.445, 0.78, 0.455, 0.145, 0.455, 0.5, 0.405, 0.655, 0.14, 0.19, 0.385, 0.044, 0.775, 0.
46, 0.83, 0.58, 0.695, NaN, 0.635, 0.31, 0.39, 0.345, 0.245, 0.335, 0.084, 0.41, 0.73, 0.855, 0.48, 0.8
95, 0.465, 0.3, 0.64, 0.11, 0.545, 0.165, 0.18, 0.695, 0.525, 0.22, 0.28, 0.375, 0.29, 0.265, 0.115, 0.
385, 0.33, NaN, 0.66, 0.34, 0.285, 0.695, 0.905, 0.32, 0.31, 0.645, 0.295, 0.435 -
0.945, 0.57, 0.365, 0.36, 0.5, 0.19, 0.16, 0.275, 0.36, 0.079
0.185, 0.3, 0.6, 0.365, 0.335, 0.049, 0.044, 0.044, 0.24, 0.069, 0.11, 0.33, 0.215, 0.36, 0.064, 0.03
9, 0.029, 0.044, 0.059, NaN, 0.029, 0.029, 0.16, 0.13, 0.245, 0.125, 0.11, 0.27, 0.285, 0.27, 0.21, 0.3
2, 0.175, 0.37, 0.265, 0.19, 0.155, 0.059, 0.69, 0.069, 0.094, 0.205, 0.31, 0.23, 0.45, 0.365, 0.36, 0.
495, 0.18, NaN, 0.089, 0.765, 0.305, 0.45, 0.355, 0.285, 0.365, 0.485, 0.255, 0.285 -
0.58, 0.61, 0.059, 0.23, 0.11, 0.099, 0.099, 0.074, 0.3, 0.25
0.99, 0.76, 0.87, 0.375, 0.51, 0.285, 0.78, 0.76, 0.63, 0.699, 0.305, 0.089, 0.235, 0.385, 0.33, 0.37,
0.069, 0.18, 0.145, NaN, 0.655, 0.56, 0.395, 0.425, 0.315, 0.31, 0.53, 0.392, 0.28, 0.125, 0.044, 0.38
, 0.255, 0.054, 0.26, 0.225, 0.039, 0.059, 0.14, 0.084, 0.034, 0.26, 0.029, 0.28, 0.029, 0.079, 0.395,
0.745, 0.36, NaN, 0.47, 0.915, 0.94, 0.99, 0.8, 0.695, 0.65, 0.505, 0.695, 0.785 -
0.955, 0.95, 0.79, 0.78, 0.615, 0.81, 0.22, 0.625, 0.77, 0.965

```

Coming of age: ten years of next-generation sequencing technologies

Sara Goodwin¹, John D. McPherson² and W. Richard McCombie¹

NATURE REVIEWS | GENETICS

VOLUME 17 | JUNE 2016 | 333

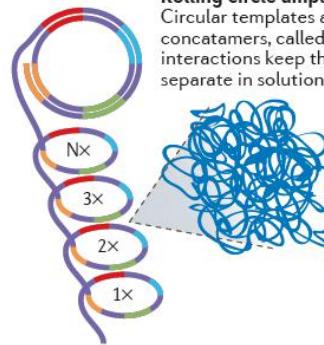
Platform	Read length (bp)	Throughput	Reads	Runtime	Error profile	Instrument cost (US\$)	Cost per Gb (US\$, approx.)
BGISEQ-500 FCS ¹⁵⁵	50–100 (SE/PE)*	8–40 Gb*	NA [†]	24 h*	≤0.1%, AT bias [‡]	\$250 (REF. 155)	NA [†]
BGISEQ-500 FCL ¹⁵⁵	50–100 (SE/PE)*	40–200 Gb*	NA [†]	24 h*	≤0.1%, AT bias [‡]	\$250,000 (REF. 155)	NA [†]

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



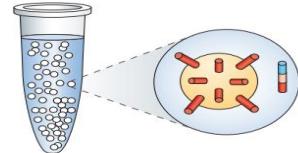
Rolling circle amplification

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

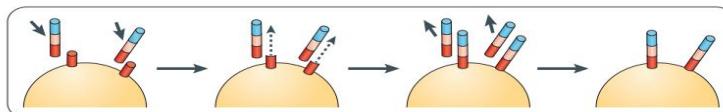
Emulsion PCR

(454 (Roche), SOLiD (Thermo Fisher), GeneReader (Qiagen), Ion Torrent (Thermo Fisher))



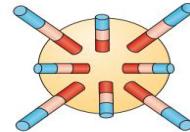
Emulsion

Micelle droplets are loaded with primer, template, dNTPs and polymerase



On-bead amplification

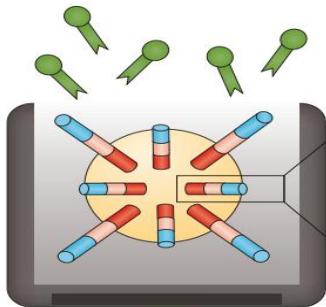
Templates hybridize to bead-bound primers and are amplified; after amplification, the complement strand disassociates, leaving bead-bound ssDNA templates



Final product

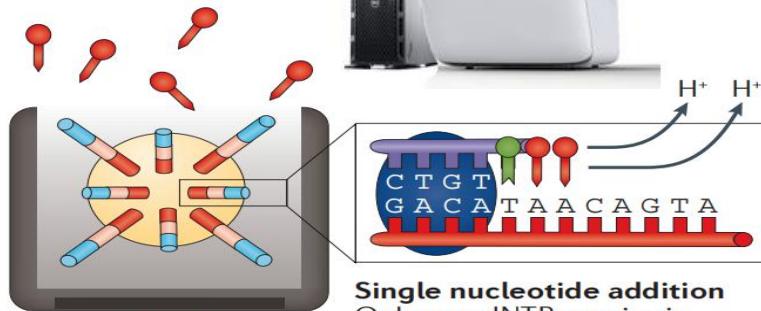
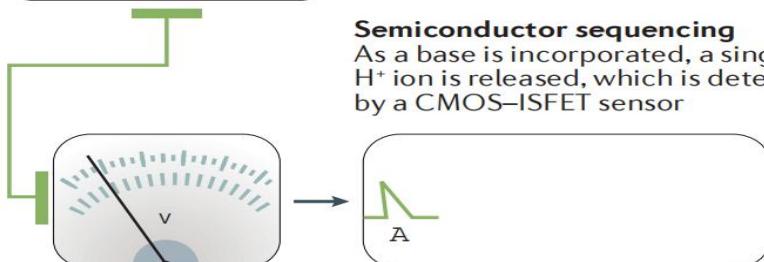
100–200 million beads with thousands of bound template

b Ion Torrent (Thermo Fisher)



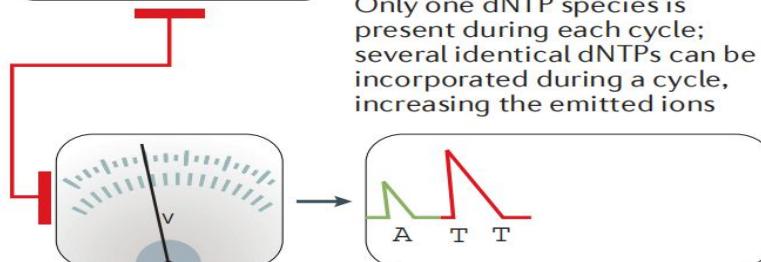
Semiconductor sequencing

As a base is incorporated, a single H^+ ion is released, which is detected by a CMOS-ISFET sensor

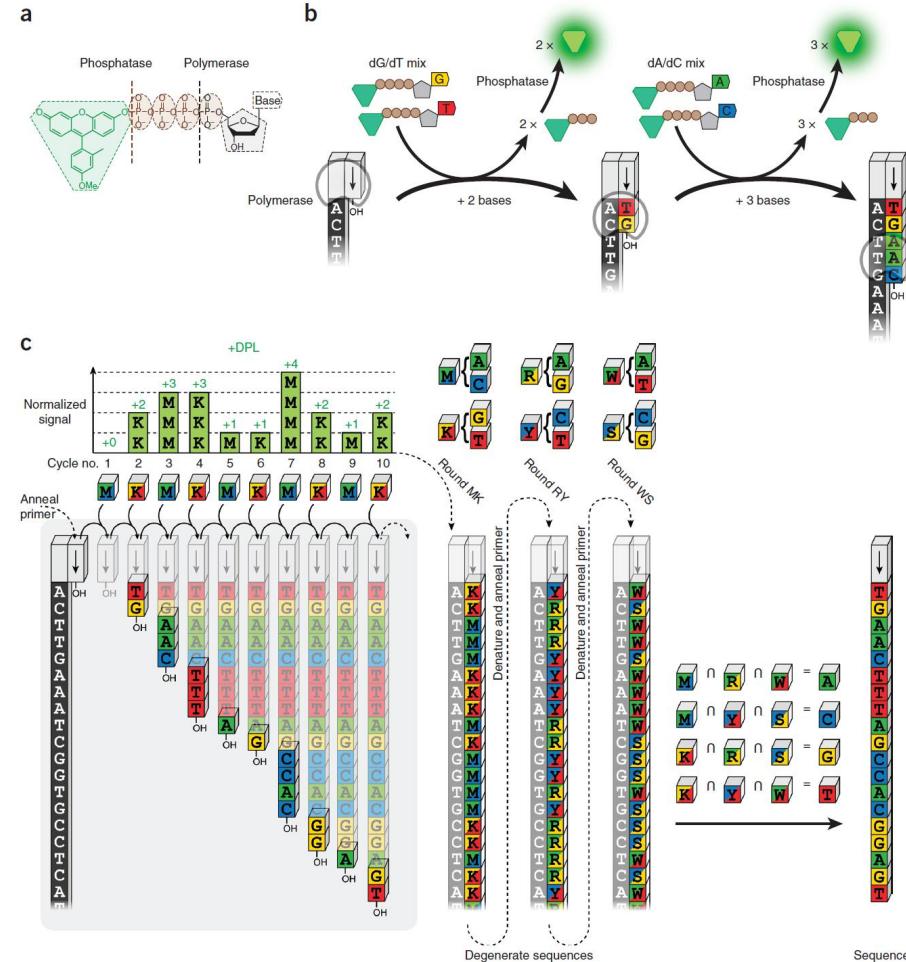


Single nucleotide addition

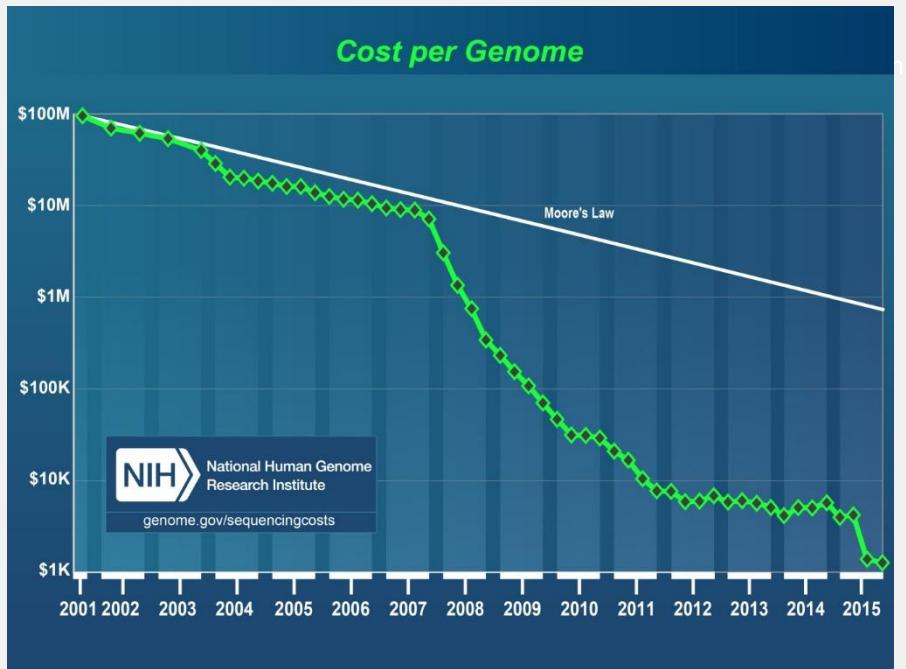
Only one dNTP species is present during each cycle; several identical dNTPs can be incorporated during a cycle, increasing the emitted ions



Error-correction code sequencing



Massive parallel sequencing VS single molecule real time sequencing



Easy to use

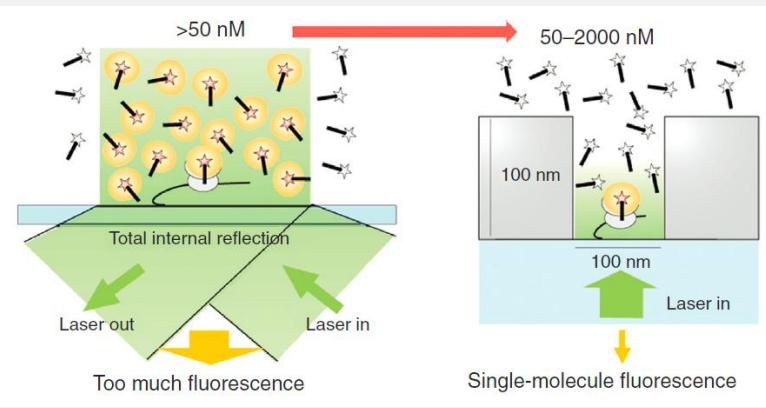
Signal amplification needs
extra library preparation



Fast
Long reads

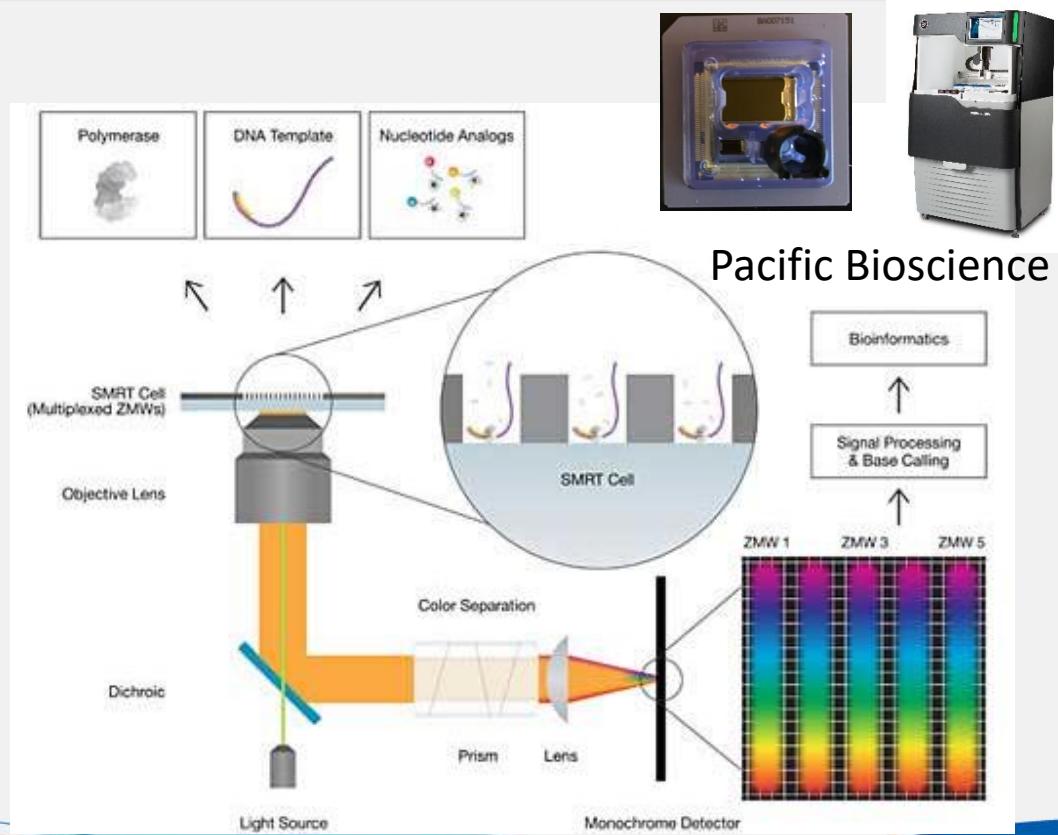
Synchronization slows
down and can not last long

Single polymerase, optical read out



通过零模波导（ZMW）限制激发光范围，获得单分子信号。

31 JANUARY 2003 VOL 299 SCIENCE



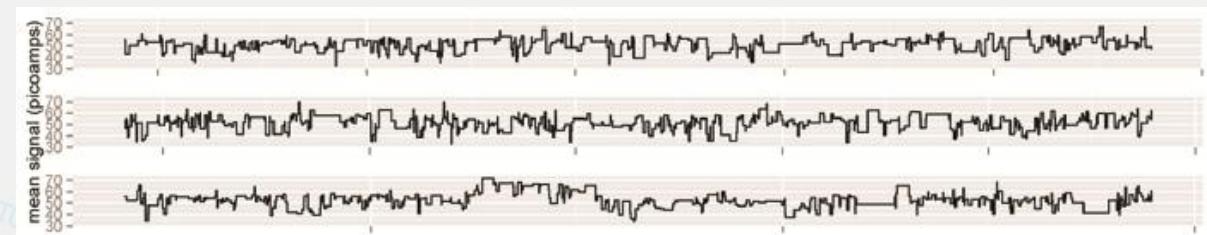
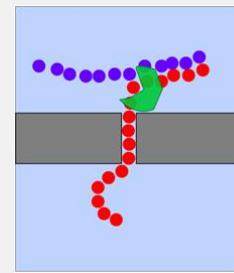
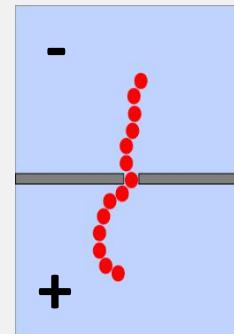
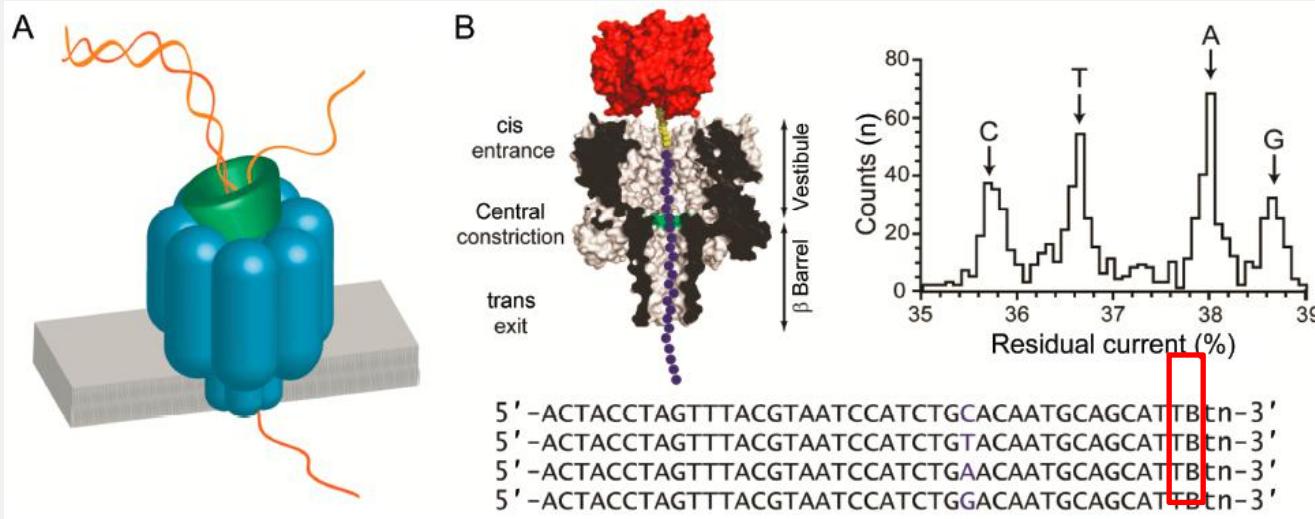
No polymerase, electric read out



Oxford
Nanopore
Technology

Ideal

Reality



读：MPS测序仪

国际进展

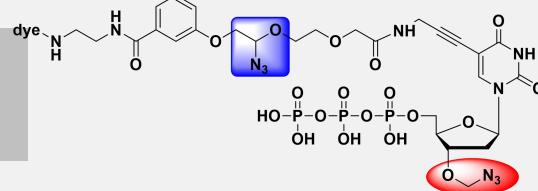
Illumina



300G/Run, 400M;
6T/Run, 8k-10k M

255台; 1228台

illumina®



GenapSys



1.2-2G/Run
10-13M

电测序

Element
Biosciences

C轮
2.76亿美金

/

SBB测序

Singular
Genomics

G4

Purpose-built, next
generation sequencer

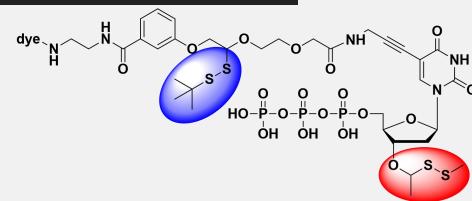
HD-Seq
SLR
Rapid
Seq
Seq



4*45Gb/Run
50M

SBS测序，单细胞

**SINGULAR
GENOMICS**



Omniome

PacBio以8亿美
元收购

180G/Run

SBB测序

Genemind



300G/Run
250-500M

SBS, 单分子

**Element
biosciences**

Fapon



120G/Run
100-400M



Molly He, PhD
CEO, Co-Founder & Board Member



Michael Previde,
PhD
CTO & Co-Founder

关键指标	参数
聚合时间	60s
切除时间	90s
PE100全流程	46h
Q30	95%

2021.5美股上市，
市值17.5亿美元，
真实仪器参数不详

累计超过4亿美元
融资， illumina
原班人马，技术路
线与 illumina类似

读：单分子测序



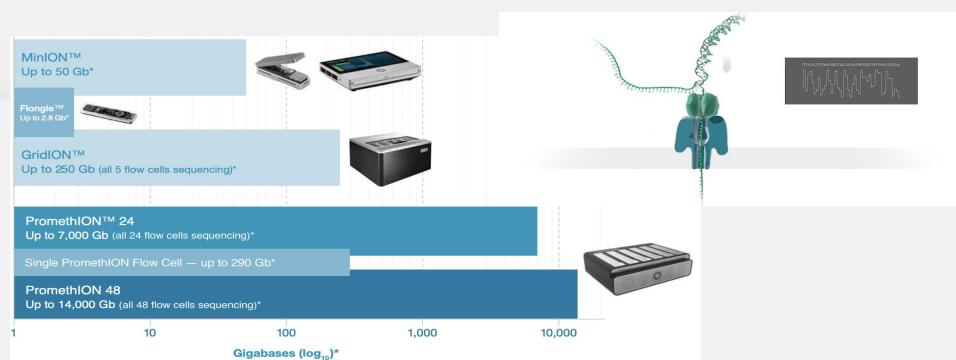
预计下半年IPO上市，当前估值 £2.48B

融资共计 £994.6M

读长4.15Mb，准确率99.3% (Q20)，通量10Tb (P48)

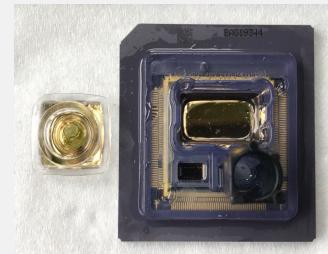
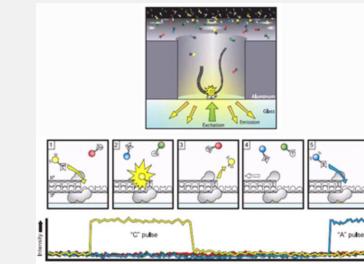
发布新冠即时检测解决方案 LamPORE

登陆伦敦证交所，市值一度接近50亿欧元



市值 \$6.22B 股价 \$31.33

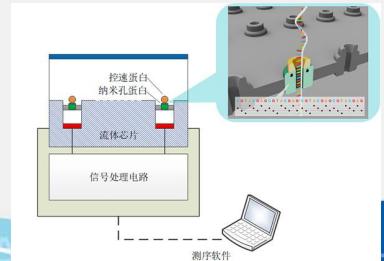
读长20kb，准确率99%，通量 4M reads (CCS模式)
\$800M收购Omnionome，形成长短读长测序技术整合



2021年融资 ￥400M，融资共计 ￥543M

2020年09月发布第一款产品QNome-9604，读长150kb，
通量500Mb，准确率90%

GMP生产基地2021年9月底在成都建成，总投资 ￥200M，
总面积 4000m² QNome-9604



展望

1. 测序通量，成本将不断逼近理论极限，千元人民币基因组时代来临。
2. 样品处理与计算存储面临更大挑战。
3. 小型化，便携式，集成样品处理的测序方案推广。
4. 单分子测序走向成熟。
5. 国内测序仪研发热潮来临。



Thanks!