高通量测序技术及其原理

刘赟
复旦大学基础医学院
2024. 9. 20

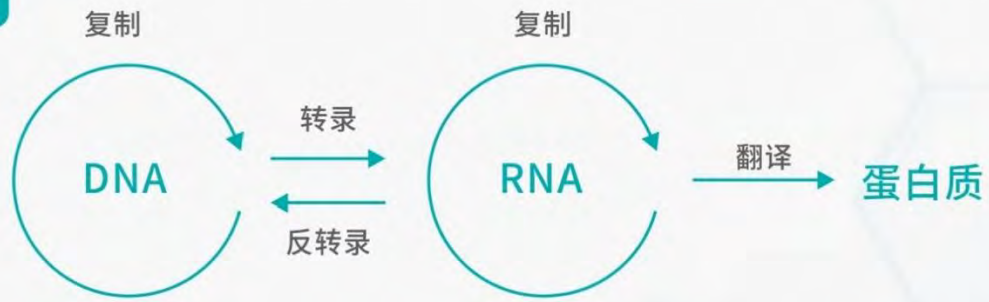
目录

01 测序技术发展史

生命从何而来？

生命之源，源于基因，基因遵循**生命中心法则**，与内外环境互动，使生命在时空中繁衍与传承

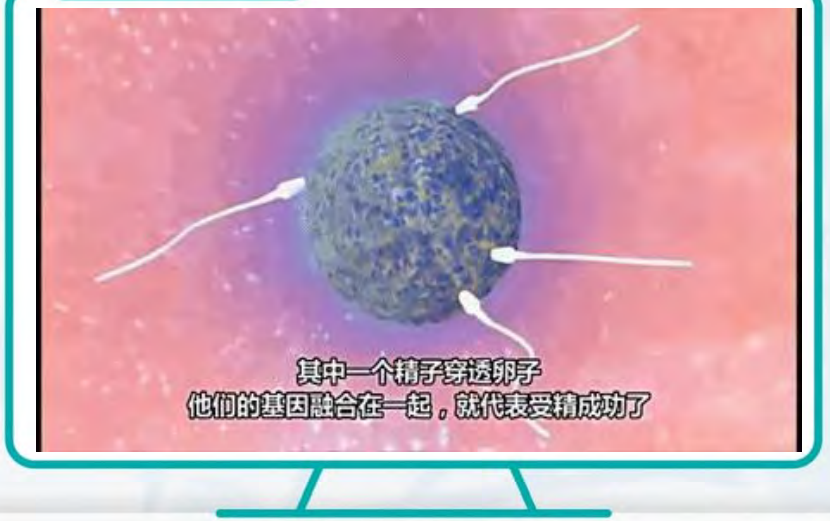
生命中心法则

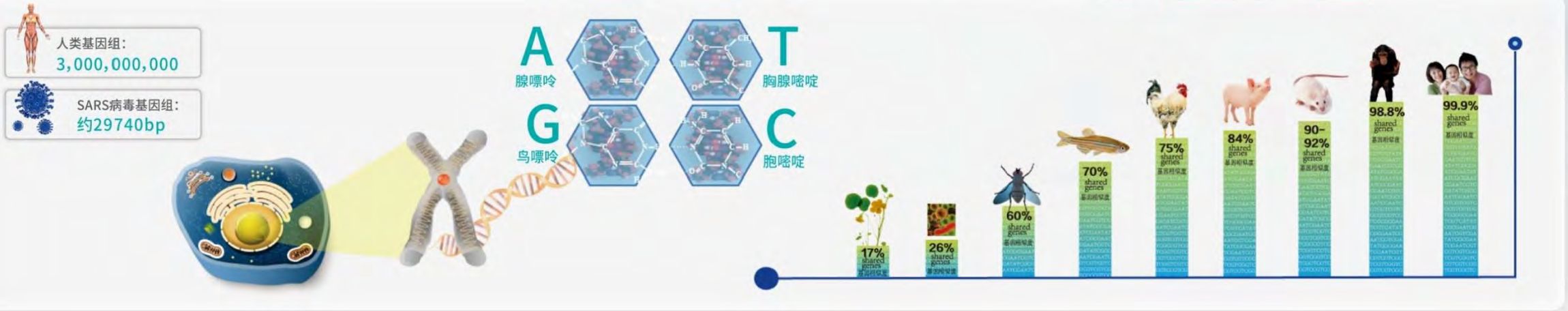


基因由ATCG组成，**生命**由简到繁，**天书**由小到大

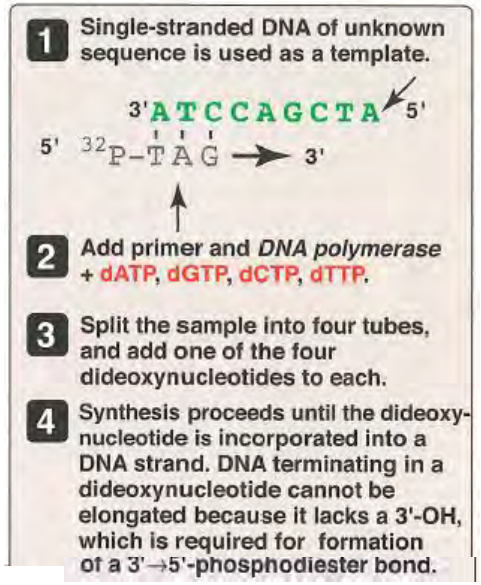
SARS病毒与人类：只是ATCG排序和多少不同

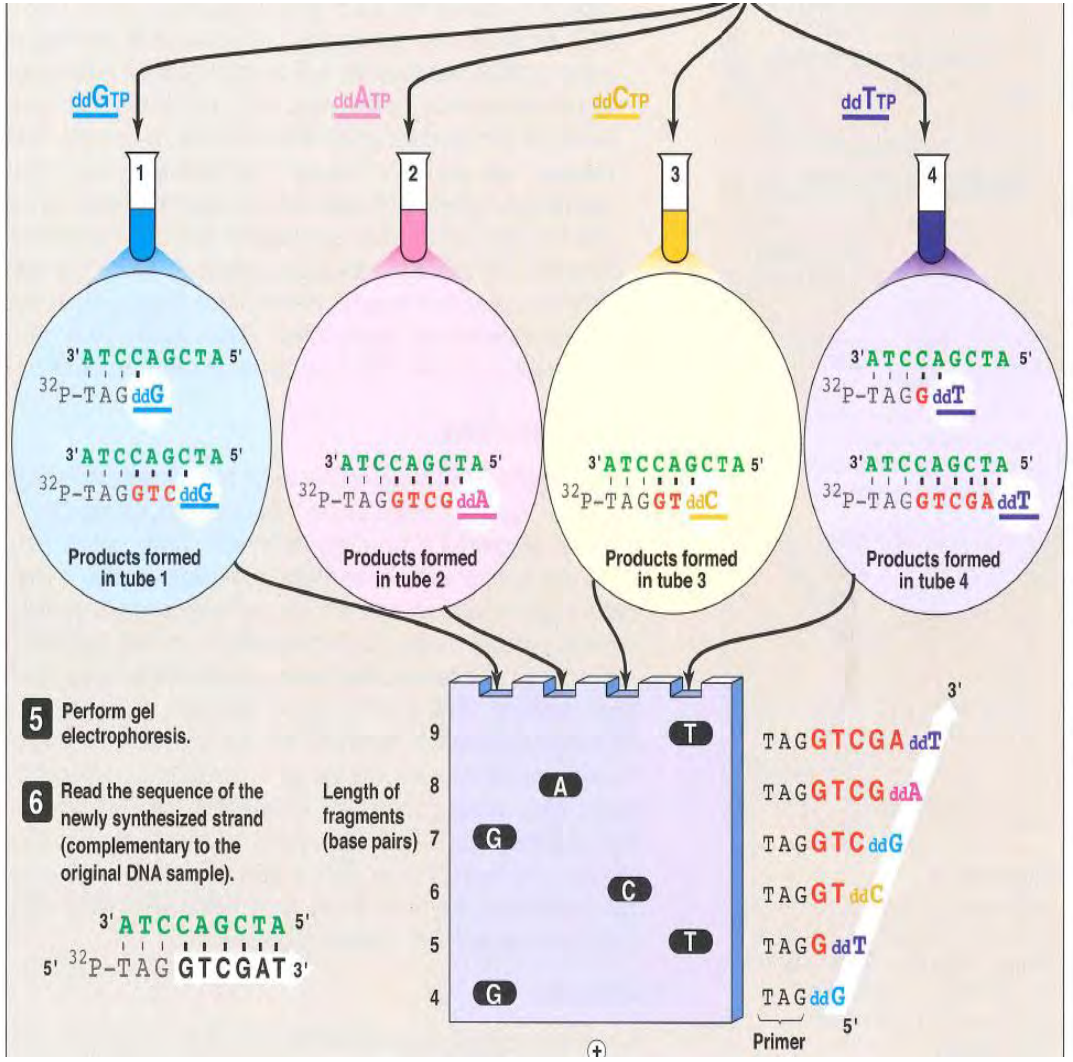
生命的形成



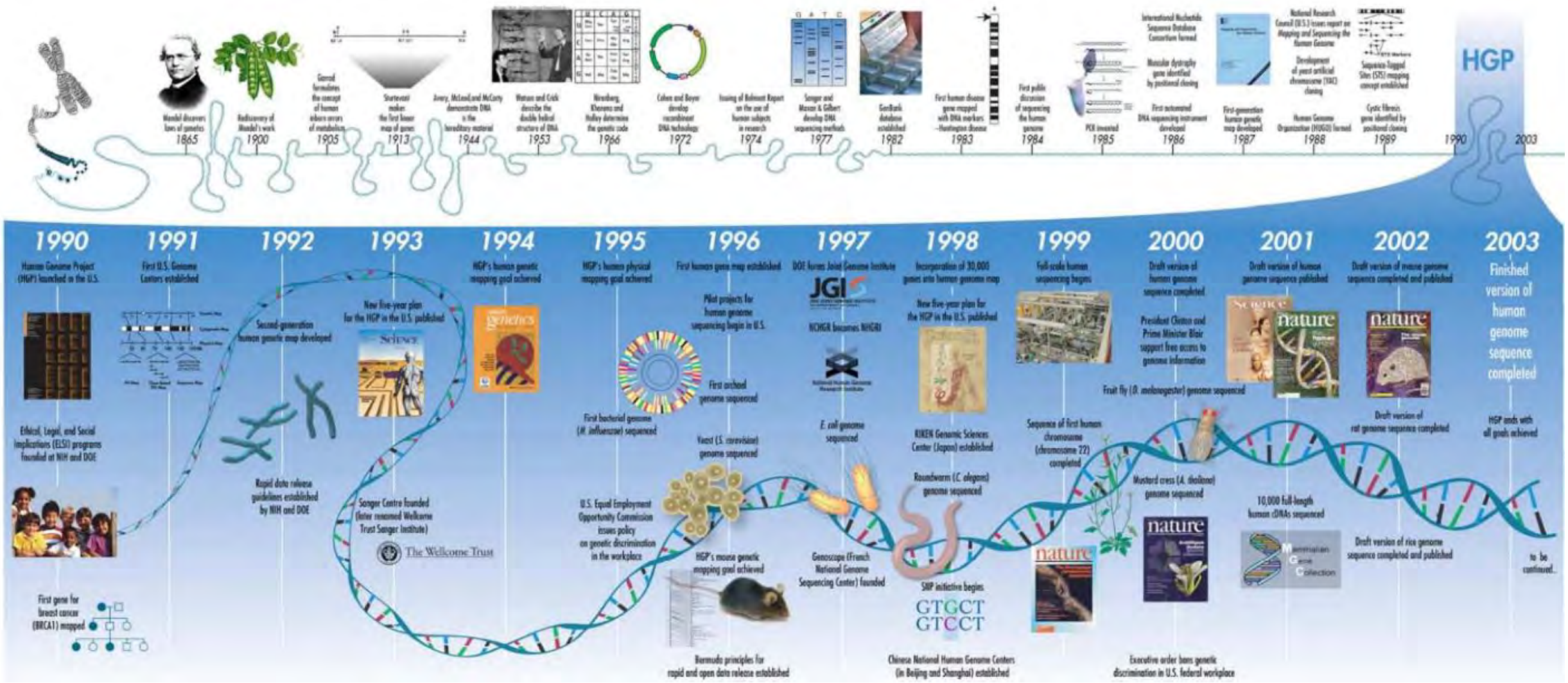


一代测序技术

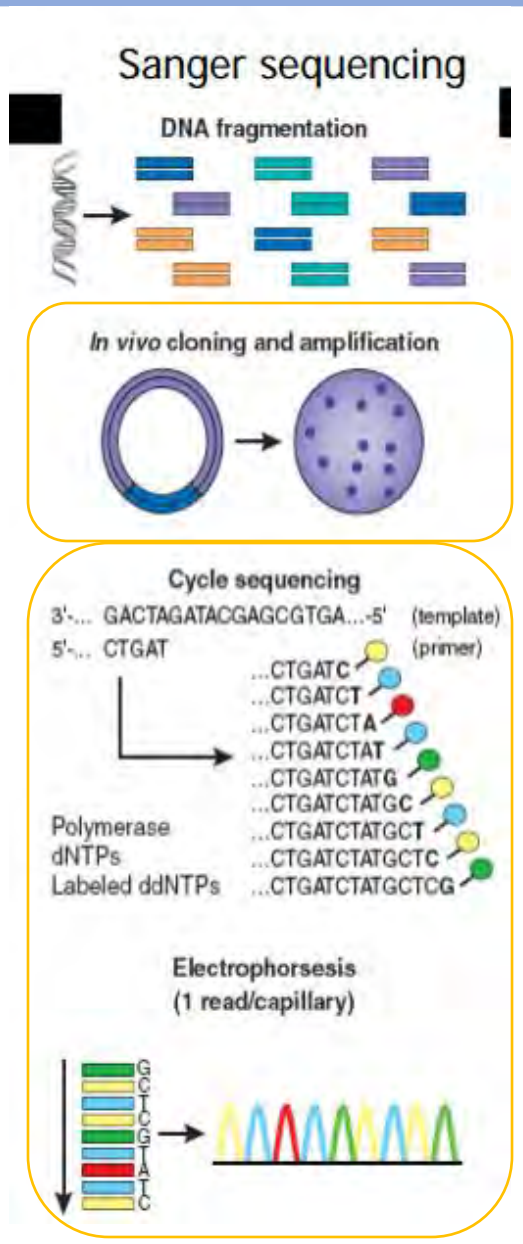

1 Single-stranded DNA of unknown sequence is used as a template.
3' ATCCAGCTA 5'
5' 32P-TAG → 3'
2 Add primer and DNA polymerase + dATP, dGTP, dCTP, dTTP.
3 Split the sample into four tubes, and add one of the four dideoxynucleotides to each.
4 Synthesis proceeds until the dideoxynucleotide is incorporated into a DNA strand. DNA terminating in a dideoxynucleotide cannot be elongated because it lacks a 3'-OH, which is required for formation of a 3'→5'-phosphodiester bond.

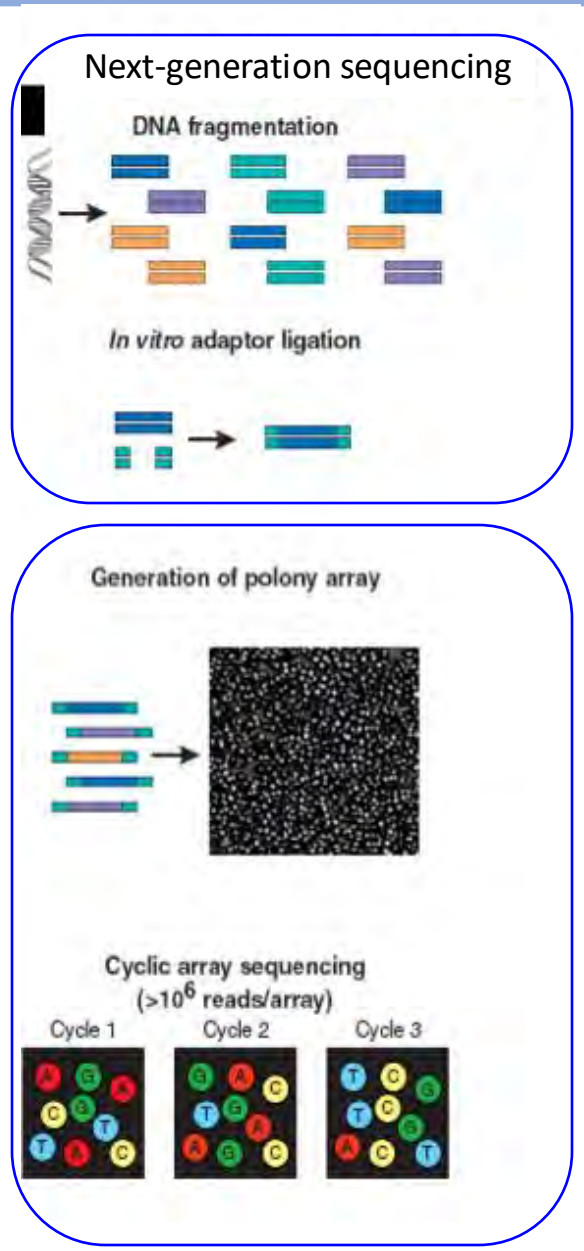

ddGTP
ddATP
ddCTP
ddTTP
1
2
3
4
3'ATCCAGCTA 5'
32P-TAGddG
3'ATCCAGCTA 5'
32P-TAGGTCddG
Products formed in tube 1
3'ATCCAGCTA 5'
32P-TAGGTCGddA
Products formed in tube 2
3'ATCCAGCTA 5'
32P-TAGGTddC
Products formed in tube 3
3'ATCCAGCTA 5'
32P-TAGGddT
3'ATCCAGCTA 5'
32P-TAGGTCGAddT
Products formed in tube 4
5 Perform gel electrophoresis.
6 Read the sequence of the newly synthesized strand (complementary to the original DNA sample).
3'ATCCAGCTA 5'
5' 32P-TAG GTCGAT 3'
Length of fragments (base pairs)
9
8
7
6
5
4
T
A
G
C
T
G
3'
TAGGTCGAddT
TAGGTCGddA
TAGGTCddG
TAGGTddC
TAGGddT
TAGddG
Primer
5'

The Human Genome Project



二代测序技术





Advantages of NGS

- Construction of a sequencing library -> clonal amplification to generate sequencing features

  No in vivo cloning, transformation, colony picking…

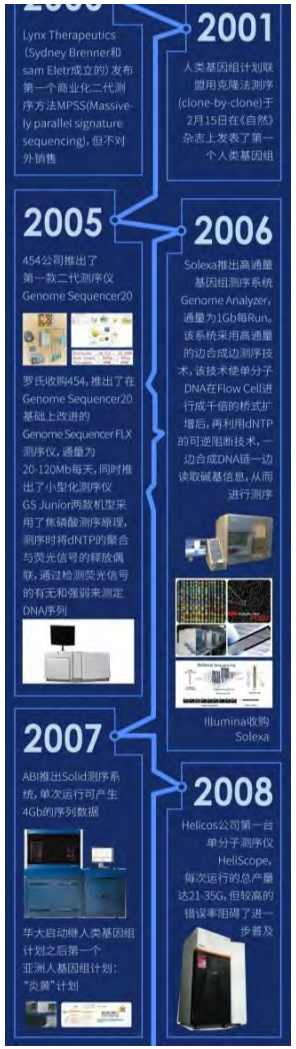
- Array-based sequencing

  Higher degree of parallelism than capillary-based sequencing

02 高通量测序技术原理

不同的二代测序平台

2006-2007阶段旧三足鼎立：
罗氏收购454、ABI推出Solid、Illumina收购Solexa

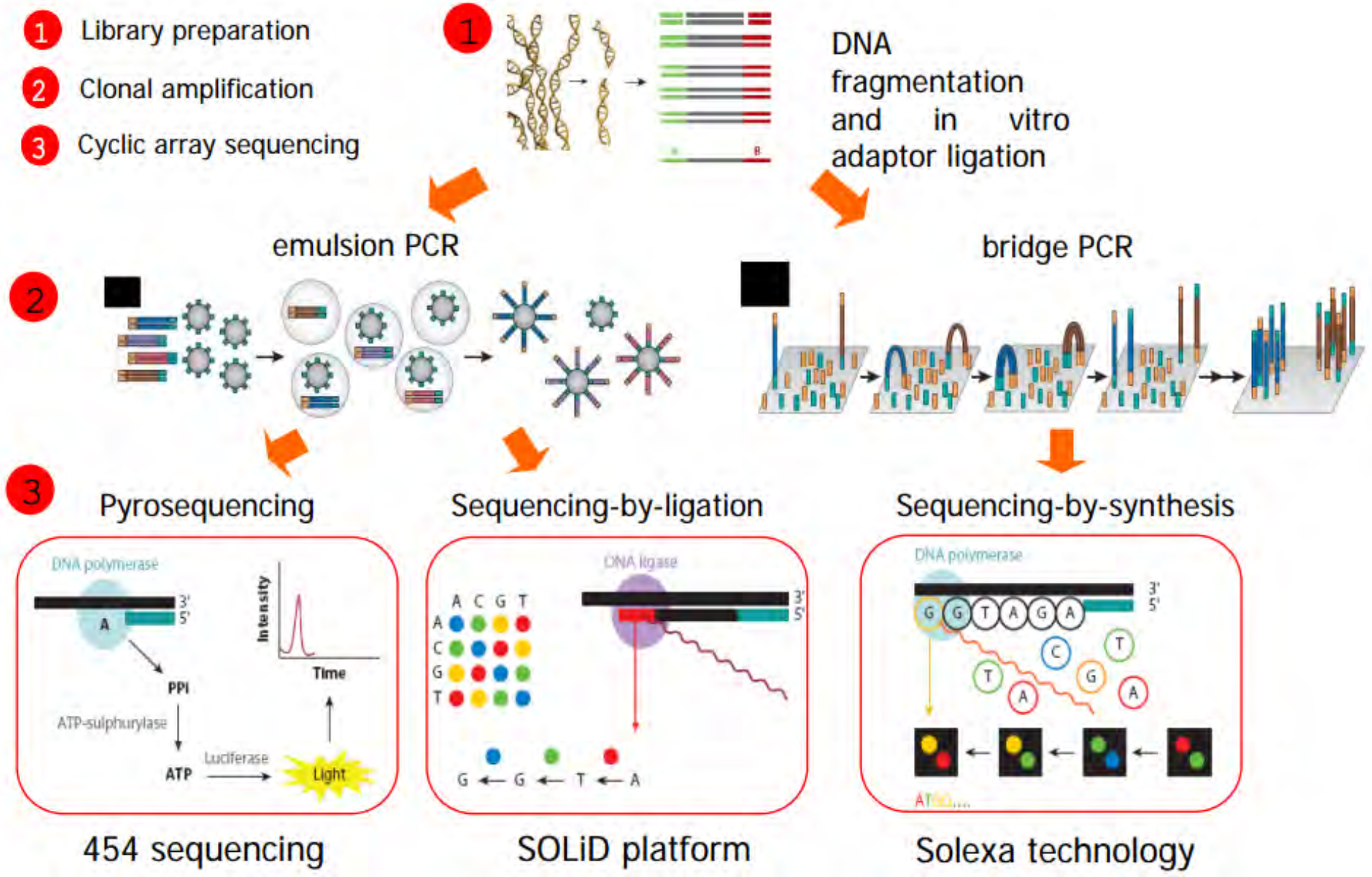


2013~2015阶段
新三足鼎立：

- 赛默飞收购Life
- 华大收购CG后推出MGI
- Illumina推出NextSeq等

Illumina测序平台

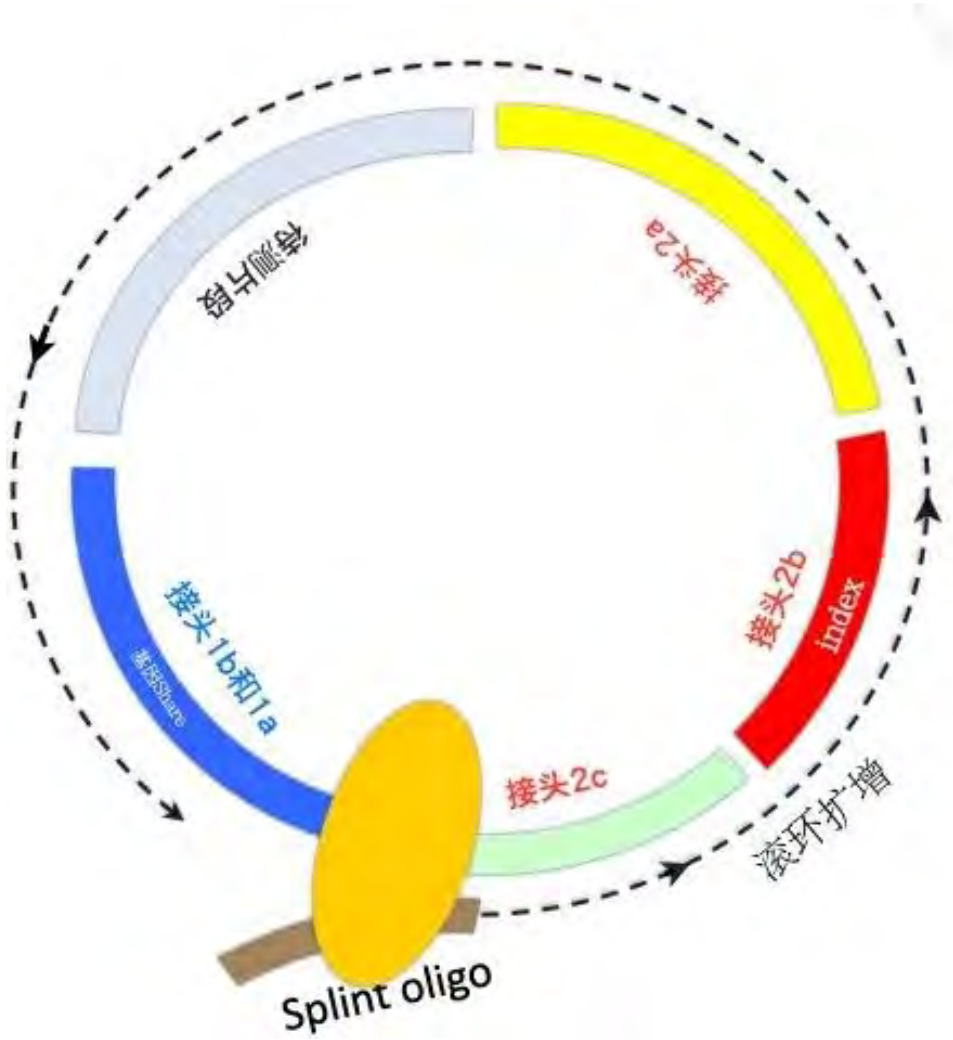
不同的二代测序平台


1 Library preparation
2 Clonal amplification
3 Cyclic array sequencing
1
DNA fragmentation and in vitro adaptor ligation
emulsion PCR
bridge PCR
2
3
Pyrosequencing
Sequencing-by-ligation
Sequencing-by-synthesis
DNA polymerase
A
PPi
ATP-sulphurylase
ATP
Luciferase
Light
Intensity
Time
DNA ligase
A C G T
G G T A
DNA polymerase
G G T A G A
454 sequencing
SOLiD platform
Solexa technology

华大DNB测序技术

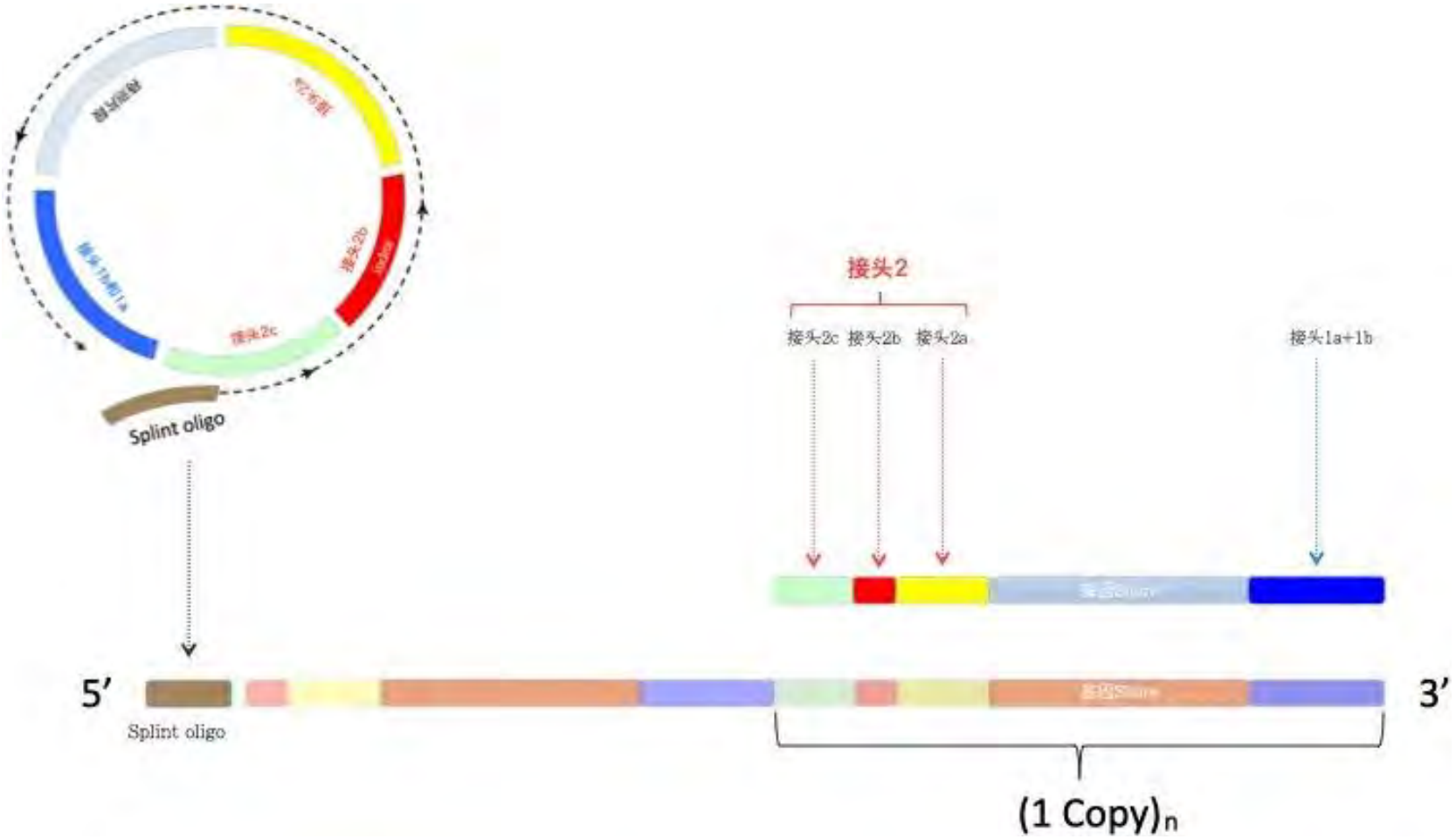
a) 纳米球制备：

荧光信号采集单元的制备（制备DNA 纳米球）



华大DNB测序技术

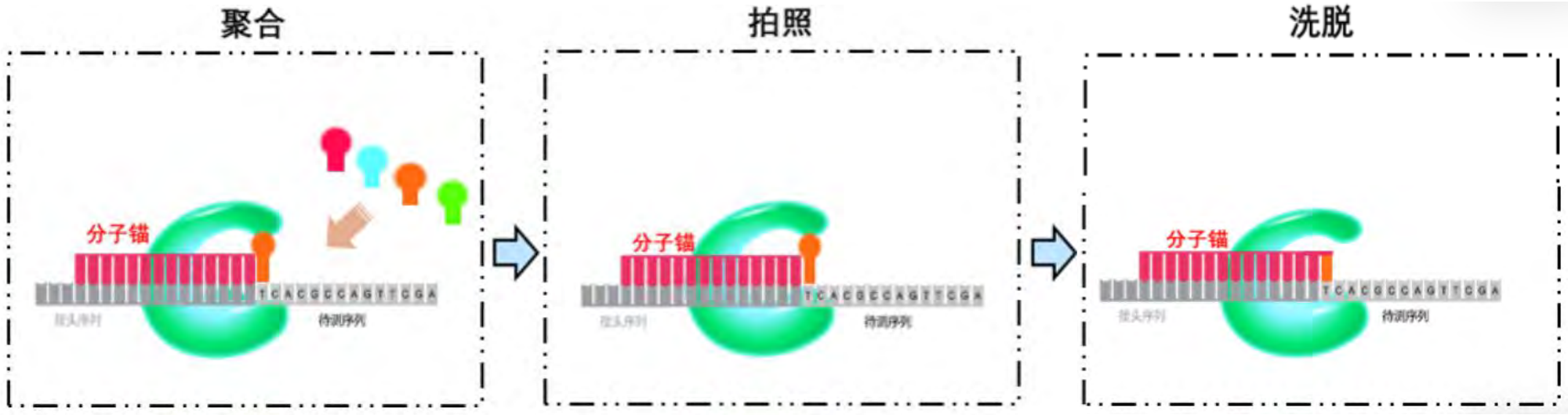
b) DNA纳米球“拉直”后即下图结构，待测片段“串联重复”



华大DNB测序技术

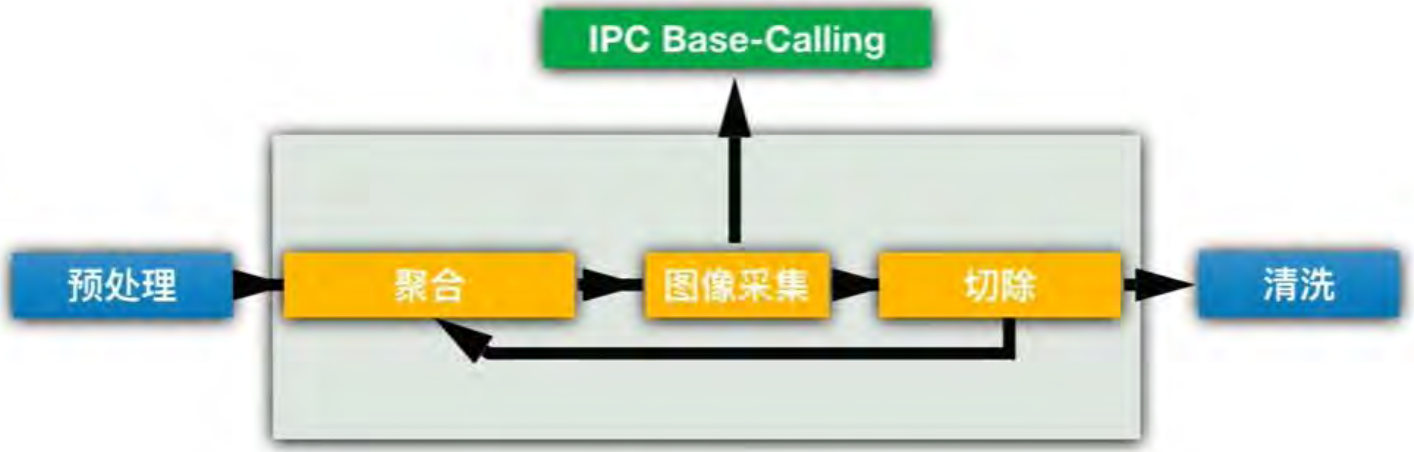
联合探针锚定聚合技术(Combinatorial Probe-Anchor Synthesis, **cPAS**)：

末端特殊修饰的碱基（AGCT) 分别被标记为不同的荧光探针，依次流过流动池，在DNA聚合酶的催化下，DNA分子锚和荧光标记核苷酸于DNB上进行聚合，洗脱掉未结合的核苷酸后，在激光的作用下荧光信号被激发，随后高分辨率成像系统对光信号进行采集，光信号经过数字化处理后，获得当前待测碱基的信息。然后加入再生洗脱试剂，去除荧光基团，进入下一个循环检测，SE50测序一共61个循环。



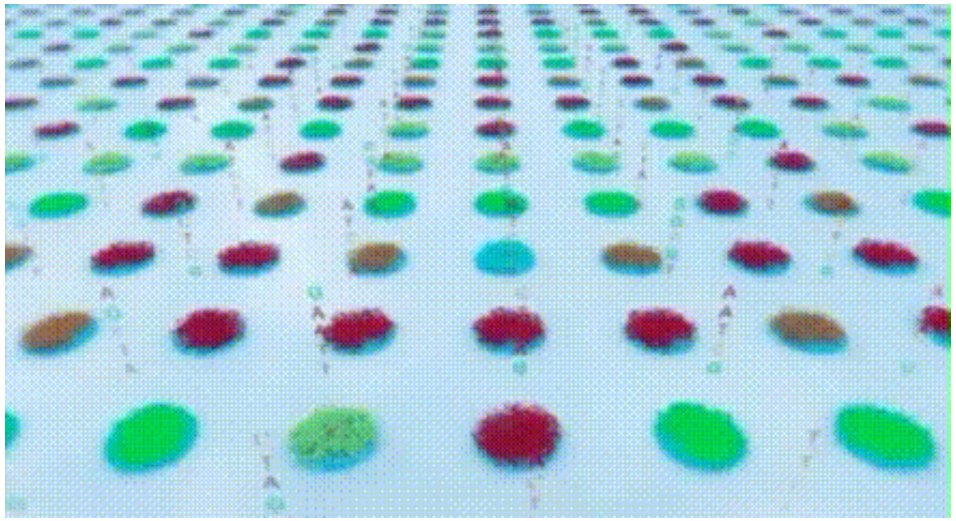
华大DNB测序技术

测序原理



联合探针锚定聚合技术
cPAS (combinatorial Probe Anchor Synthesis)

Improved from cPAL technology(Complete Genomics)



华大DNB测序技术

测序原理

双色荧光				
碱基	A	T	C	G
图像 No.1	🔴		🔴	
图像 No.2	🟢	🟢		
结果	A	T	C	G

T 碱基能被 绿色 激光激发

C 碱基能被 红色 激光激发

A 碱基能被 红色 和 绿色 激光激发

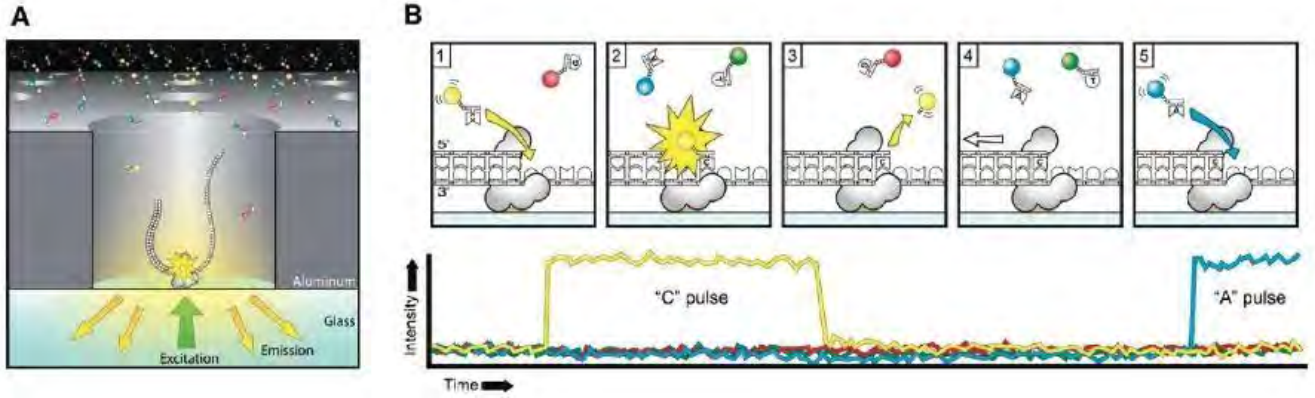
G 碱基不被激发

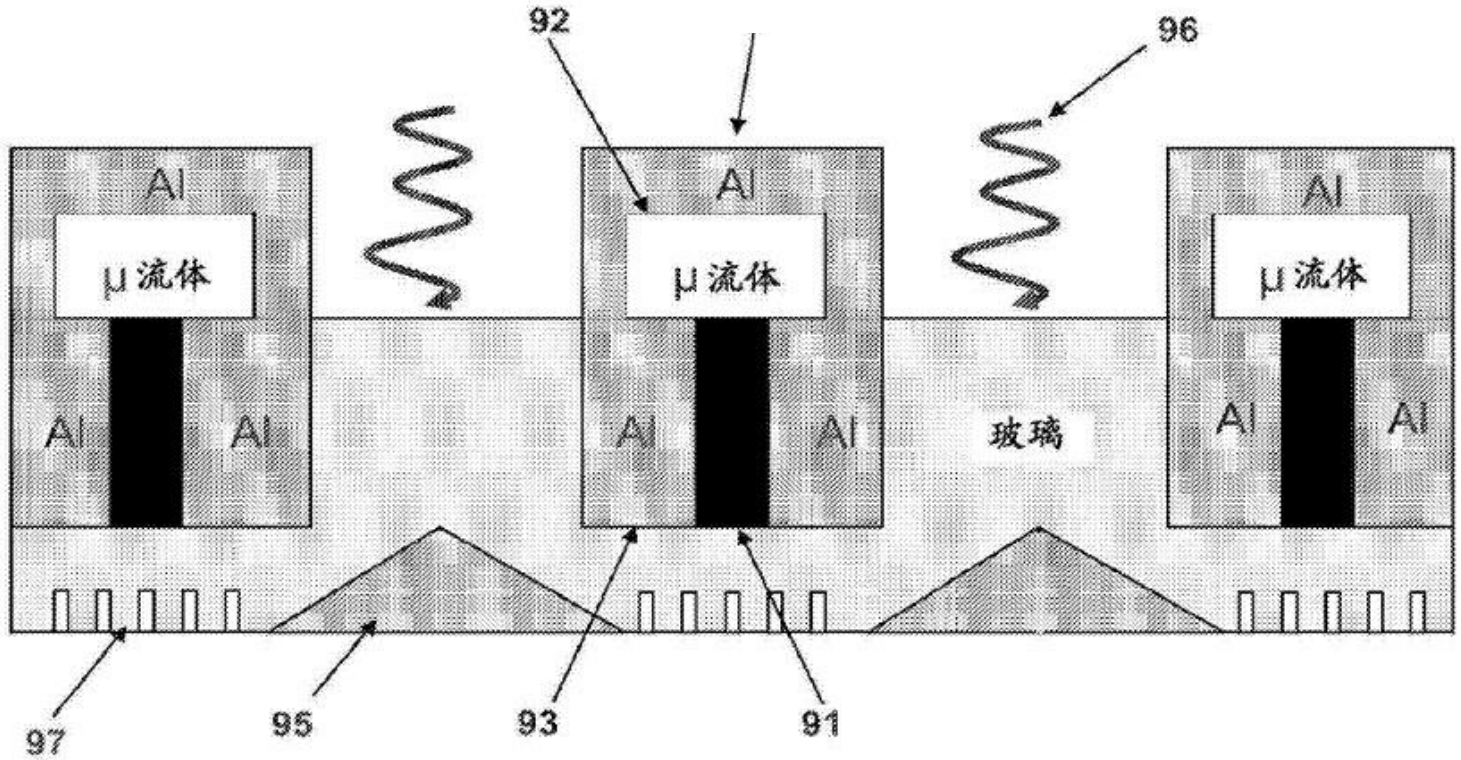
BGISEQ-50

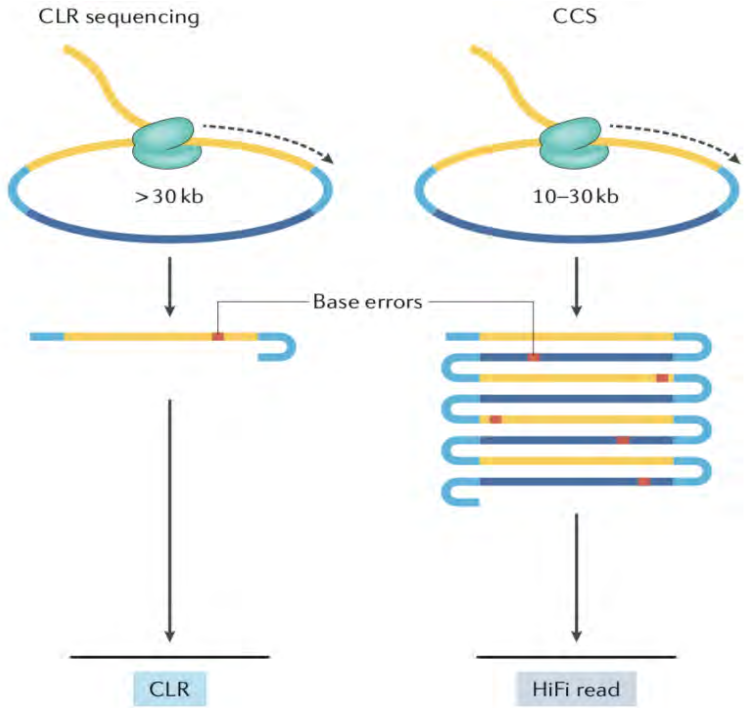
MGISEQ-200

华大DNB测序技术

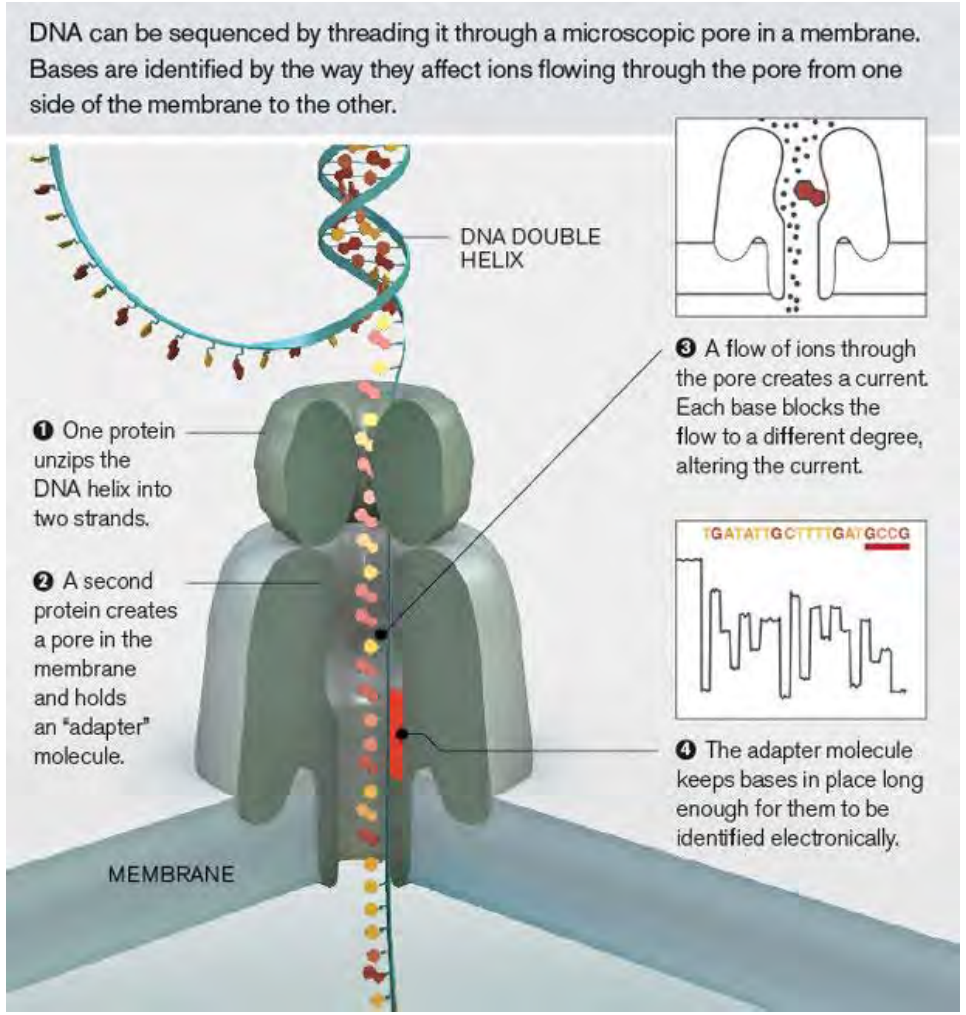
三代测序技术 - PacBio





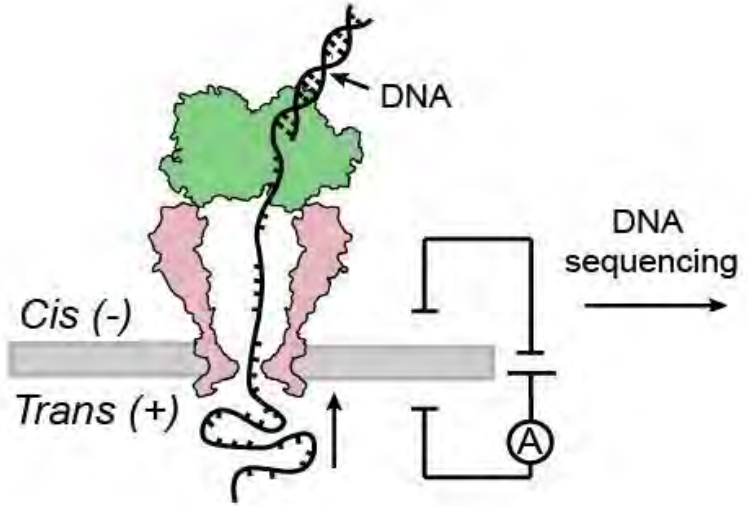


三代测序技术 - Nanopore



- 纳米孔技术起源于1996年，在一个典型的纳米孔测序实验中，纳米孔（粉色）是磷脂膜（灰色）两侧离子通过的唯一通道。

- 测序酶（绿色）充当DNA的马达蛋白，拉动DNA链使其以单个核苷酸的步长依次通过纳米孔，每当一个核苷酸穿过纳米孔，相应的堵孔信号会被记录下来。通过分析这些序列相关的电流信号，我们可以反推出DNA的序列。



不同测序方法优劣势对比

测序方法		准备方法	检测方法	主要优点	主要缺点	应用场景
一代	Sanger (毛细血管电泳)	PCR+电泳	同位素/ 荧光标记	读长较长、准确度高，重复、多聚序列测得好	通量低、成本高	司法
二代	Illumina (Solexa)	芯片+桥式PCR	荧光标记	准确度较高，通量高，成本低	读长短	临床+科研
	BGI/MGI (CG)	芯片+DNB	荧光标记	准确度较高，通量高，成本低	读长短	
	Thermo Fisher (Ion Torrent)	芯片+乳滴PCR	PH变化	时间短	读长短、成本较高、通量低	
三代	Pacific Biosciences	芯片+SMRT bell	ZMW+荧光标记	读长长(30-50 kb) 样本制备简单	成本高，文库制备复杂	科研
	Oxford Nanopore	芯片+leader-hairpin DNA	电流变化	读长长(100 kb) 样本制备简单	错误率较高，成本高	

03 高通量测序技术应用

高通量测序技术应用

Sequence DNA

- *De novo* sequencing
- Reference-based re-sequencing
  - SNP, CNV, Indels
- Metagenomics
  - Identify “who is there?” in a mixture of microbes

Sequence RNA

- RNA-Seq (transcriptome-wide sequencing)
- smRNA-Seq
- novel ncRNAs

Study Protein-DNA/RNA interaction

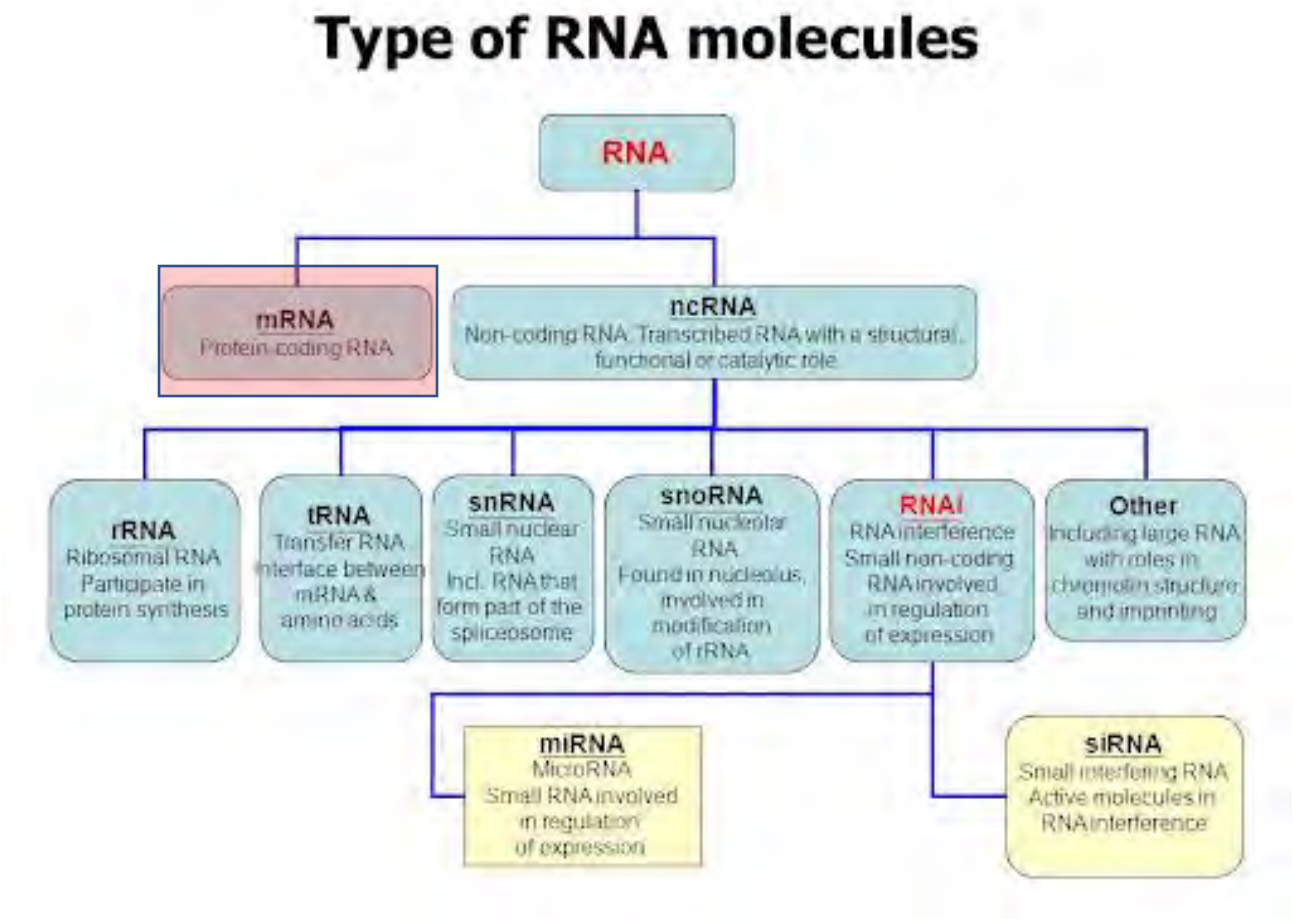
- ChIP-Seq (for TF, Pol II binding)
- CLIP-Seq (for RNA binding proteins)

Epigenetics

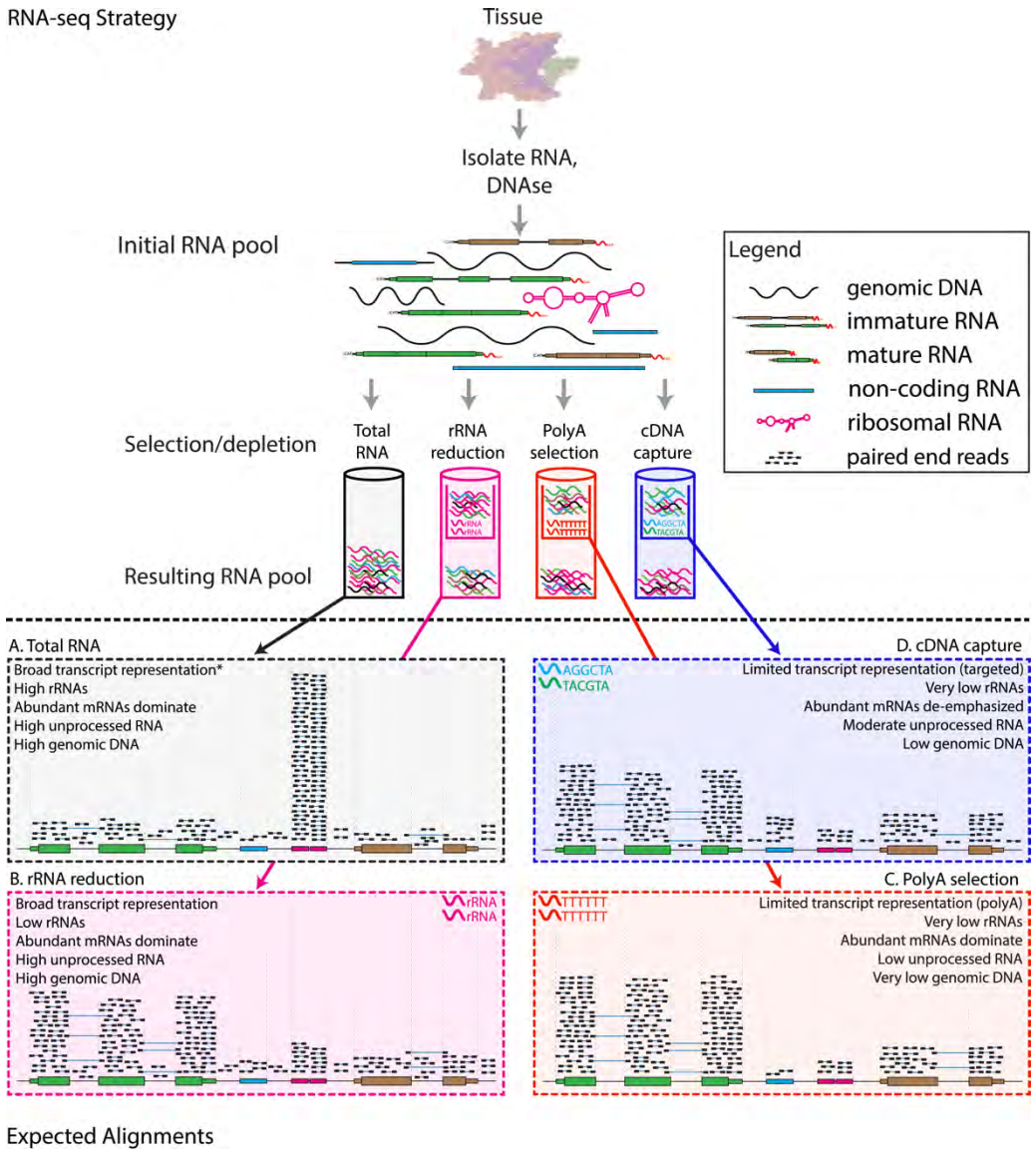
- DNA methylation
- Histone modification (ChIP-Seq)
- Nucleosome positioning
- Chromosome looping (Hi-C)

RNA-Seq

RNA sequencing (RNA-Seq) is a sequencing technique which uses next-generation sequencing (NGS) to reveal the presence and quantity of RNA in a biological sample at a given moment.

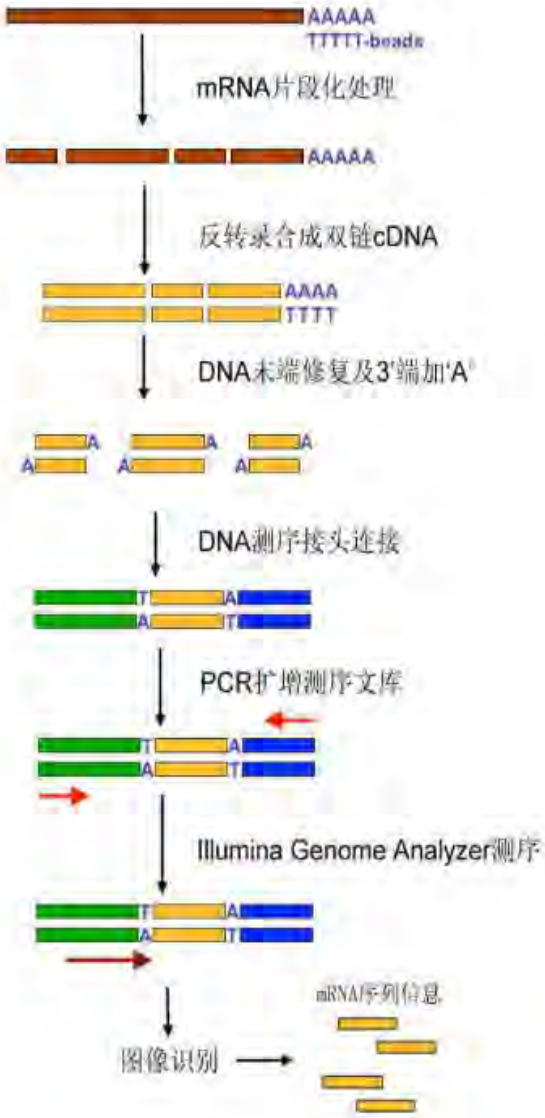


不同的RNA-Seq策略



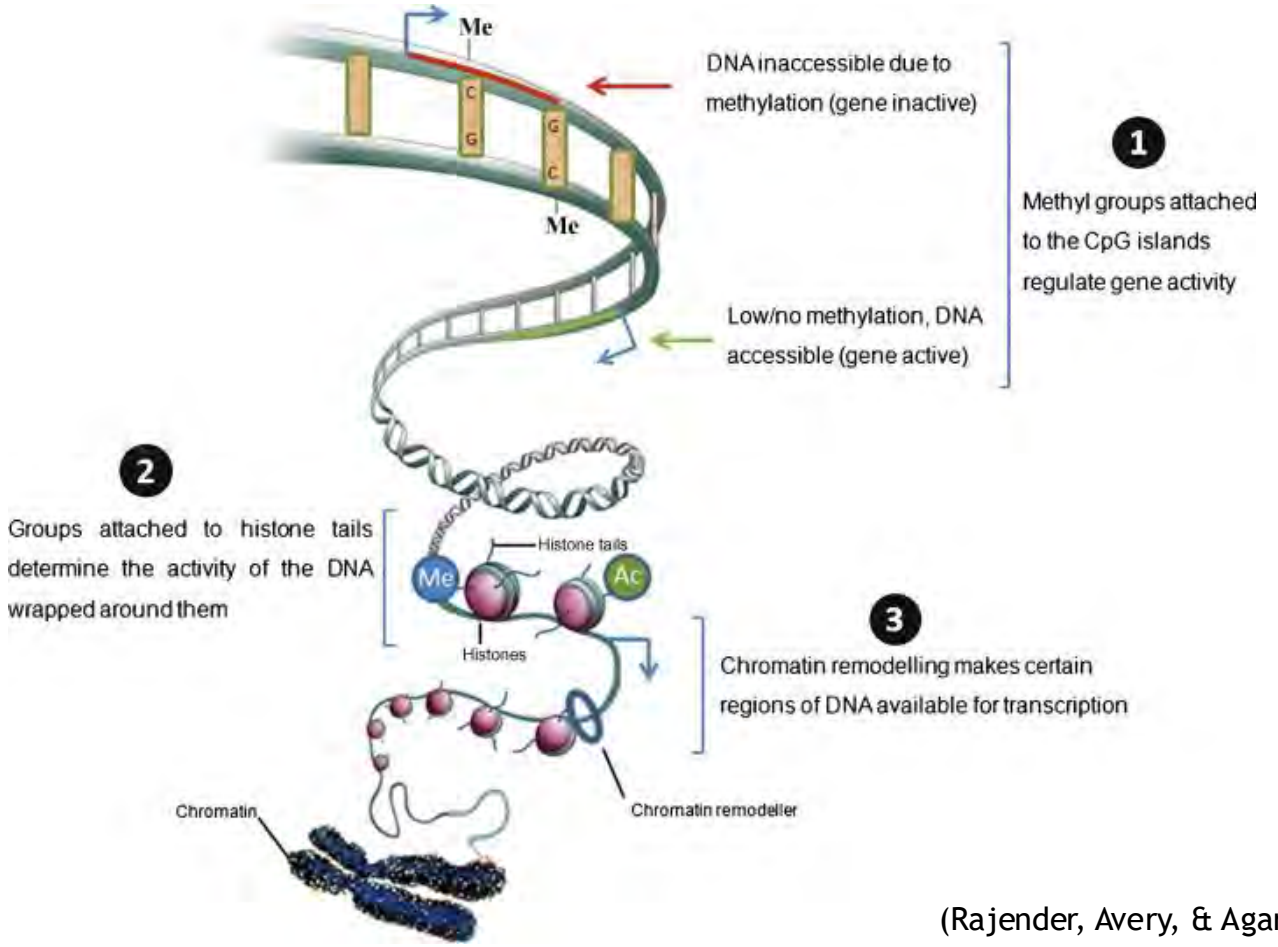
RNA-Seq for mRNA

1. 样品**RNA**准备
2. 测序文库构建
   - 使用oligo dT微珠纯化mRNA
   - mRNA片段化处理
   - 反转录反应合成合成双链cDNA
   - 双链DNA末端修复及3’末端加‘A’
   - 使用特定的测序接头连接DNA片段两端
   - 高保真聚合酶扩增构建成功的测序文库
3. **DNA**成簇（**Cluster**）扩增
4. 高通量测序（**Illumina Genome Analyzer IIx**）
5. 数据分析
   - 原始数据读取
   - 与数据库比对并进行注释
   - 深层次数据分析



Epigenomics

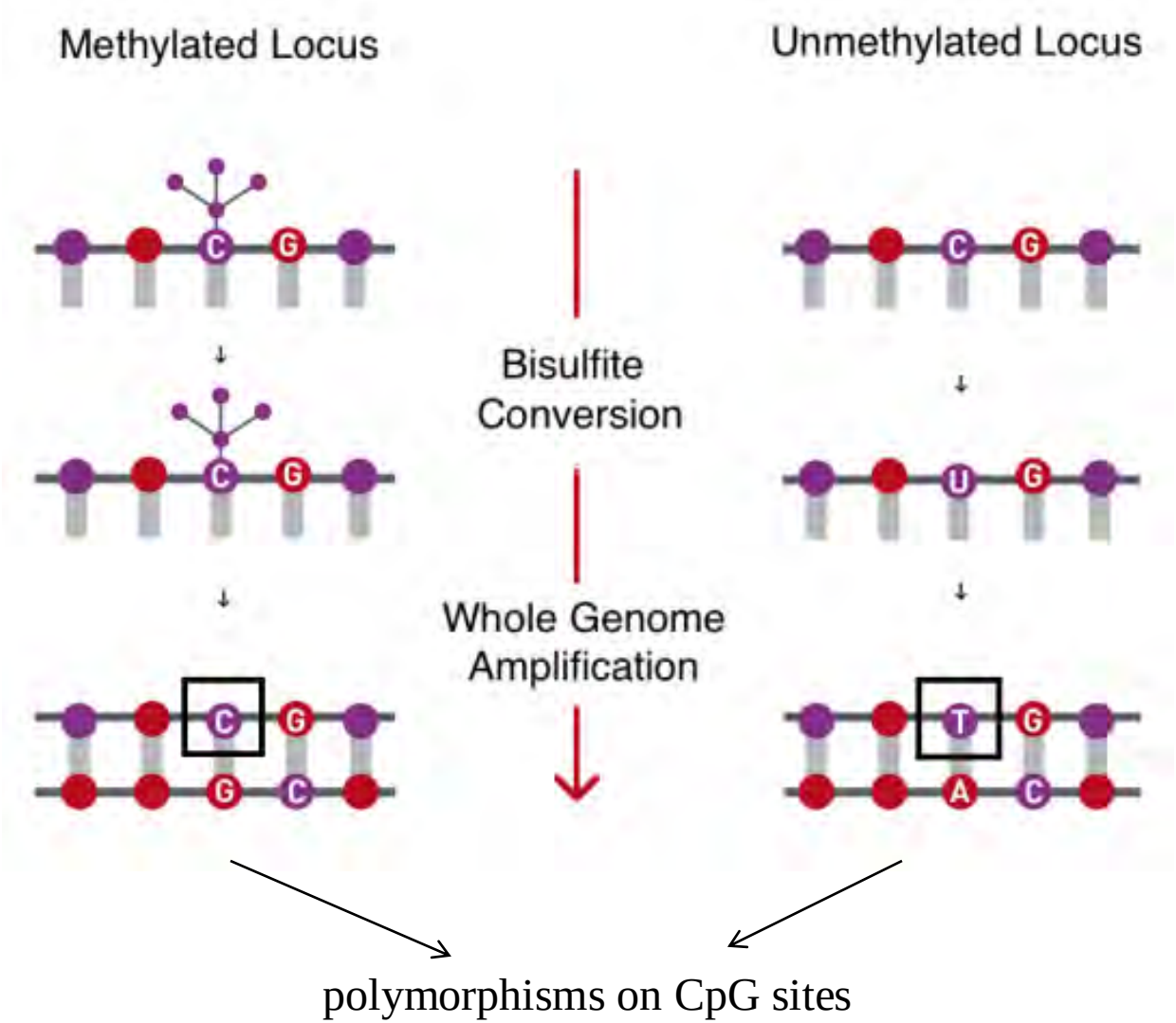
The study of reversible heritable changes in gene function that occur without a change in the sequence of nuclear DNA.



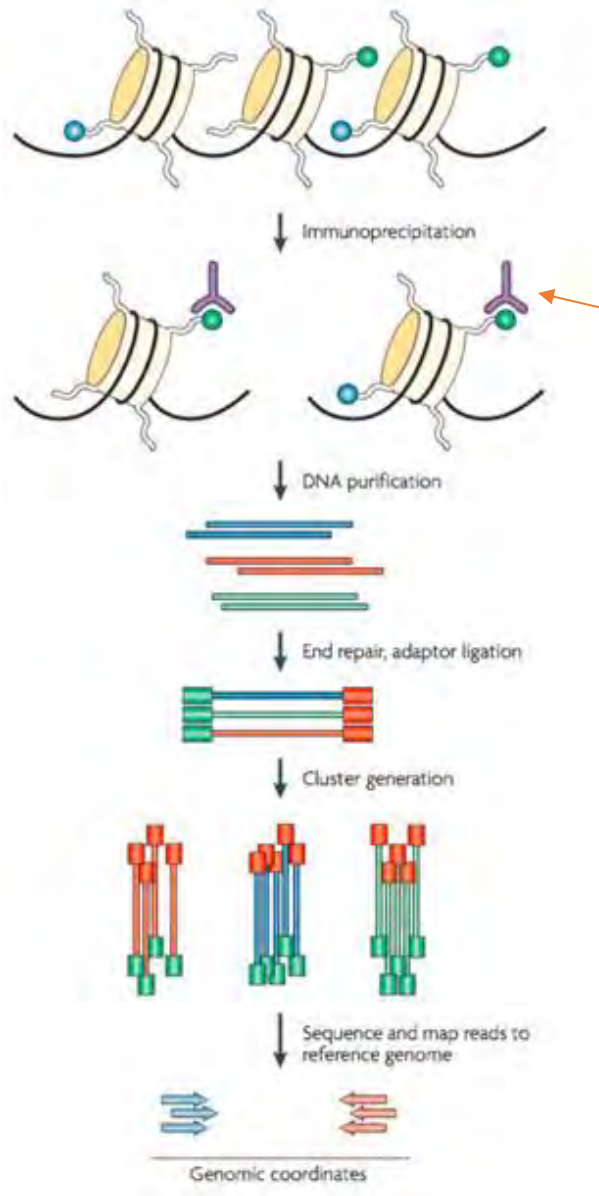
(Rajender, Avery, & Agarwal, 2011)

DNA甲基化测序（WGBS）

Bisulfite Conversion

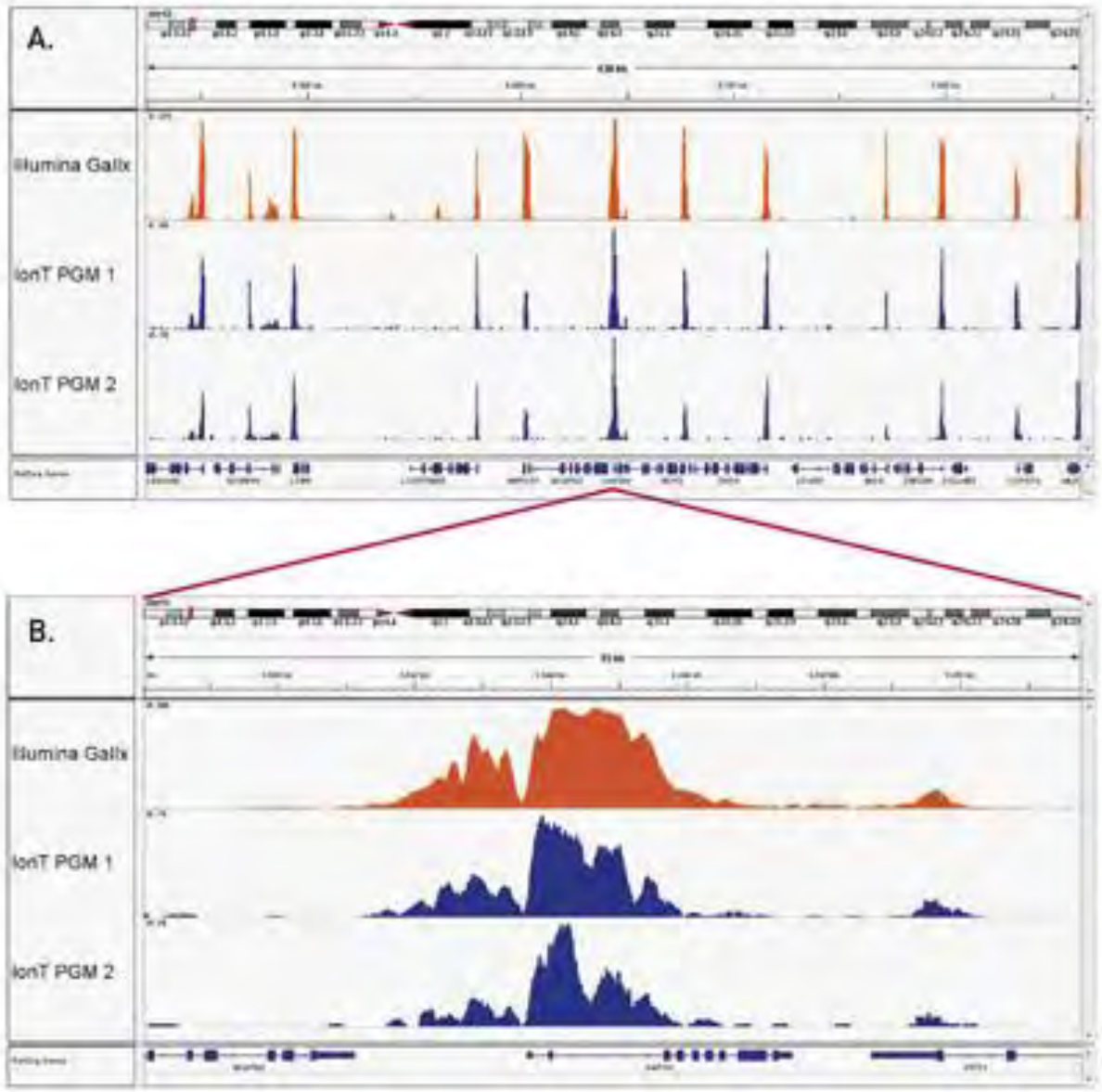


组蛋白修饰（ChIP-Seq）



Antibody specific to
one type of histone modification

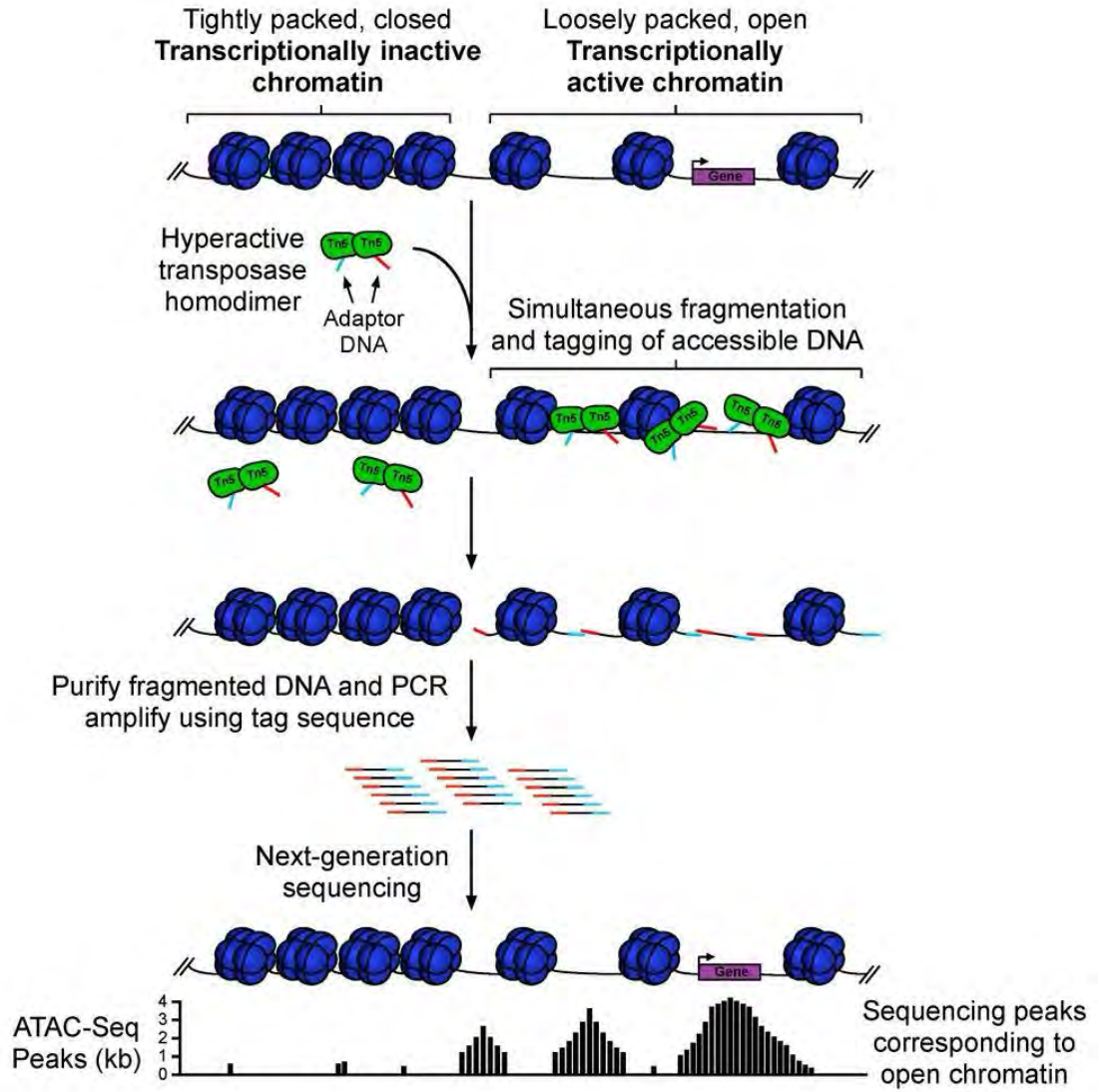
组蛋白修饰（ChIP-Seq）



染色质可及性（ATAC-Seq）

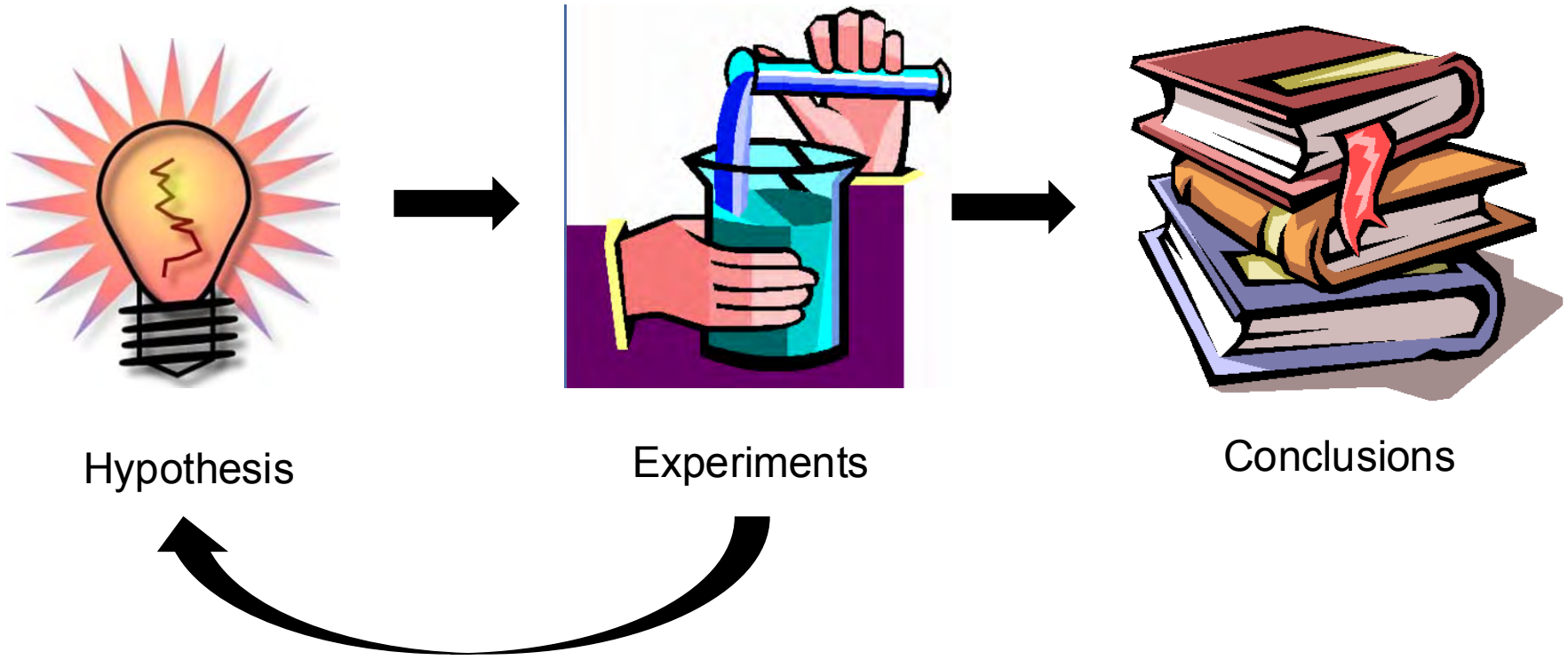
ATAC-Seq: ATAC-Seq is a popular method for determining chromatin accessibility across the genome.

By sequencing regions of open chromatin, ATAC-Seq can help you uncover how chromatin packaging and other factors affect gene expression.



Different ways of thinking in genomic era

Hypothesis driven



Different ways of thinking in genomic era

Data driven

Experiments

Data

Analysis

Sequence
Mutation
Expression
Structure
Function
Evolution
Pathway
Interaction
…

Results

Hypothesis

谢 谢!