



高通量测序技术及其原理

刘赞

复旦大学基础医学院

2024. 9. 20



目录



01

测序技术发展史

02

高通量测序技术原理

03

高通量测序技术应用

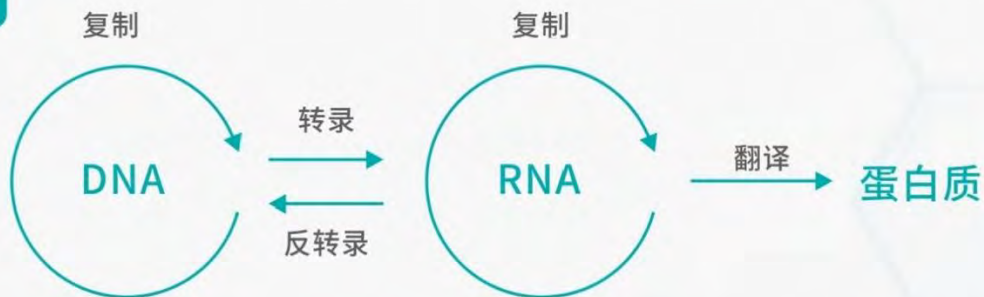
01

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

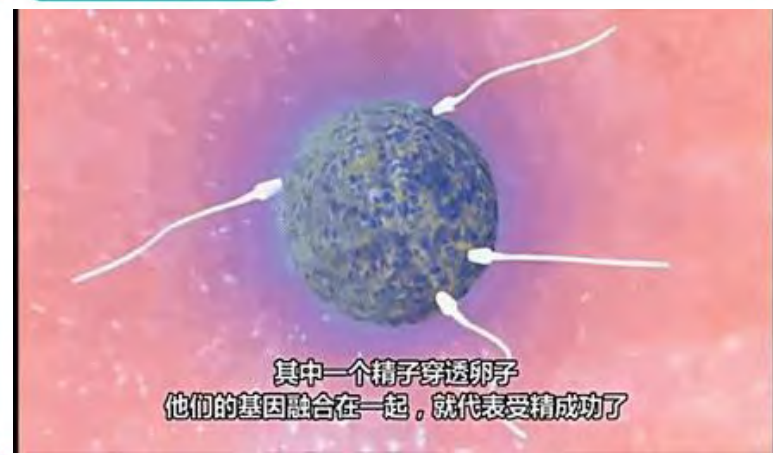
生命中心法则



基因由ATCG组成，生命由简到繁，天书由小到大

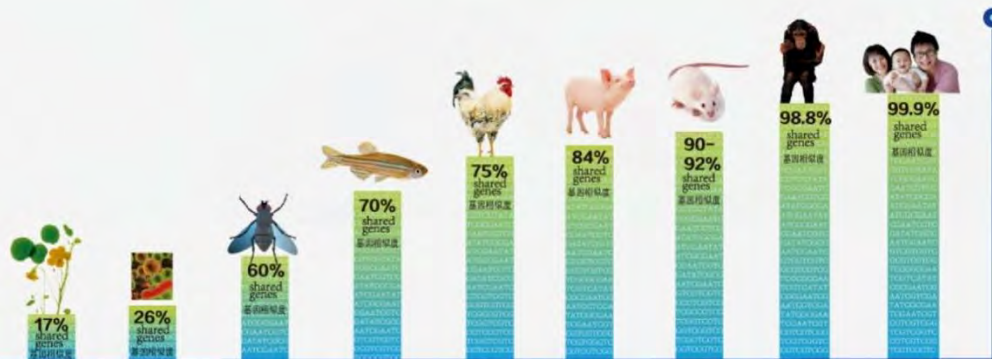
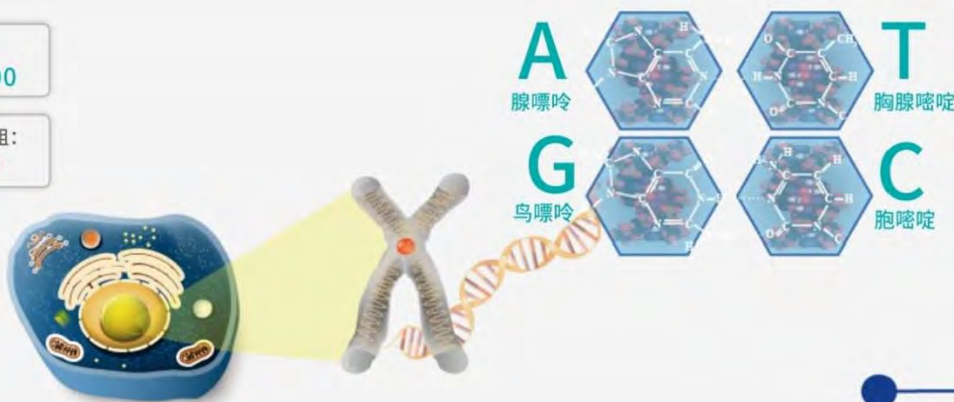
SARS病毒与人类：只是ATCG排序和多少不同

生命的形成



人类基因组：
3,000,000,000

SARS病毒基因组：
约29740bp

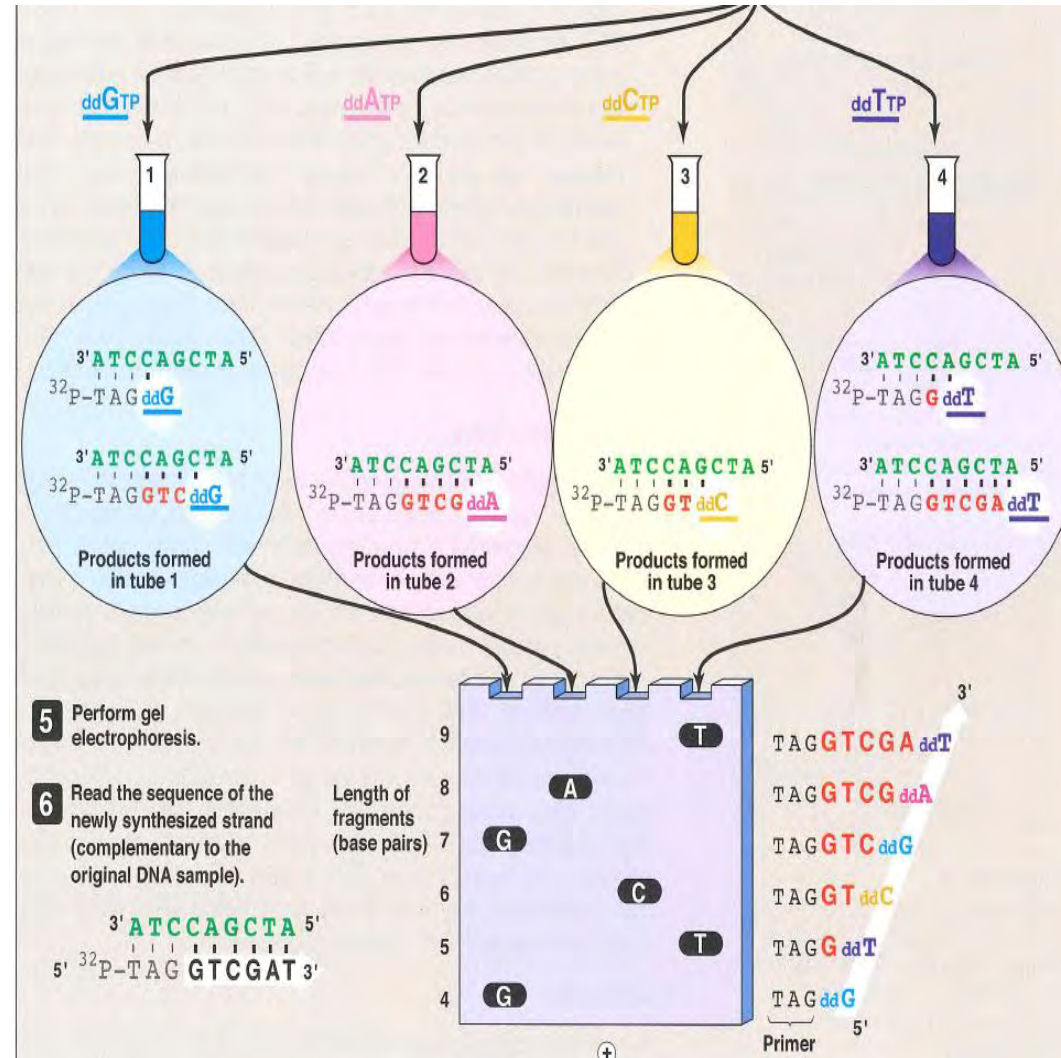


一代测序技术

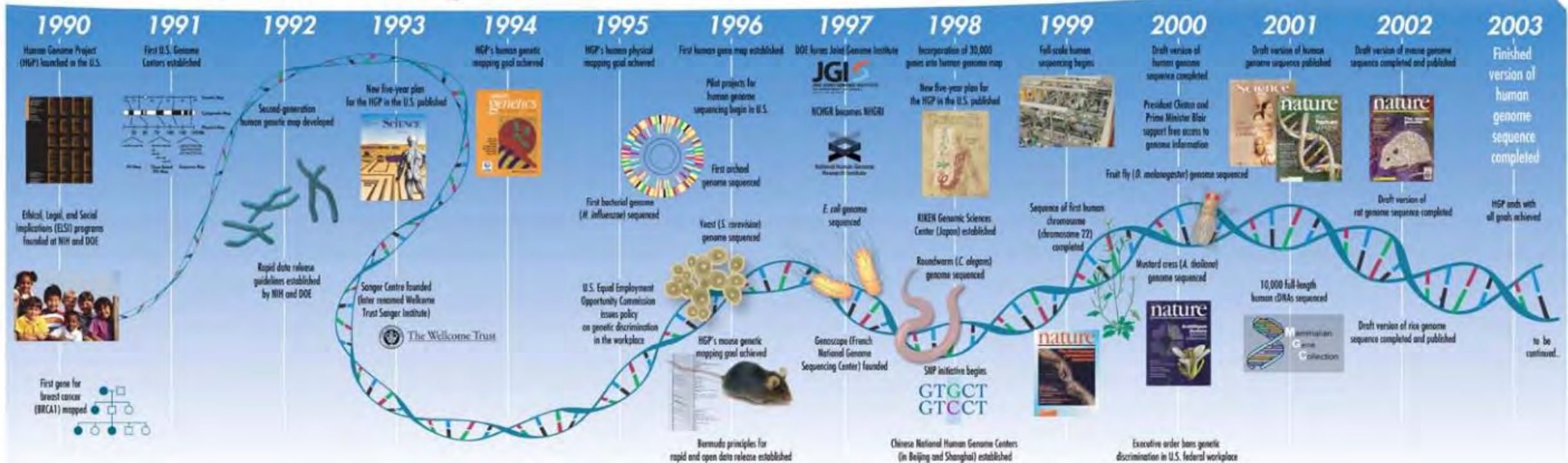
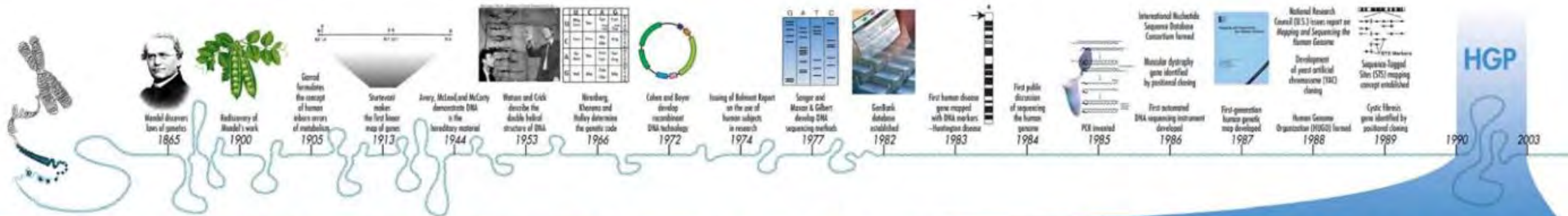
- Single-stranded DNA of unknown sequence is used as a template.

$$3' \text{ATCCAGCTA} 5'$$

$$5' \text{}^{32}\text{P-TAG} \rightarrow 3'$$
- Add primer and DNA polymerase + dATP, dGTP, dCTP, dTTP.
- Split the sample into four tubes, and add one of the four dideoxynucleotides to each.
- 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.

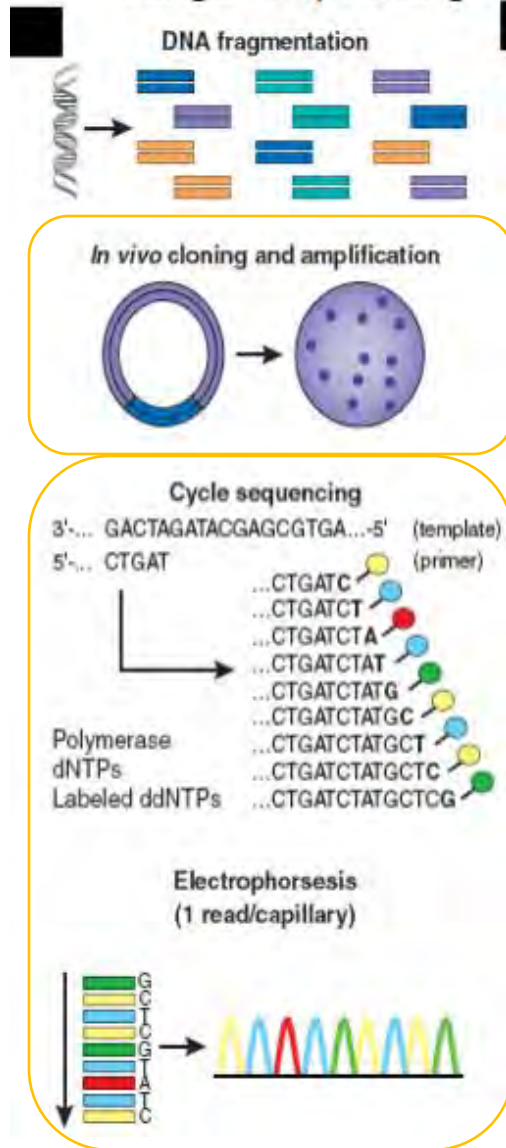


The Human Genome Project

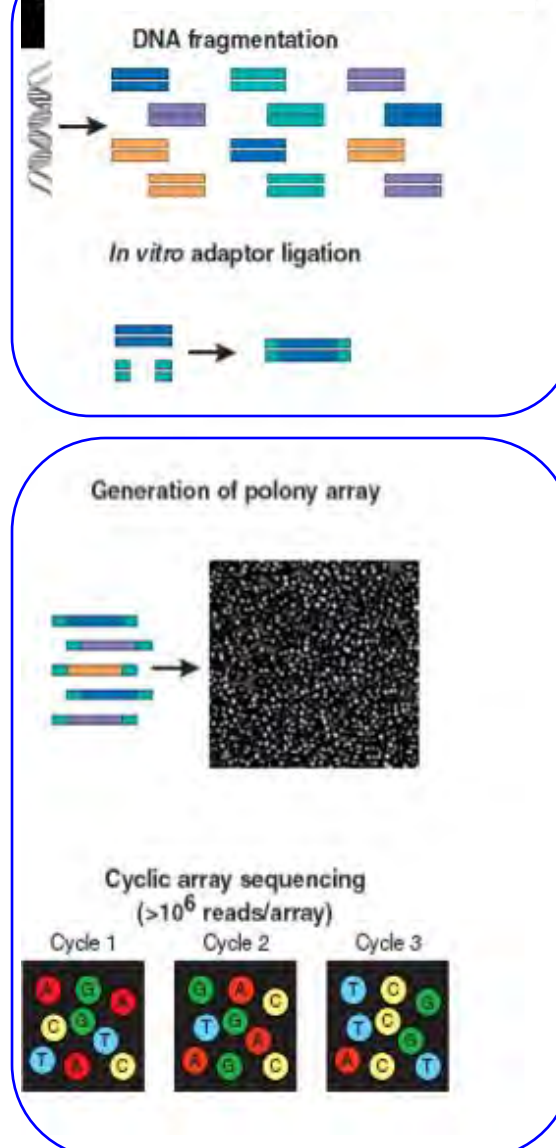


二代测序技术

Sanger sequencing



Next-generation sequencing



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

1953 James Watson和Francis Crick提出DNA双螺旋结构“生命是序列的，生命是数据的”

1964 康奈尔大学的研究者第一次分析了酵母的核苷酸序列

1965 Robert Holley完成丙氨酸的tRNA测序(77nt),耗时7年(3年分离RNA,4年测序)

1977 Frederick Sanger发明链终止法或桑格法Walter Gilbert, Allan M. Maxam发明化学裂解法测序技术

1979 R Staden 提出用电脑程序处理测序结果拼接,奠定了shotgun测序的理论基础。

1983 Mullis发明了PCR技术

1986 荧光标记测序法诞生

1987 Applied Biosystems (简称ABI) 推出第一台商业化的DNA测序仪ABI 370A, 装载96个通道, 产量为1000bp/天

1987 Leroy E.Hood 开发第一台半自动测序仪

1990 美国正式启动人类基因组计划(Human genome project) 提出双端shotgun测序,

1986 荧光标记测序法诞生

1987 Applied Biosystems (简称ABI) 推出第一台商业化的DNA测序仪ABI 370A, 装载96个通道, 产量为1000bp/天

1990 美国正式启动人类基因组计划(Human genome project) 提出双端shotgun测序, 应用于测序人的HGPRT基因

1995 ABI推出Prism 310测序仪,是第一台毛细管电泳测序仪,产量为5kb-15kb每天,是人类基因组计划的基础设备

1996 Pal Nyren和Mostafa Ronaghi发明焦磷酸测序法

1997 Molecular Dynamics 公司推出MegaBACE1000测序仪, 通量为250kb-500kb/天

1998 ABI推出Prism 3700毛细管测序仪,通量为500kb-1000kb/天,初步实现了测序规模化

1999 中国加入人类基因组计划,华大基因成立并负责测定全部序列的1%

2000

2001 Lynx Therapeutics (Sydney Brenner和sam Elter成立的) 发布第一个商业化二代测序方法MPSS(Massively parallel signature sequencing), 但不对外销售

2001 人类基因组计划取得重大突破, 2月15日在《自然》杂志上发表了第一个人类基因组

2005 454公司推出了第一款二代测序仪Genome Sequencer20

2006 Solexa推出高通量基因组测序系统Genome Analyzer, 通量为1Gb每Run, 该系统采用高通量的边合成边测序技术,该技术使单分子DNA在Flow Cell中进行成千倍的模式扩增,再利用dNTP的边合成边测序技术,一边合成DNA一边读取碱基信息,从而进行测序

2007 ABI推出Solid测序系统, 单次运行可产生4Gb的序列数据

2008 Helicos公司第一台单分子测序仪Heliscope, 每次运行的总产量达21.35G,但较高的错误率阻碍了进一步普及

2010 华大买入128台Illumina测序仪, 发展成为全球最大的基因组学研究机构

2010 Life Tech推出Ion PGM桌面测序仪, 通量灵活, 运行速度快, 读长400bp, 通量分别为10M, 100M, 1G

2010 Illumina推出桌面型测序仪MISEq系列, 最大读长300bp, 通量为500M到15G每Run, 个人基因组测序降至4000美元

2011 PacBio 推出PacBio RS II测序系统, 平均读长为5000bp, 每次运行可产生100M-1G数据

2012 Life tech推出基于半导体技术的台式高通量测序仪Ion Proton和Ion Torrent, 通量分别为10G和100G

2013 华大收购Complete Genomics

2013 Thermo fisher收购Life tech

2013 Illumina推出的MISEq成为第一台通过FDA批准的体外诊断(IVD)测序系统, 最大读长300bp, 通量为540M-15G

2014 华大测序仪BGISEQ-100和BGISEQ-100通过CFDA(今NMPA)医疗器械注册认证

2014 Illumina收购Solexa

2015 罗氏公司宣布454测序技术停止开发, 2016年退出市场

2015 华大推出BGISEQ-500, 具有精准、简单、快速、灵活、可拓展的特点, 读长为2x100bp, 通量为520G, 并通过NMPA医疗器械注册认证

2016 Illumina推出NextSeq500/550台式测序仪, 具有高通量、高通量两种模式, 单次运行可获得20-120G的数据, 一天可完成一个全基因组测序和16个外显子组测序

2016 BGISEQ-500采用优化的联合探针锚定聚合技术(cPAS)和改进的DNA纳米球(DNB)核心技术, 测序时, 首先DNA分子锚和荧光探针在纳米球上进行聚合, 随后高分辨成像系统对光信号进行采集, 经过数字化处理后即可获得待测序列。DNB技术通过循环性扩增技术增强信号, 大幅降低了单碱基的错误率。DNB大小与芯片上的活性位点相匹配, 每个位点承载一个DNB纳米球, 同时保证了测序精度和芯片利用率

2016 Thermo fisher推出ionS5和ionSXL, 读长为400bp, 运行时间2-4小时, 通量为0.5-15G

2016 华大集团旗下华大智造推出BGISEQ-50, 读长为50bp, 通量为8G, 具有专注、小巧、精准的特点, 10小时即可自动完成样本加载和测序

2016 Illumina同时推出HiSeq3000/4000, 读长2x150bp, 通量高达1500Gb

2016 华大发布完全集成式的“超级测序仪”Revolyocity

2017 Illumina推出小型桌面型测序仪MiniSeq, 读长为2x150bp, 通量为1.65-7.5G

2017 华大智造发布全面型测序仪MGISEQ-2000, 采用全新芯片设计, 可灵活支持多种不同的测序模式, 同时优化了光学和生化系统, 能在较短时间内完成整个测序流程, 通量高达1080G每Run

2017 Illumina推出了由10自HiSeq系列测序仪组成的HiSeq X Ten, 读长为2x150bp, 通量高达1.8T每Run, 600G每天, 每年可产出20.00个人全基因组测序, 个人基因组测序降至1000美元

2018 华大智造发布超高通量测序仪MGISEQ-T7, 采用四联芯片平台, 测序热PE100<20小时, 日产出数据高达6T, 一天最多可完成60例个人全基因组测序, 每G测序成本低至5美元, 全基因组测序成本进一步降低, 进入百美元时代

2018 Illumina推出

2019 华大智造发布全球首台模块化NGS工作站MGFLP

2019 Illumina推出NovaSeq, 具有超高通量、低成本的特点, 通量最高可达6T每Run, 进一步降低了测序成本

2019 华大智造发布高精度基因组组的“6T”版本: Contig N50 > 10, Scaffold N50 > 10, 挂载全基因组数据 > 6G, 基于“6T”标准的从基因组二倍体基因组数据, 开启了基因组测序与基因组学时代, 对基因组学直接获取遗传信息有重要程序式的突破, 释放大规模个性化医疗的发展

2020

Illumina测序平台

illumina

illumina

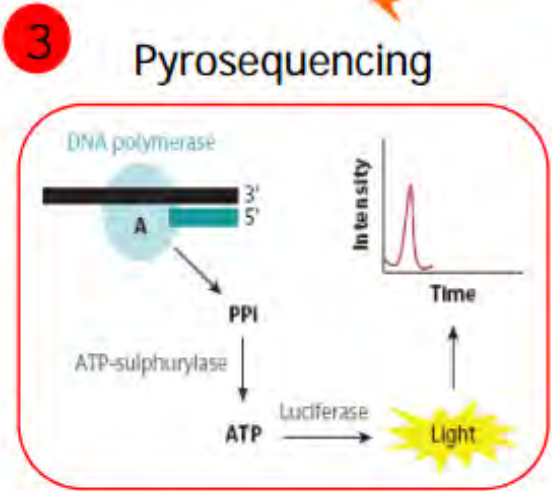
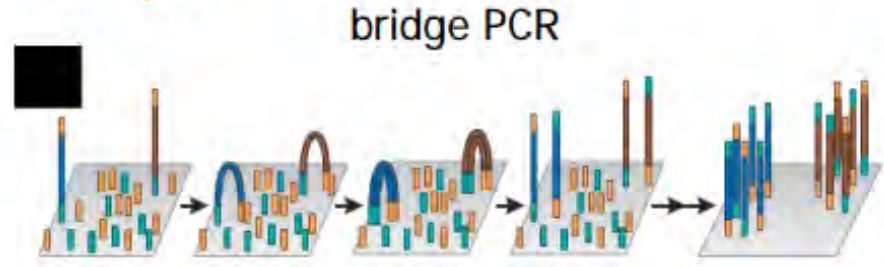
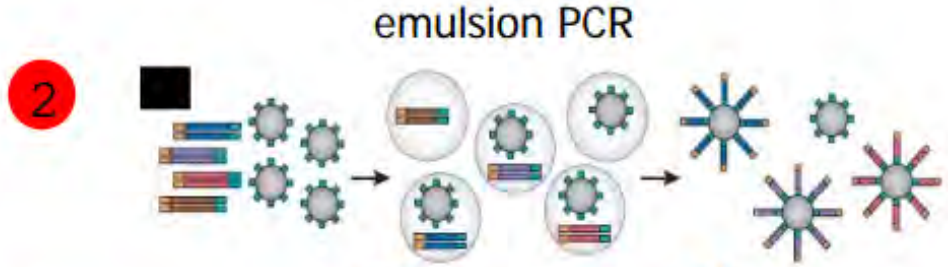
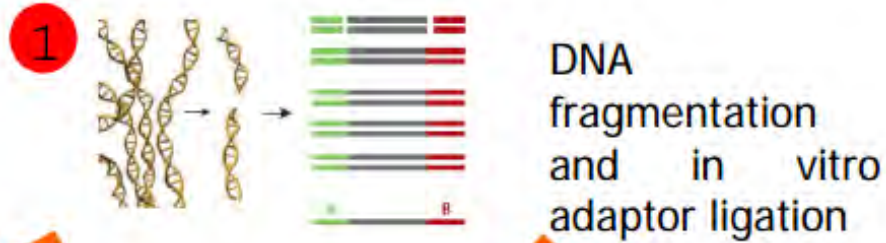


0:00 / 5:03

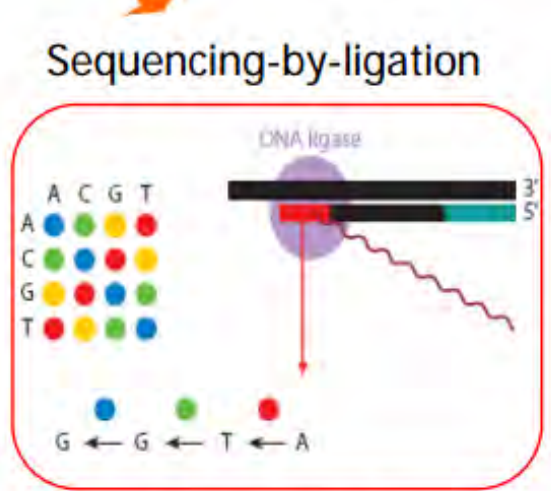


不同的二代测序平台

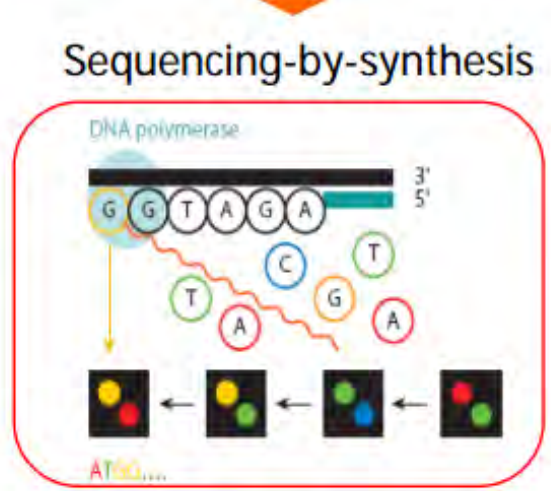
- 1 Library preparation
- 2 Clonal amplification
- 3 Cyclic array sequencing



454 sequencing



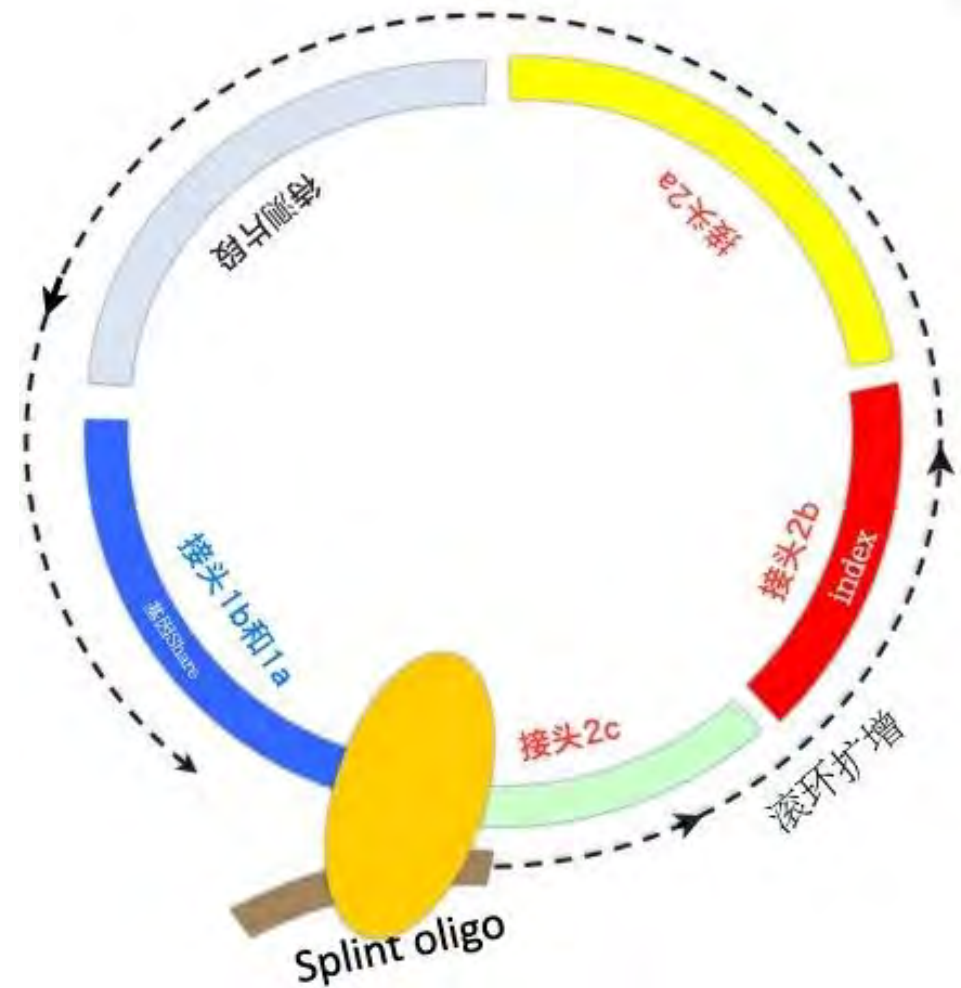
SOLiD platform



Solexa technology

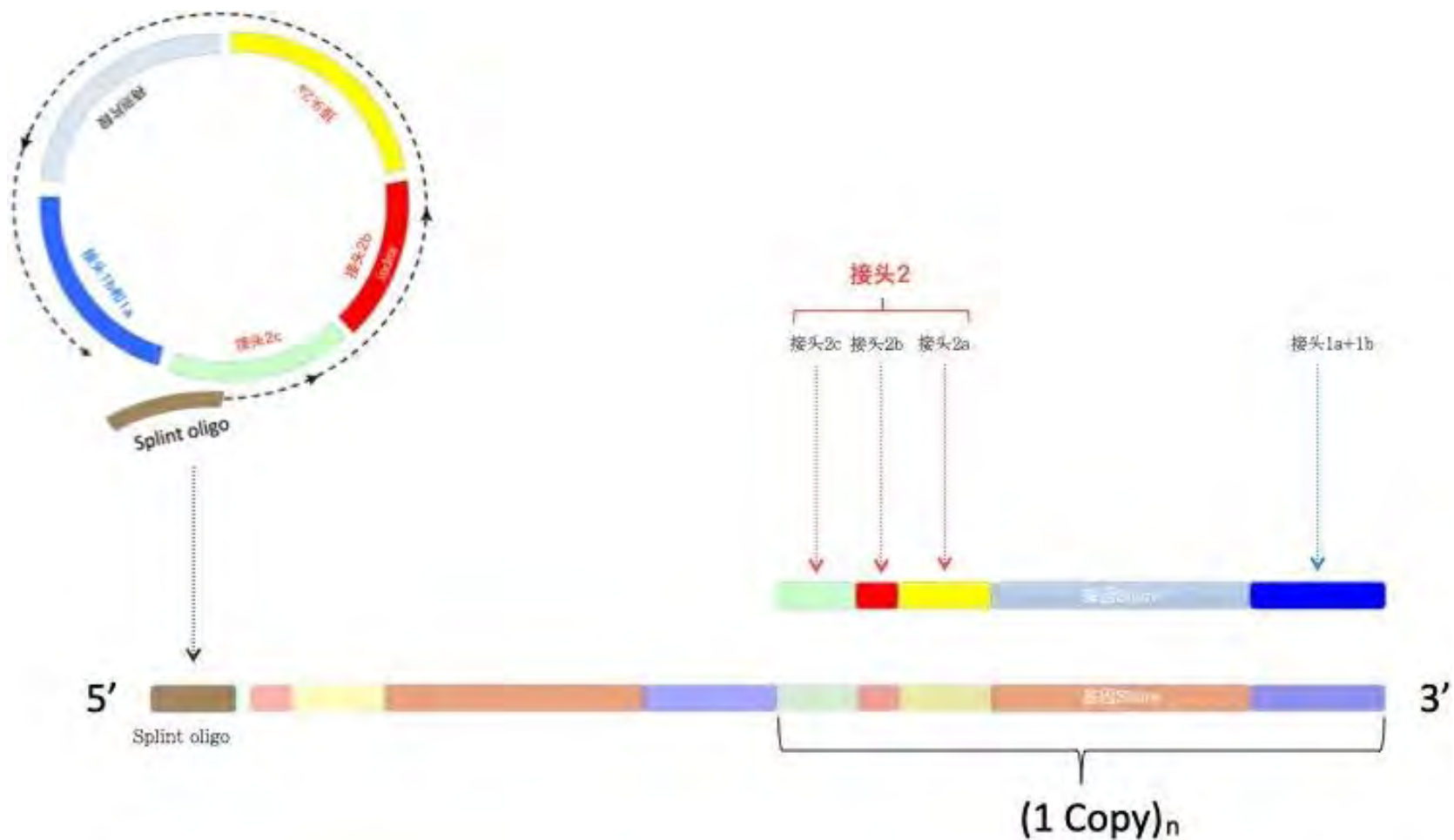
华大DNB测序技术

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

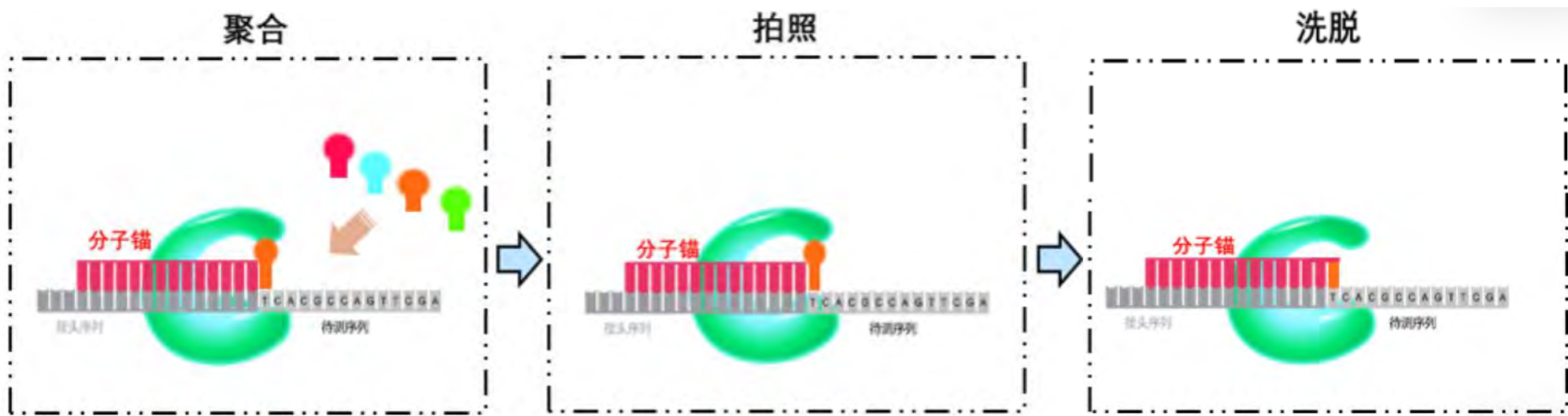
b) DNA纳米球“拉直”后即下图结构，待测片段“串联重复”



华大DNB测序技术

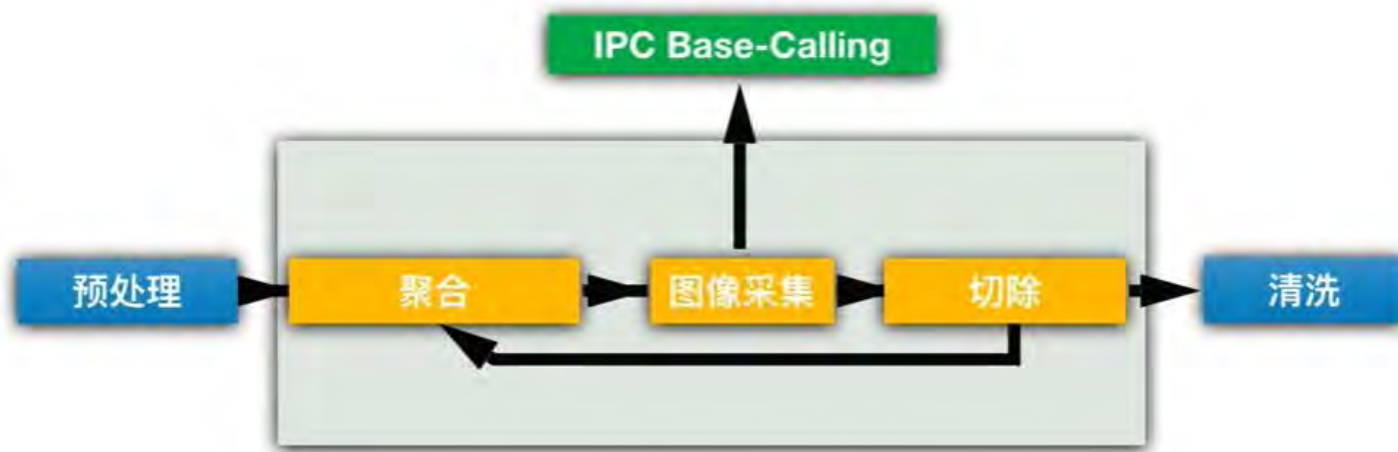
联合探针锚定聚合技术(Combinatorial Probe-Anchor Synthesis, **cPAS**):

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

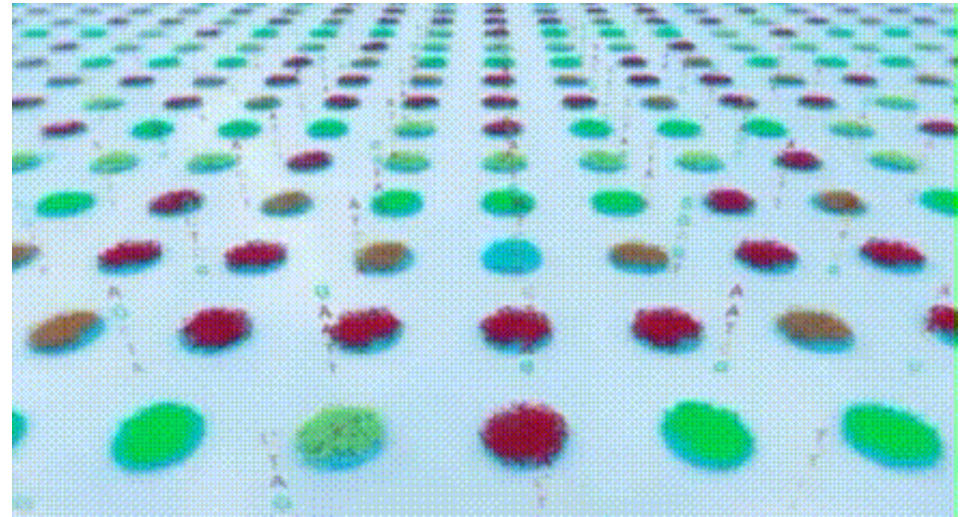
测序原理



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

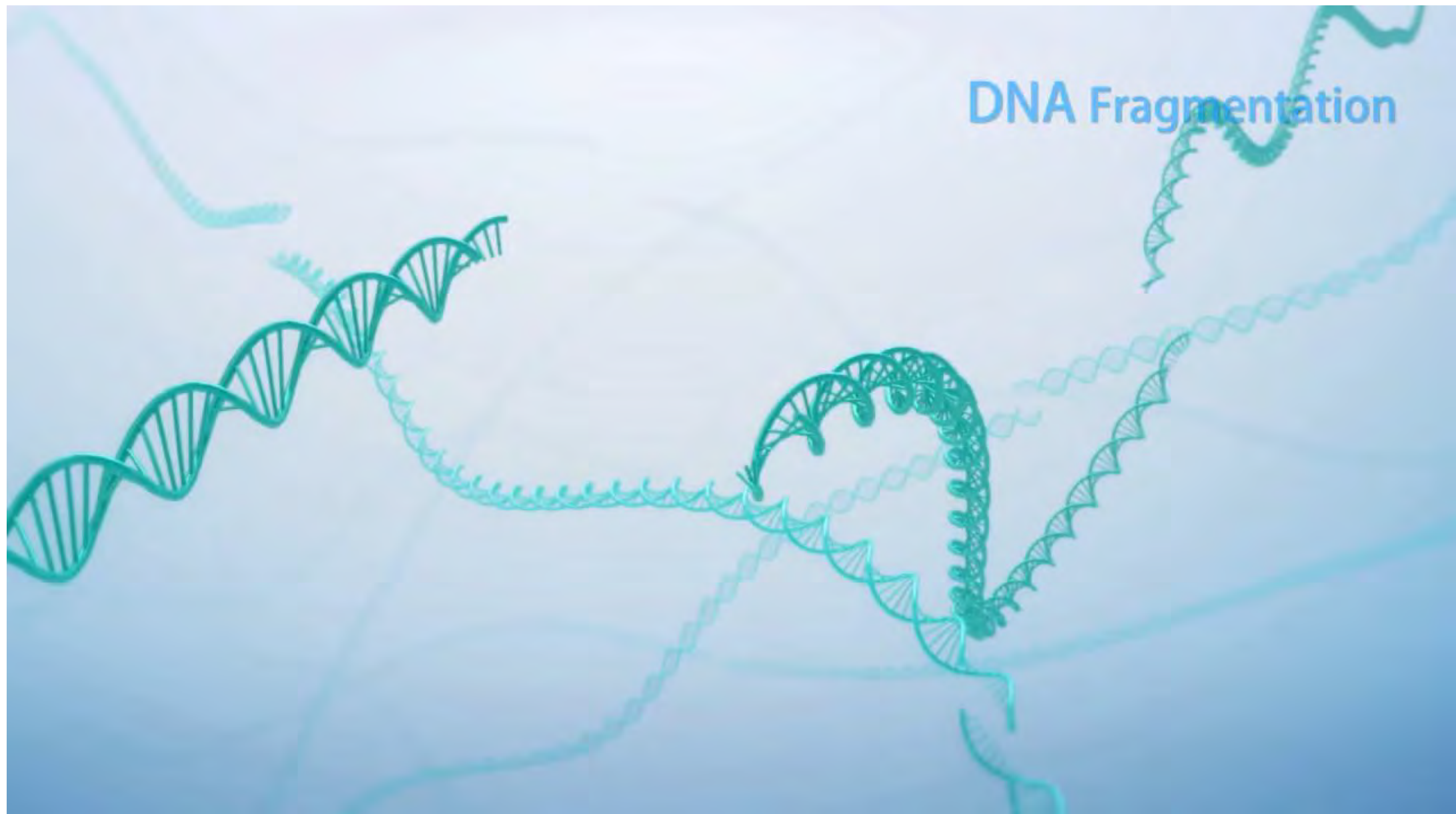


BGISEQ-50

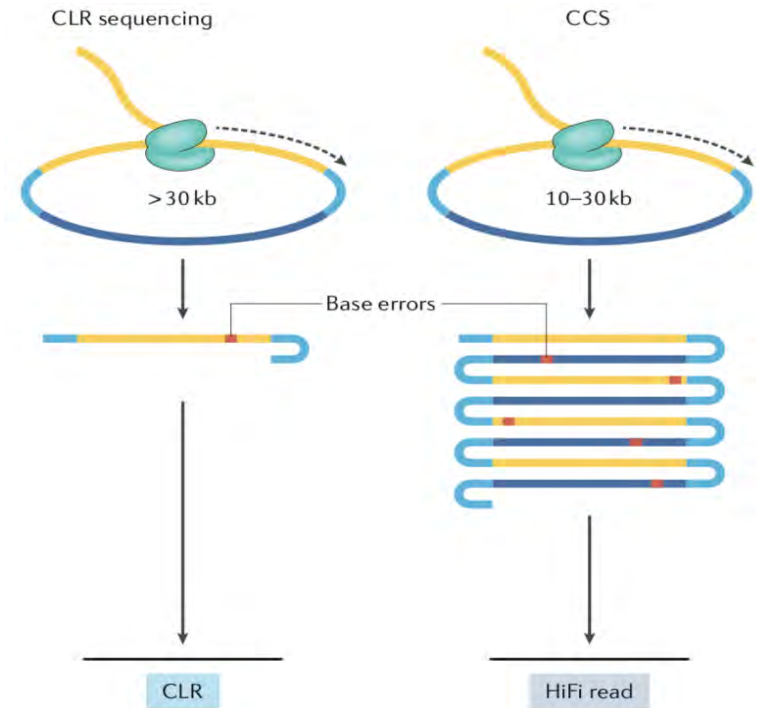
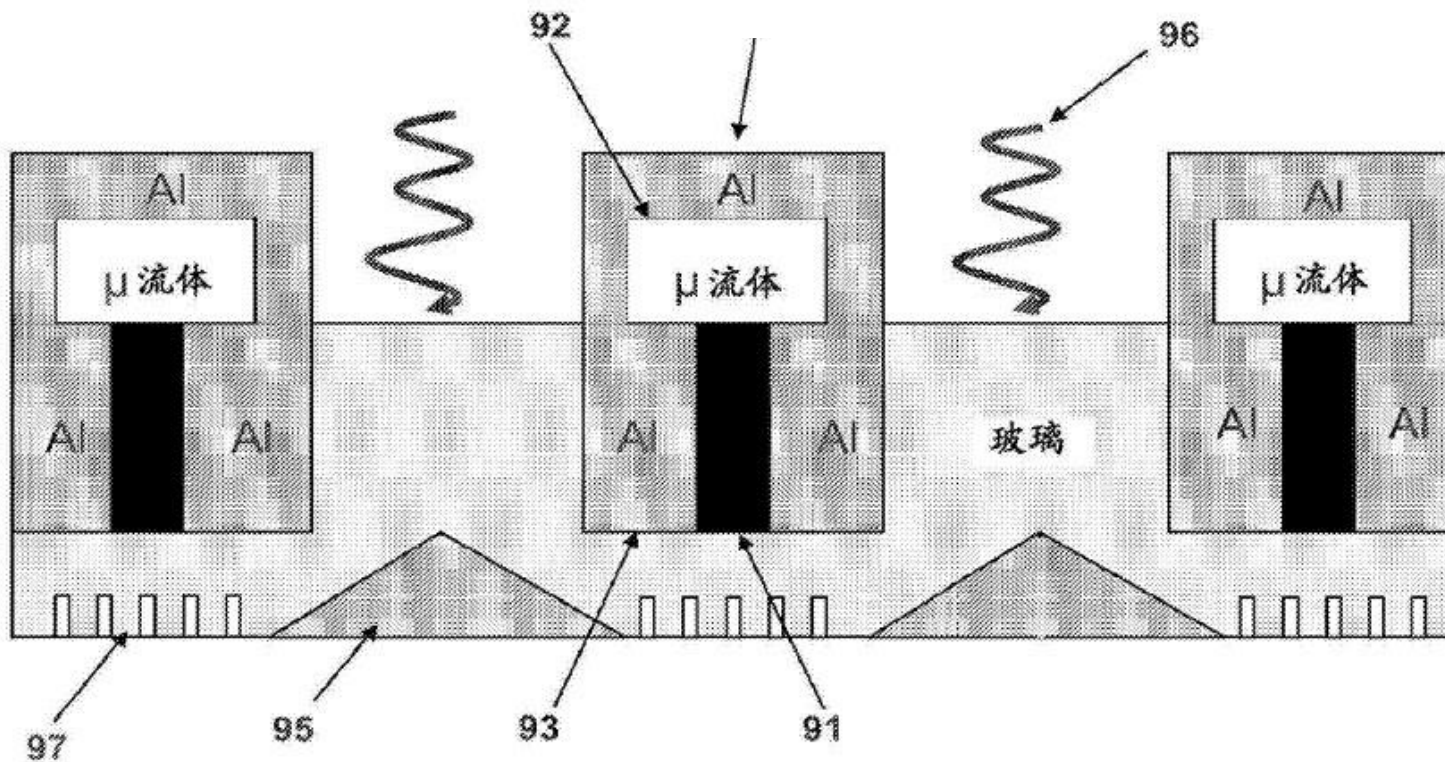
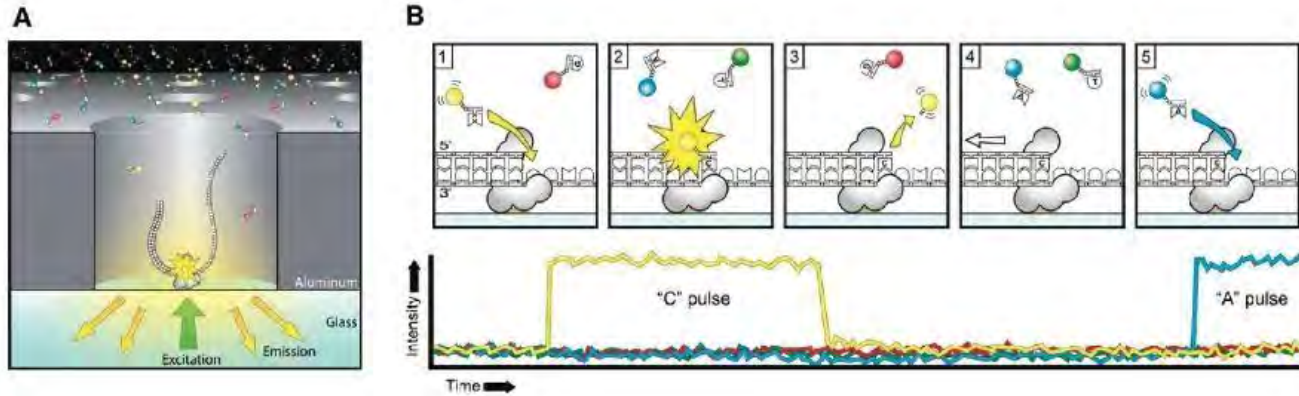


MGISEQ-200

华大DNB测序技术

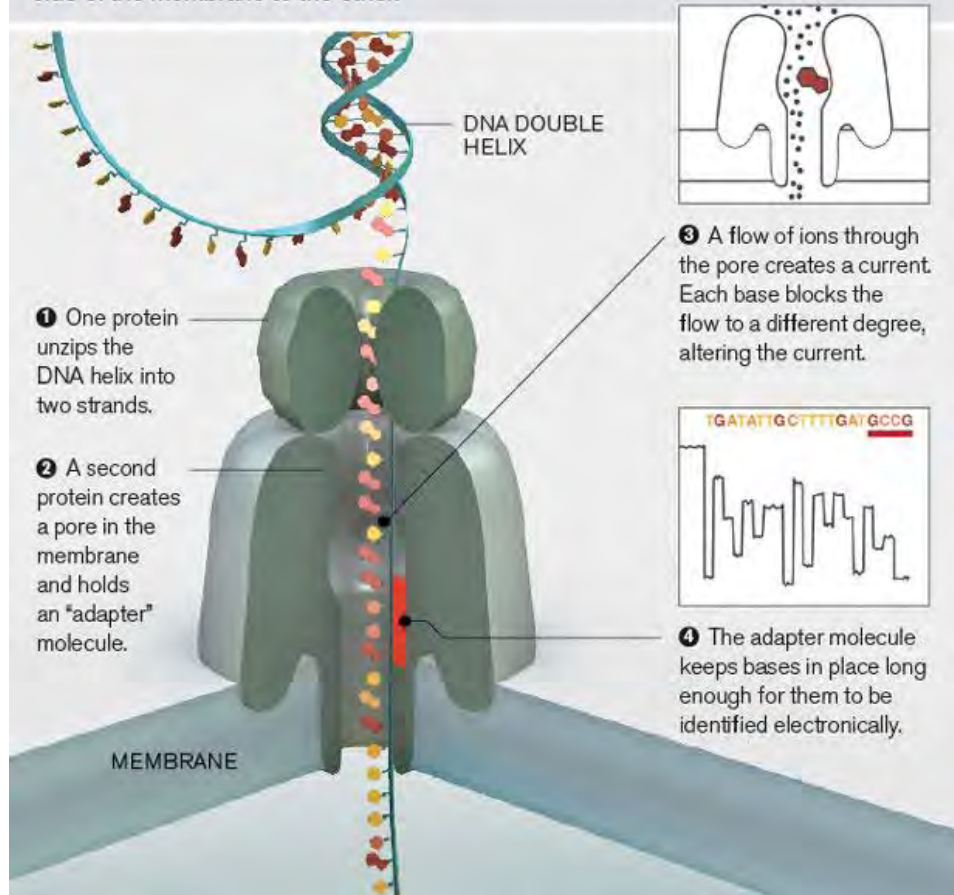


三代测序技术 - PacBio

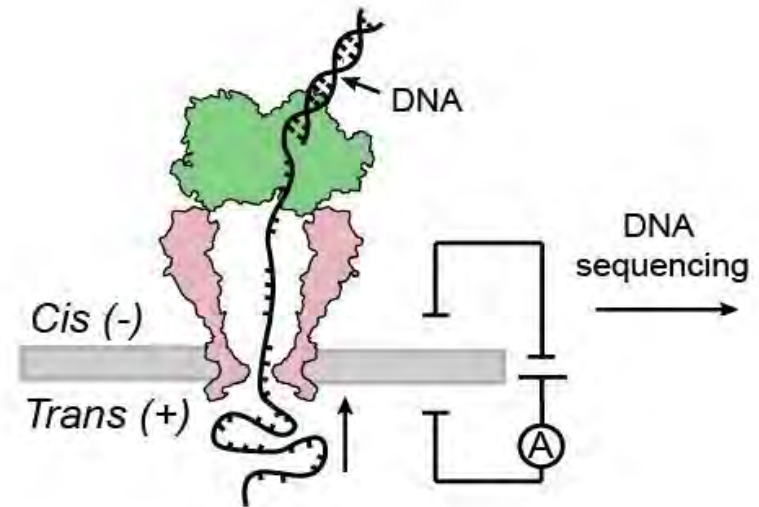


三代测序技术 - Nanopore

DNA can be sequenced by threading it through a microscopic pore in a membrane. Bases are identified by the way they affect ions flowing through the pore from one side of the membrane to the other.



- 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)

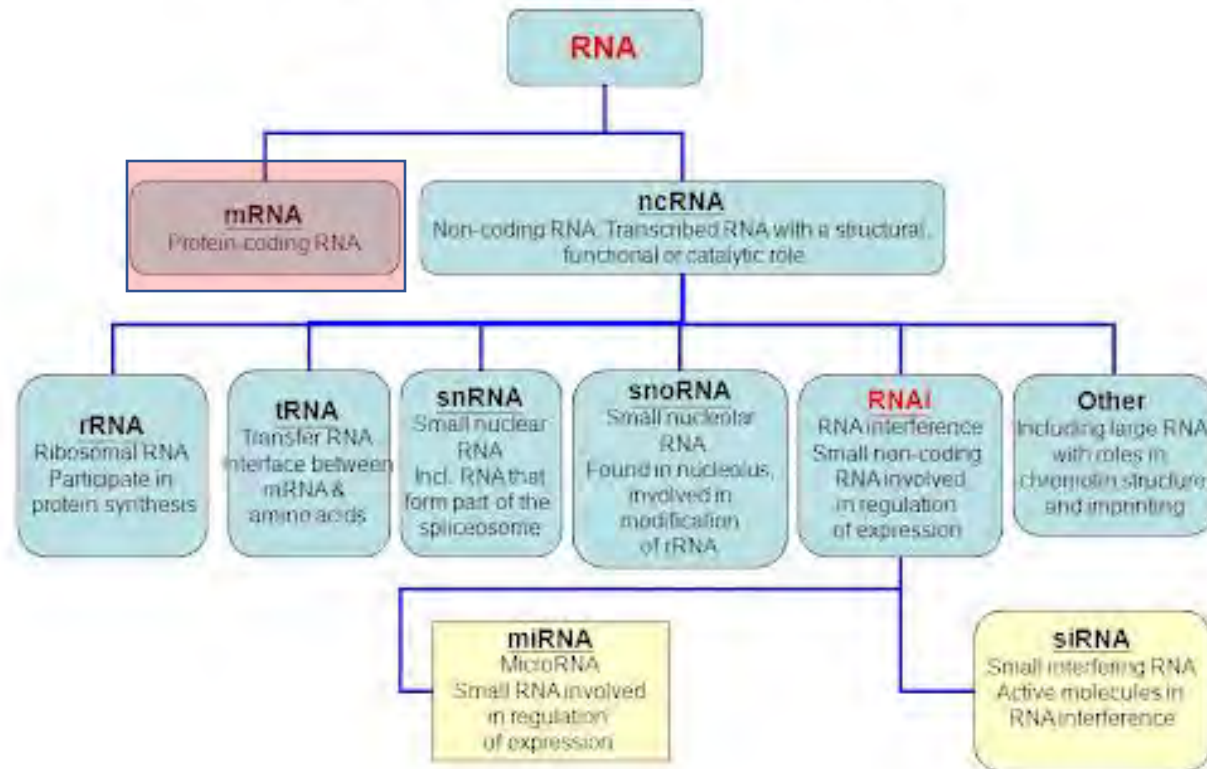
Epigenetics

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

RNA-Seq

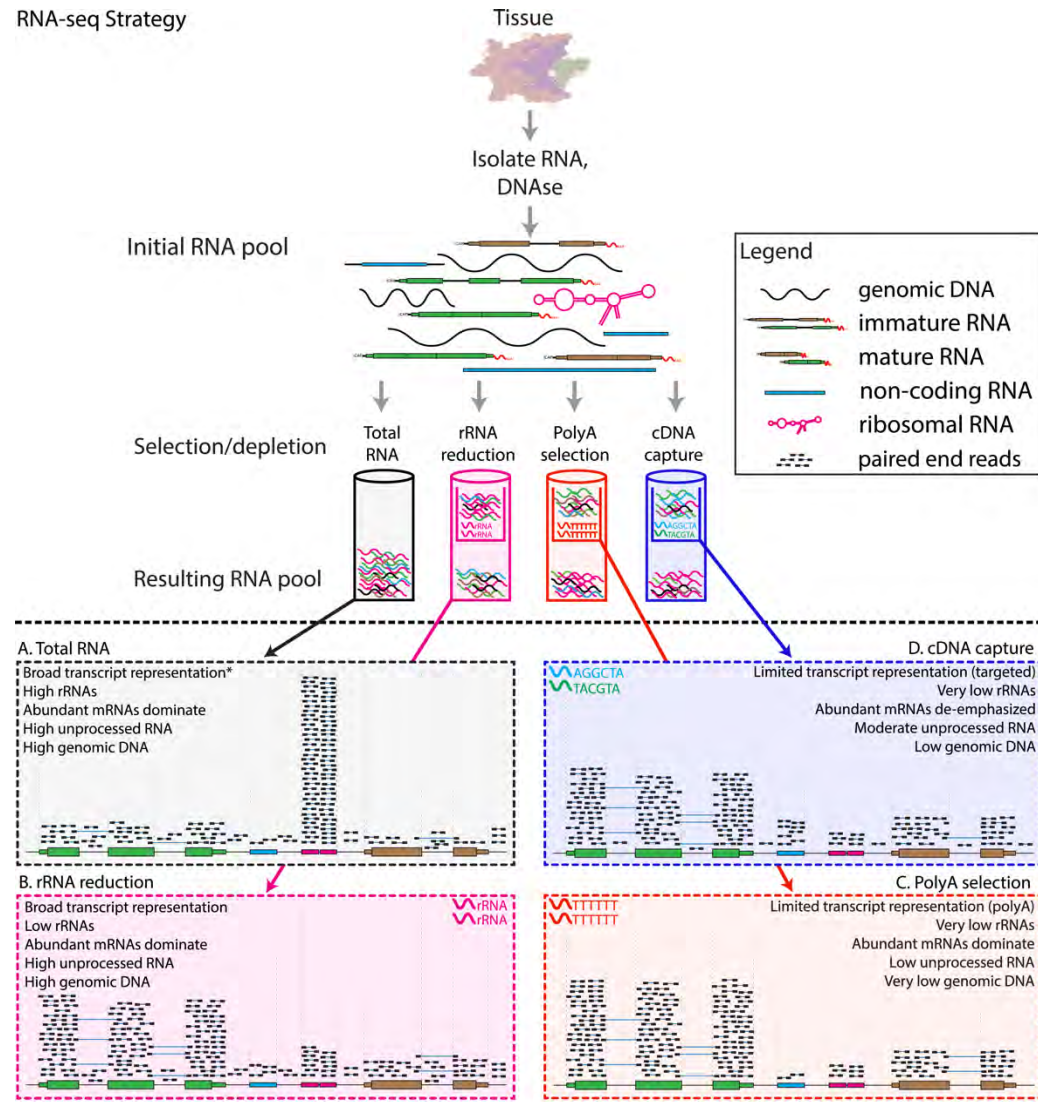
RNA sequencing (RNA-Seq) is a sequencing technique which uses next-generation sequencing (NGS) to reveal the **presence** and **quantity** of RNA in a biological sample at a given moment.

Type of RNA molecules



不同的RNA-Seq策略

RNA-seq Strategy



Expected Alignments

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

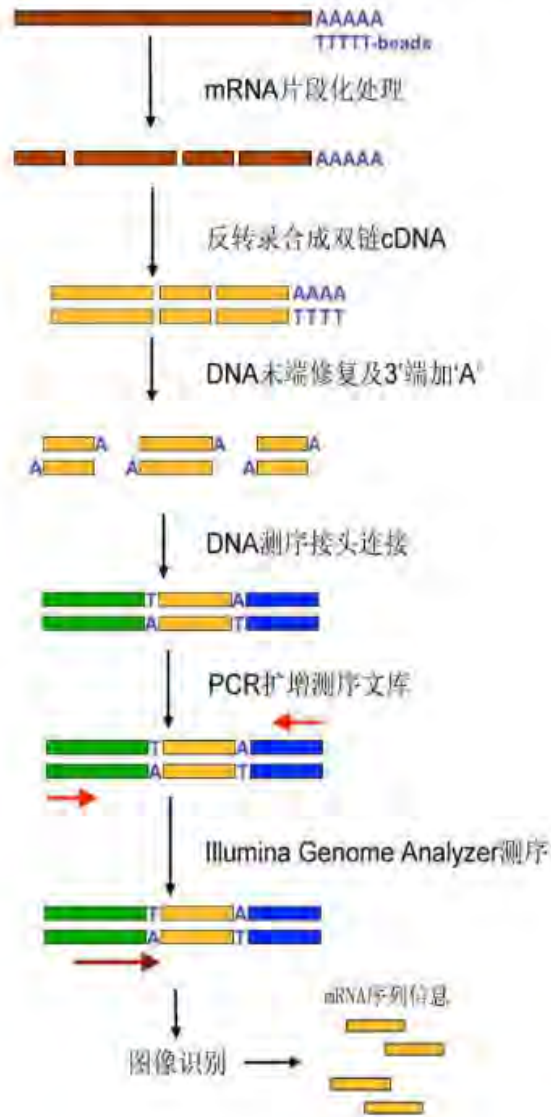
4. 高通量测序 (Illumina Genome Analyzer IIx)

5. 数据分析

原始数据读取

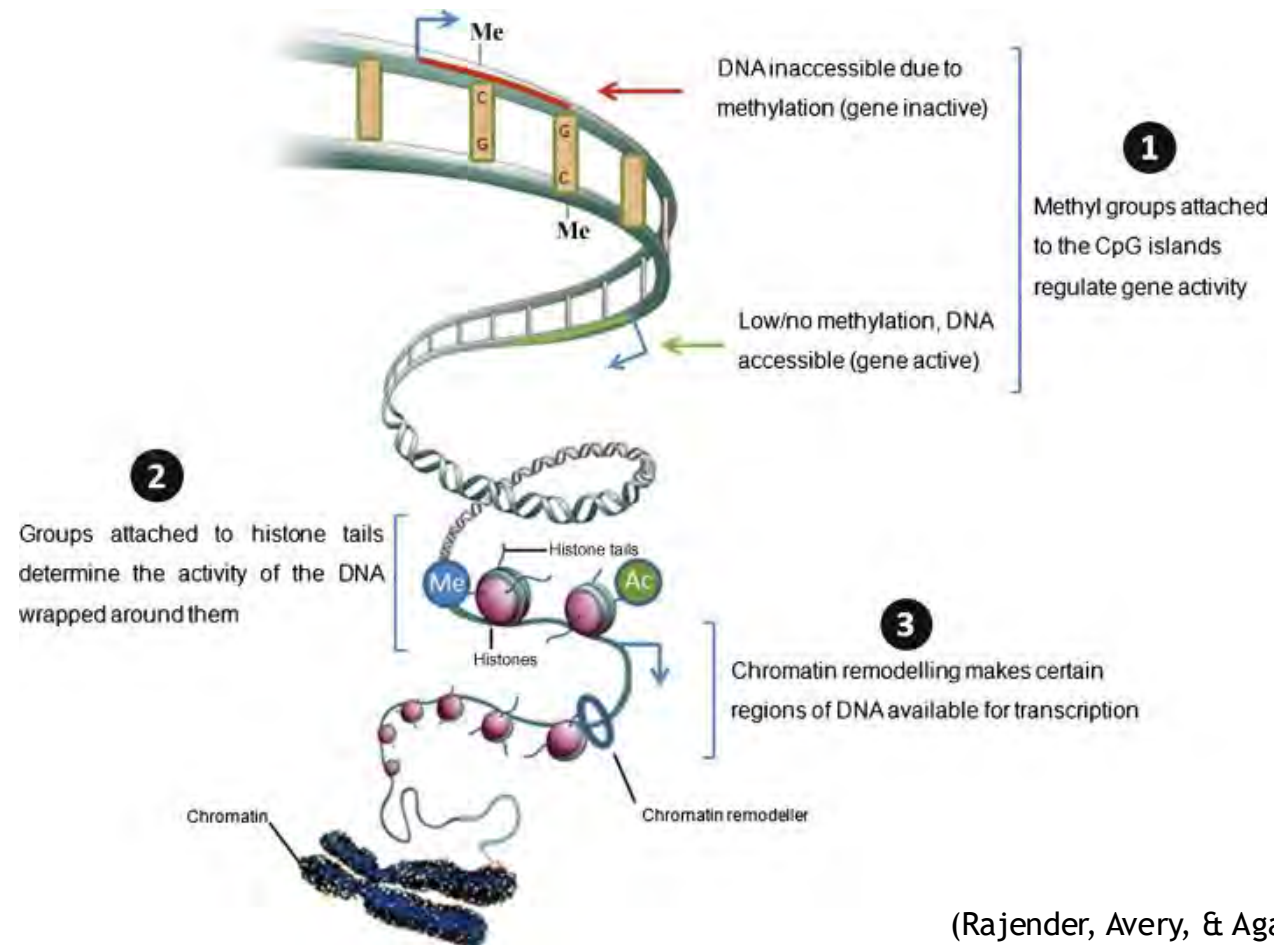
与数据库比对并进行注释

深层次数据分析



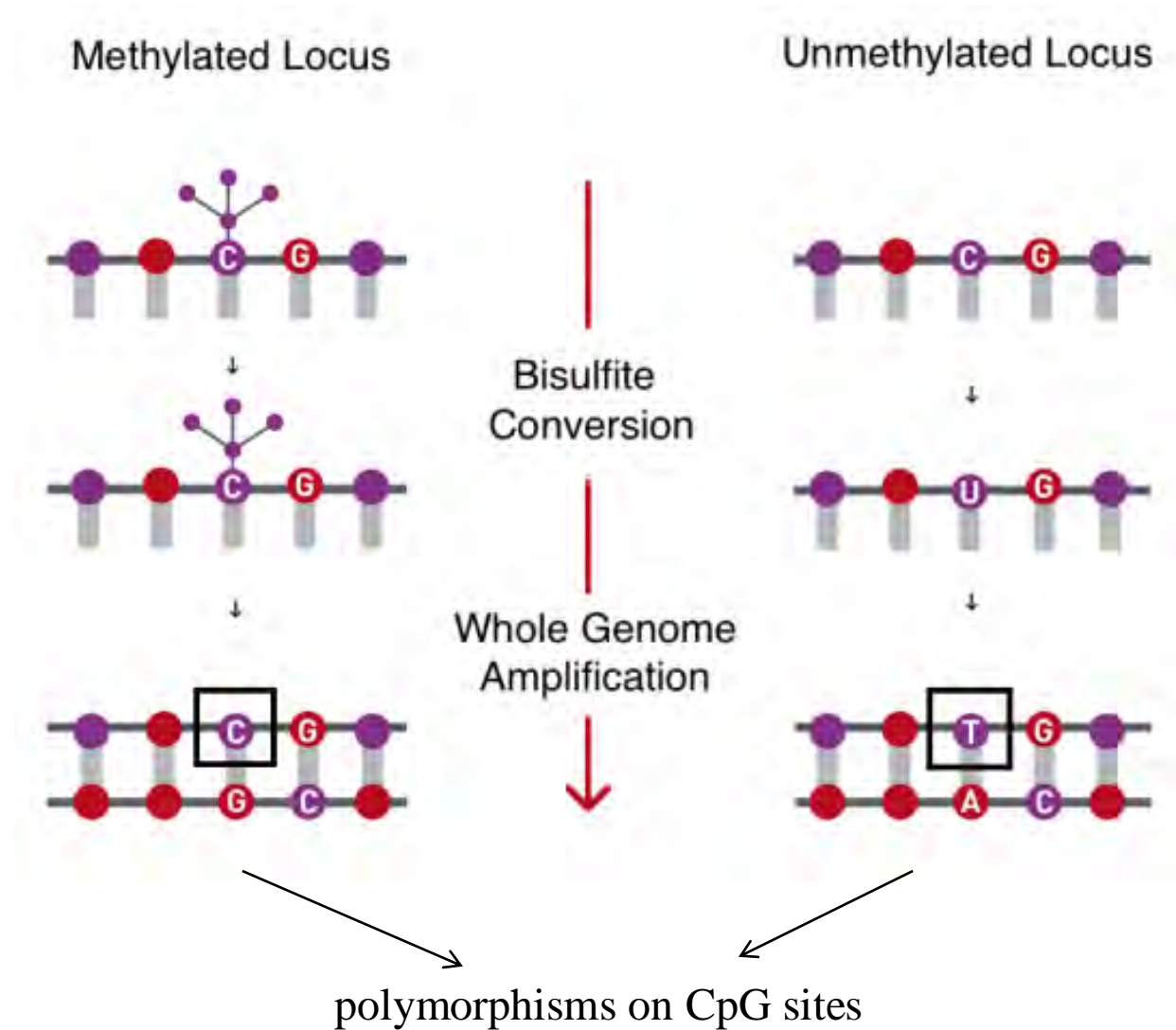
Epigenomics

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

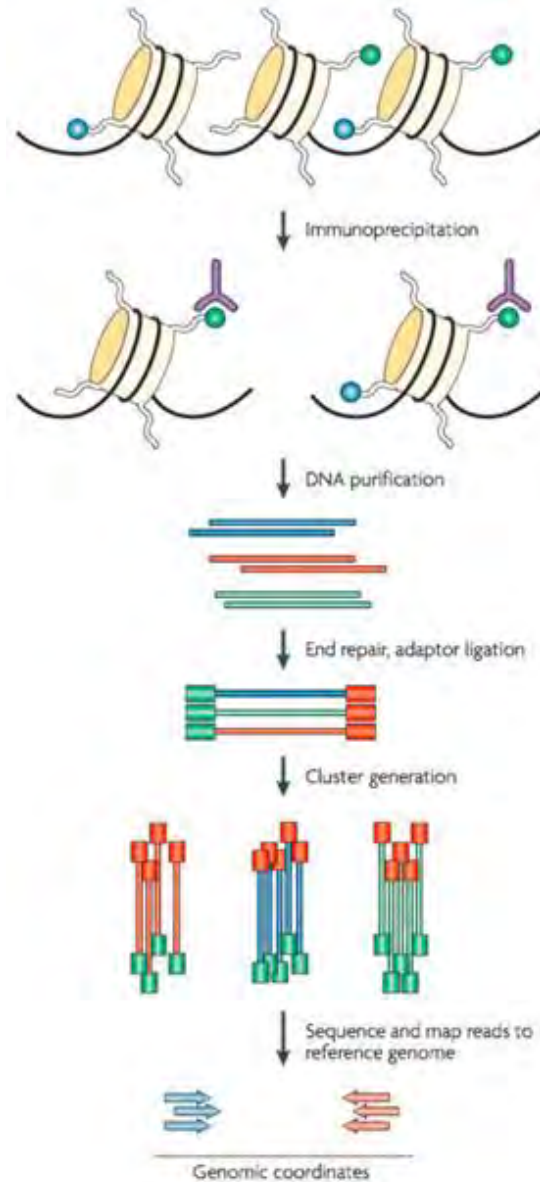


DNA甲基化测序 (WGBS)

Bisulfite Conversion

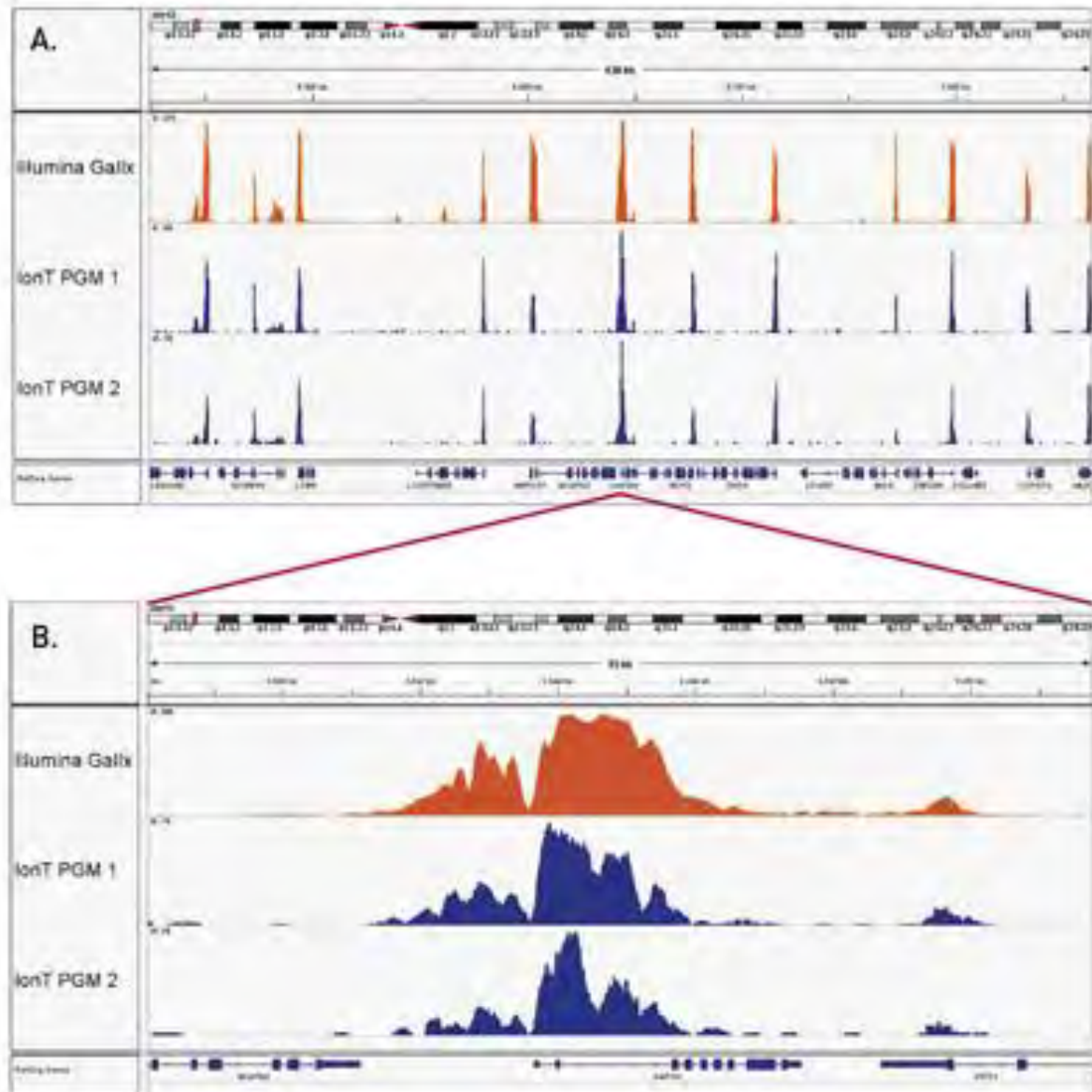


组蛋白修饰 (ChIP-Seq)



Antibody specific to
one type of histone modification

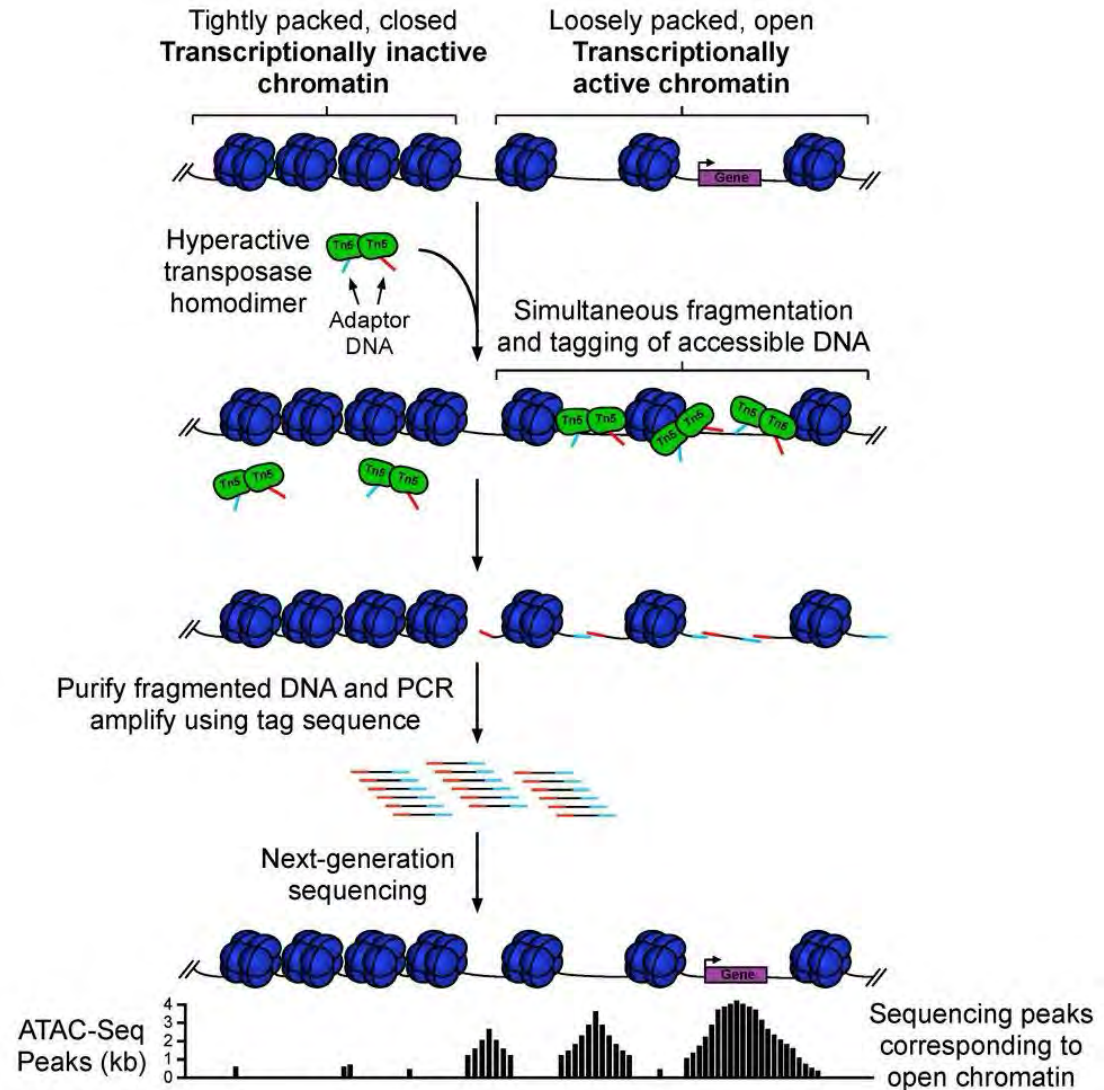
组蛋白修饰 (ChIP-Seq)



染色质可及性 (ATAC-Seq)

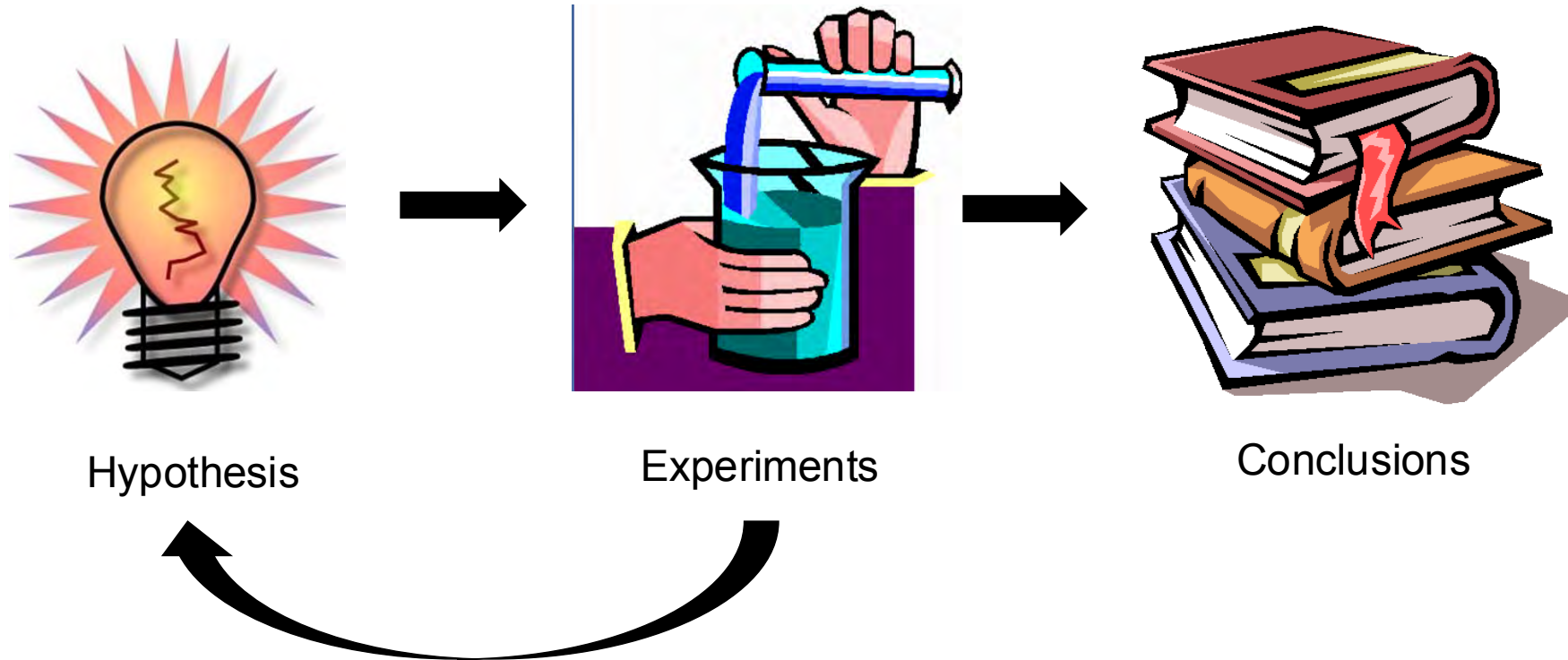
ATAC-Seq: ATAC-Seq is a popular method for determining chromatin accessibility across the genome.

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



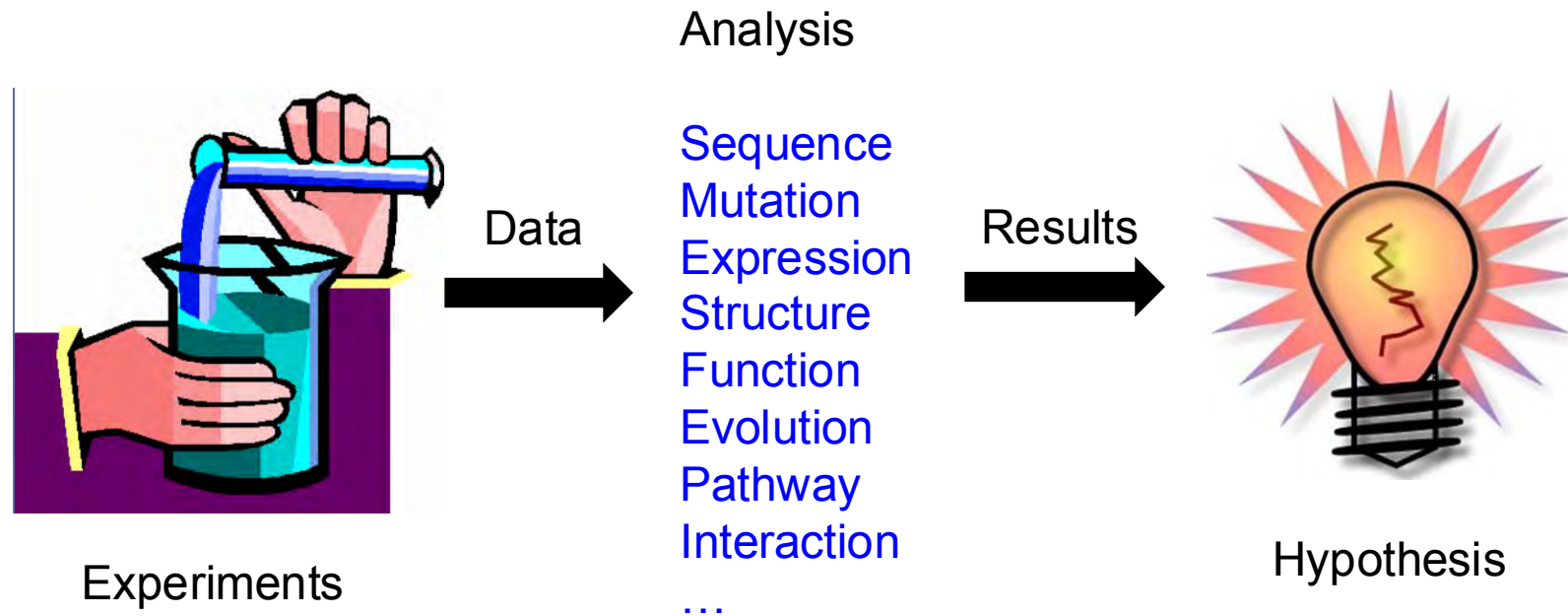
Different ways of thinking in genomic era

Hypothesis driven



Different ways of thinking in genomic era

Data driven



谢谢!