基因體學

Genomics

劉佩芬老師

高雄醫學大學
生物醫學暨環境生物學系

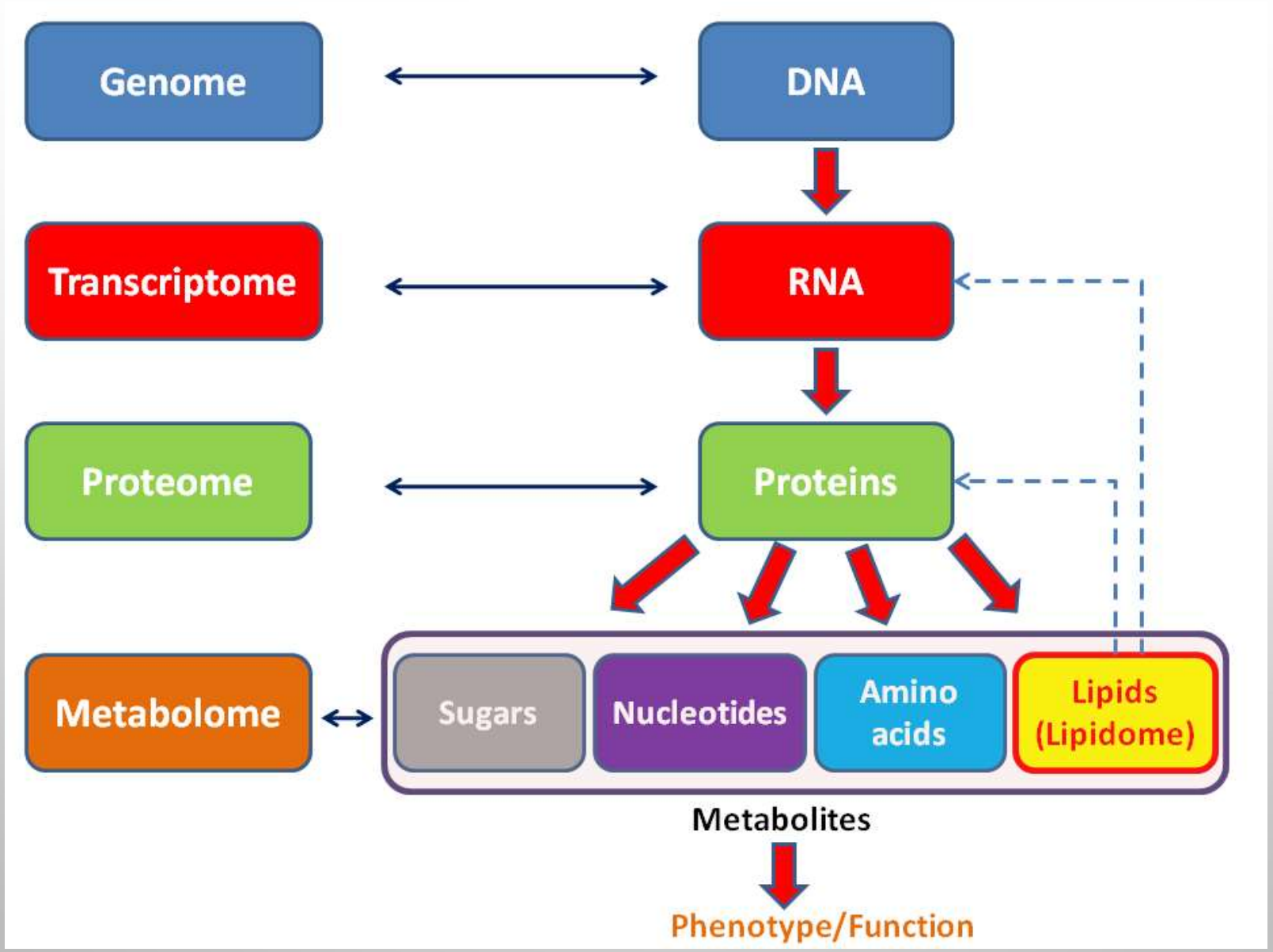
1

主題大綱

- Genomics (基因體學)
- History of DNA sequencing (DNA定序歷史)
- The first-generation DNA sequencing (第一代DNA定序)
- Next-generation sequencing (NGS,次世代定序)
- CRISPR/cas9 genome editing (基因編輯技術)
- Genome analysis (基因體分析)

Genomics
(基因體學)

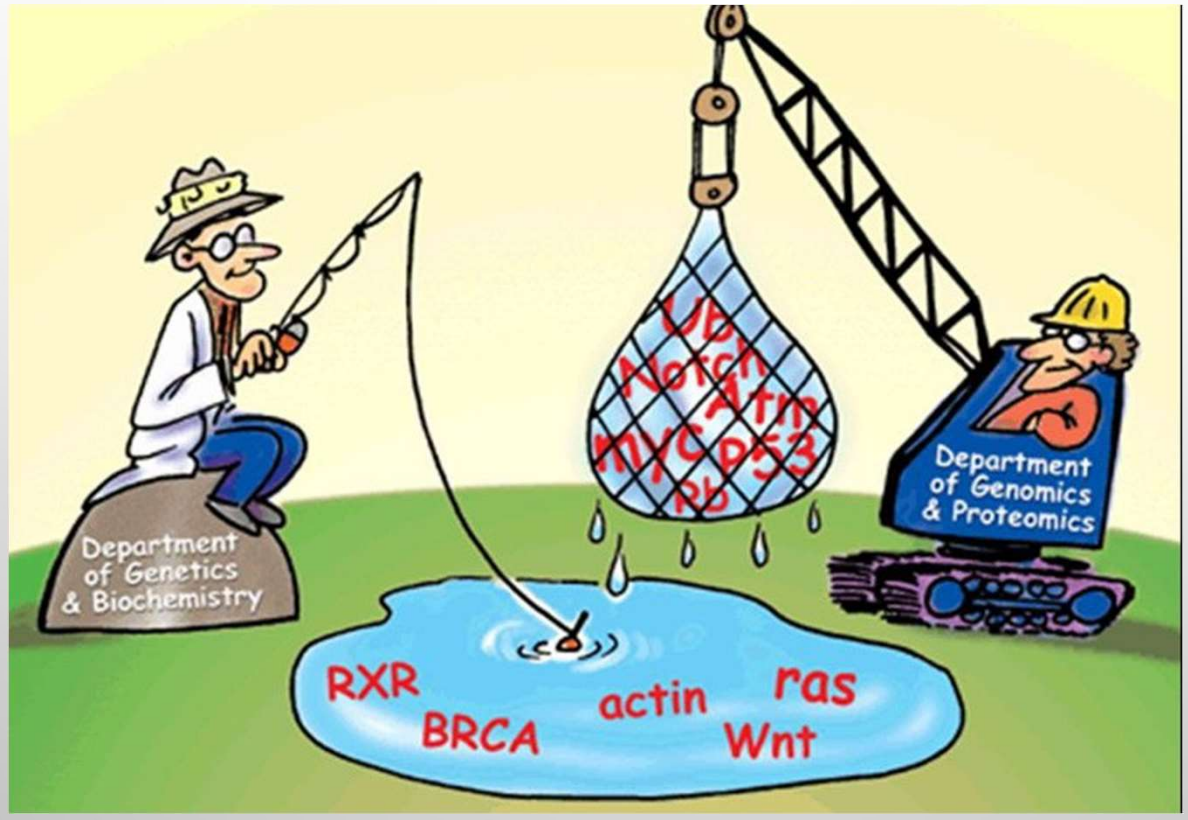
The "omics" revolution



https://en.wikipedia.org/wiki/Genomics

基因體學 (Genomics)

- 基因體(genome)：細胞內所有的DNA，包含核DNA(nuclear DNA)、葉綠體DNA、粒腺體DNA



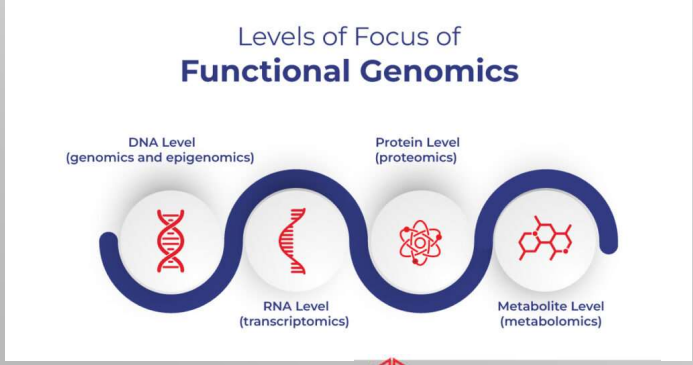
高通量(high-throughput)策略研究基因功能

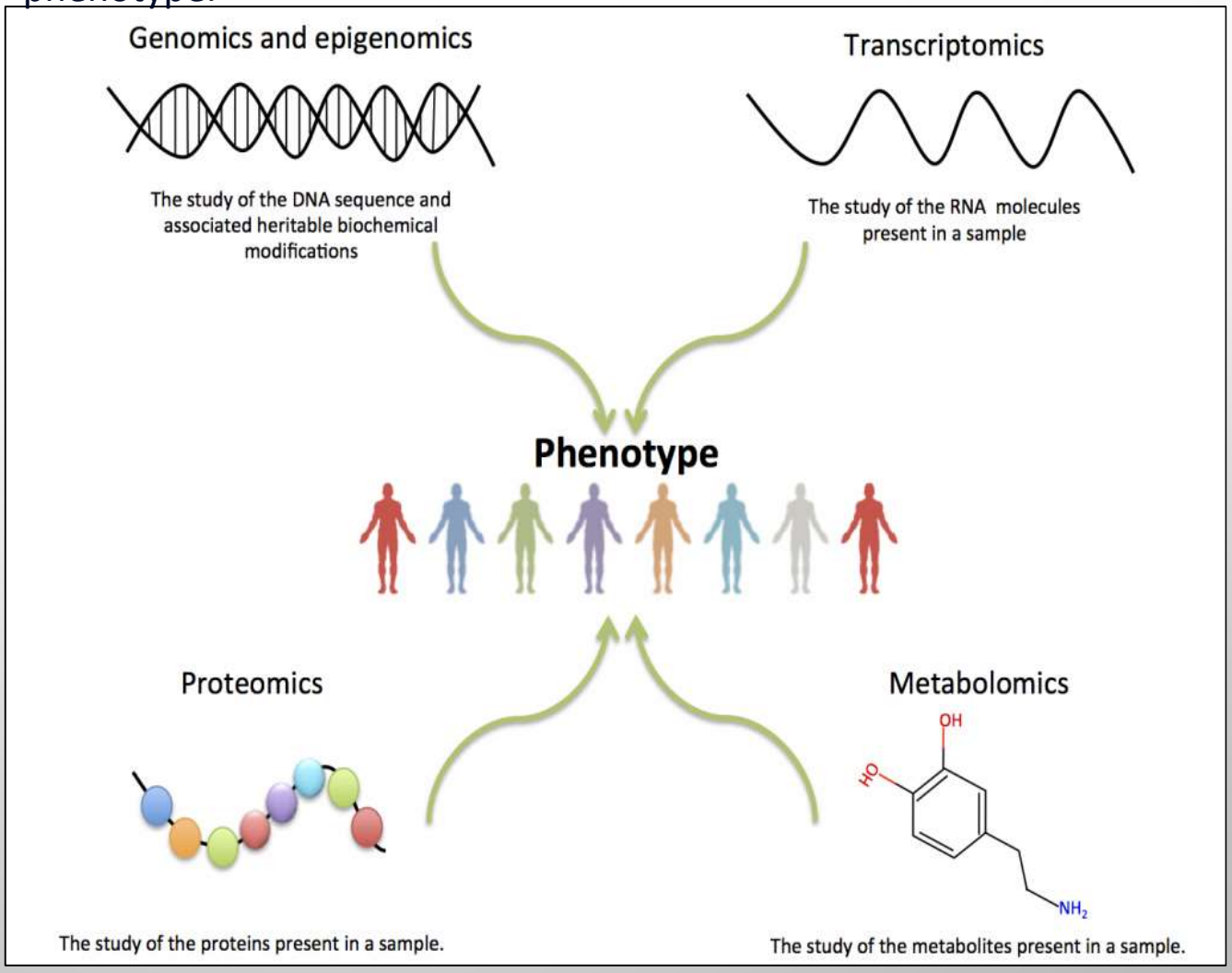
https://www.slideshare.net/uurzua/clase-genmica-presentation

功能基因體學 (Functional genomics)

The goal of functional genomics is to determine how the individual components of a biological system work together to produce a particular phenotype.

- DNA level (genomics and epigenomics)
- RNA level (transcriptomics)
- Protein level (proteomics)
- Metabolite level (metabolomics)





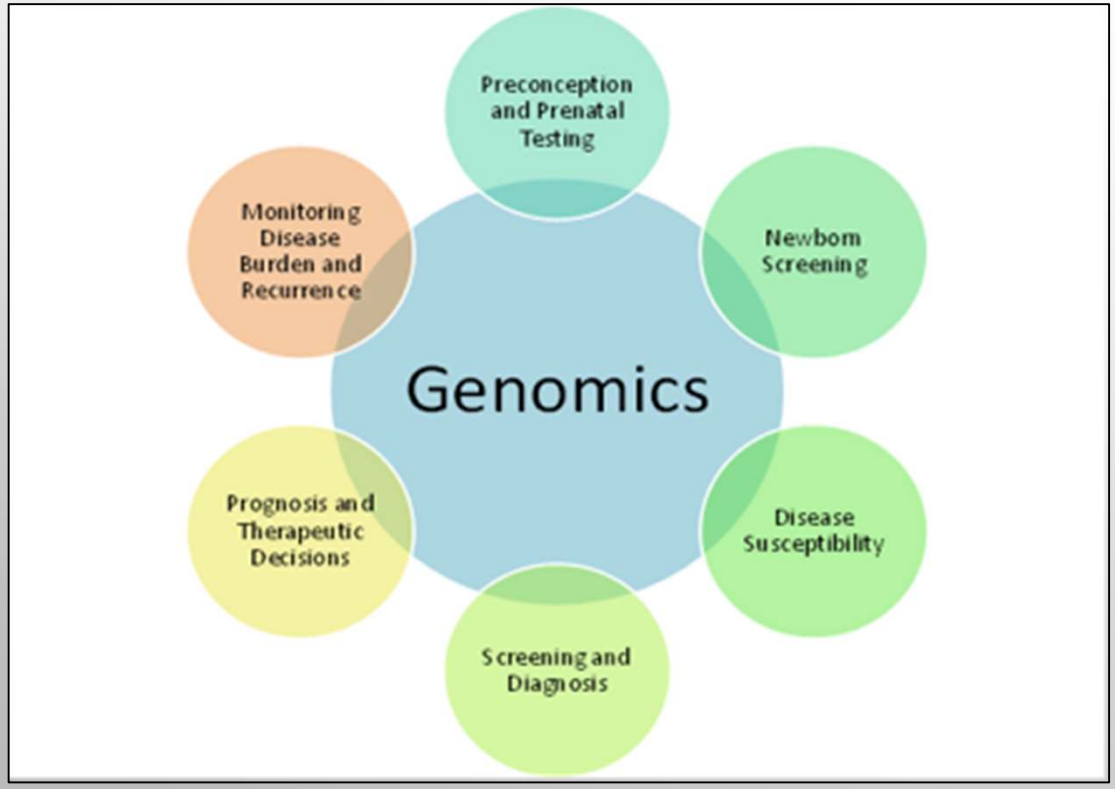
https://www.ebi.ac.uk

基因體學Applications

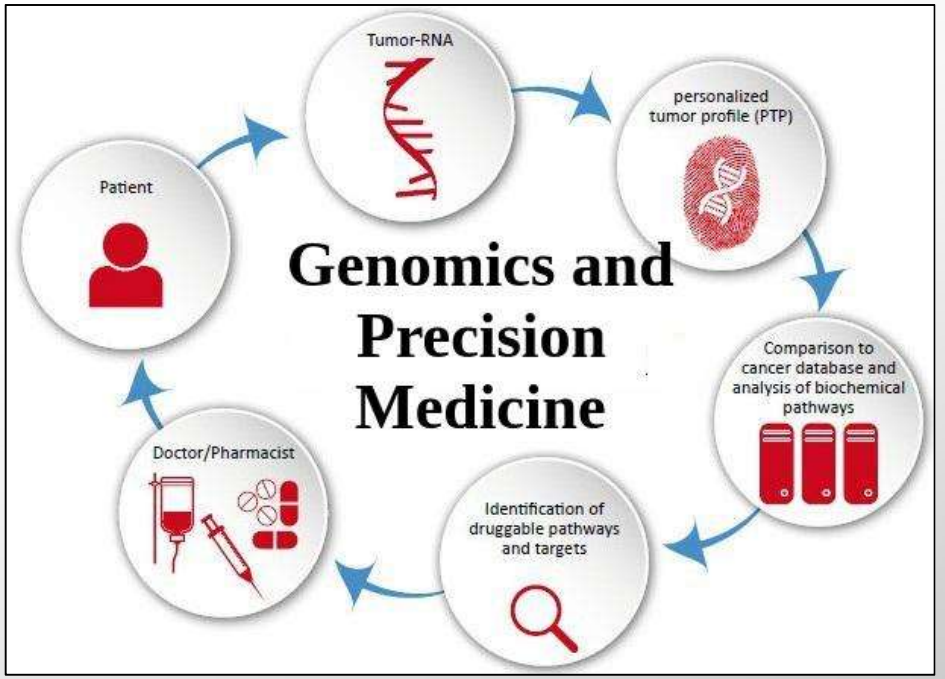
1. Gene function
2. Precision medicine
3. Genome-wide association study(GWAS)全基因組關聯分析
4. Metagenomics – 微生物菌相分析
5. Epigenomics – DNA 甲基化分析

….

單一核甘酸多型性的分析



https://www.genome.gov/minc/toolkit/why-genomics

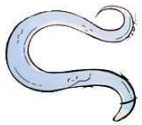


https://www.openpr.com/news/2757743/upcoming-opportunities-in-genomics-and-precision-medicine

基因體大小

- 單位(**bp / base pair**)
  1 bp = 1 bp , 1 kb = 1,000 bp,
  1GB = 1,000,000,000 bp, 1MB = 1,000,000 bp

- 基因體大小

Species	*Porcine circovirus*	*Escherichia coli*	*Caenorhabditis elegans*	*Drosophila melanogaster*	*Homo sapiens*	*Amoeba dubia*
Genome Size	1759 bp	4.6 MB	100 MB	130 MB	3.2 GB	670 GB
Common Name	 Virus	 Bacteria	 Nematode	 Fruit fly	 Human	 Ameoba
可表現蛋白基因數目	3	4288	19,000	13,600	~ 20,000	?

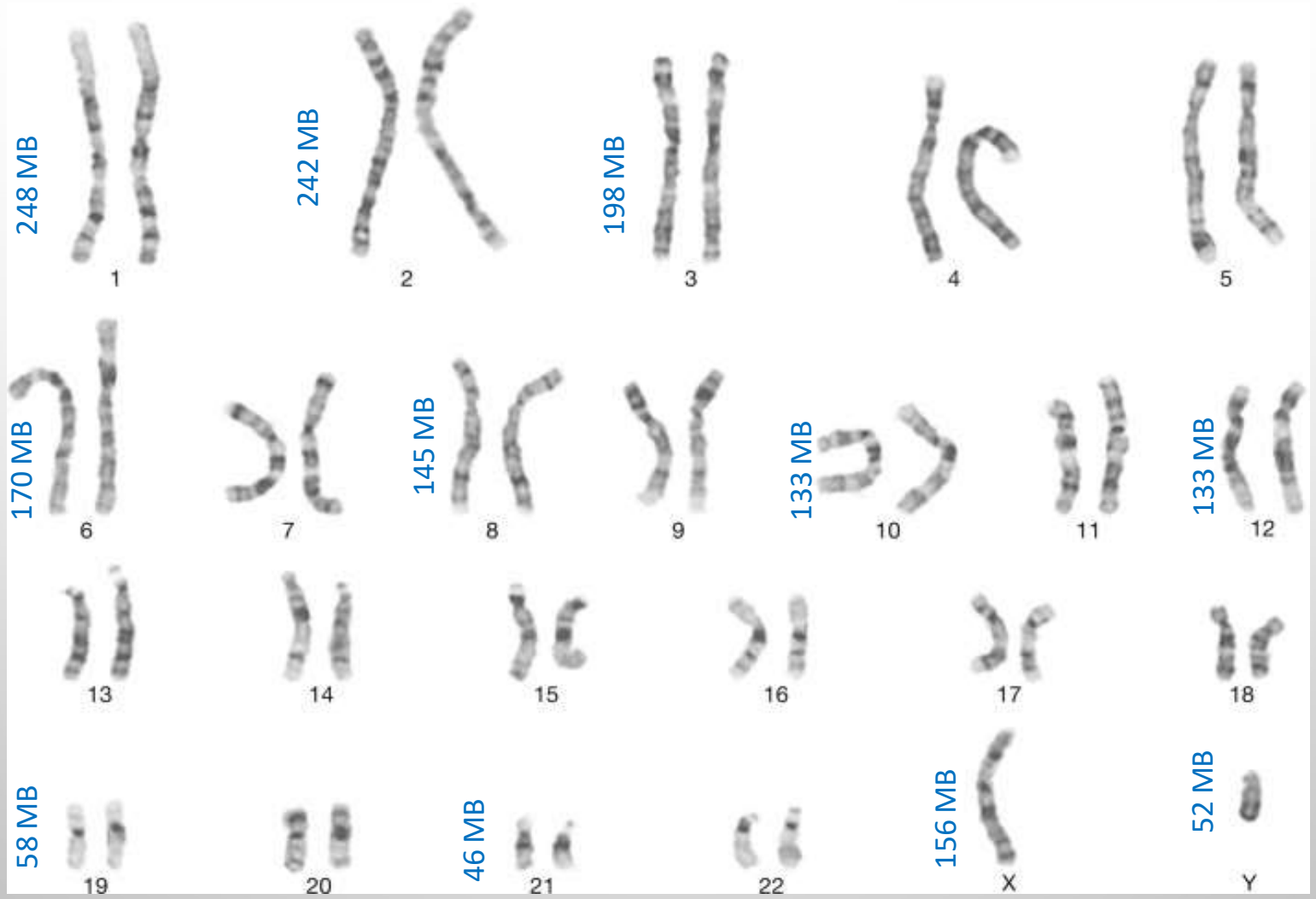
C值謎(C-value enigma): 生物的C值（基因組大小）並不與生物複雜程度相關的現象
Ex.例如植物與原生動物，可能具有比人類更大的基因組

G值謎(G-value paradox)：生物的G值（基因數量）並不與生物複雜程度相關的現象
Ex.顯微線蟲Caenorhabditis elegans由一千個細胞組成，但具有與人類大致相同數量的基因

https://old-ib.bioninja.com.au/standard-level/topic-3-genetics/32-chromosomes/genome-size.html

Chromosome (染色體)

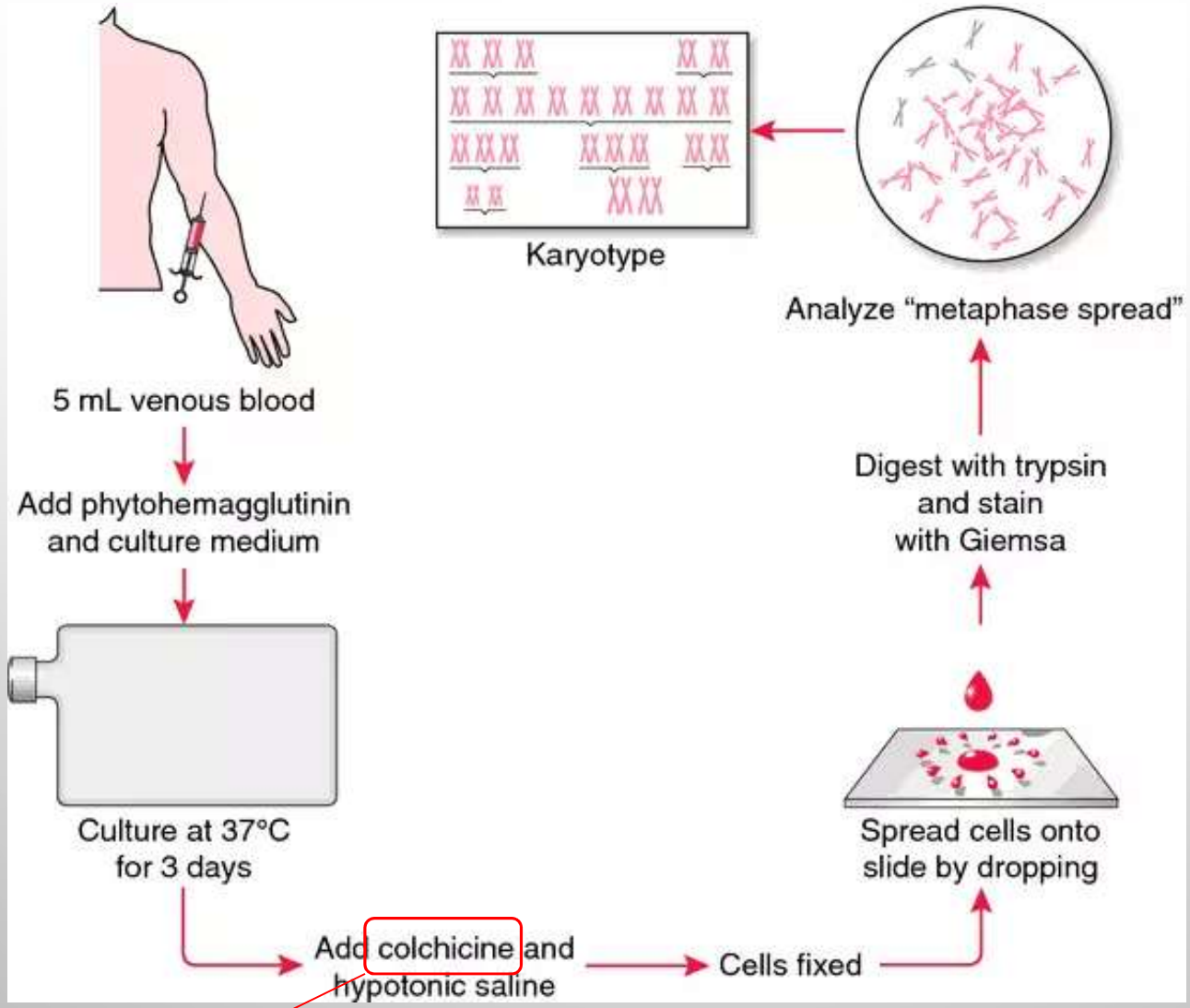
Total: 3,234.83 Mb



A.B.S Giersch, in Pathobiology of Human Disease, 2014

人類染色體

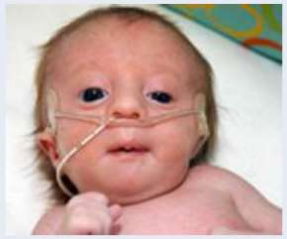
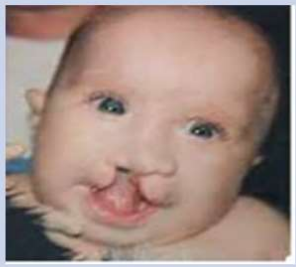
Preparation of a karyotype



. From Mueller and Young, 2001.

在有絲分裂過程中，它不會幹擾染色體的複製，但會阻止紡錘體纖維的形成。 在中期，它抑制微管的組裝，因此不會發生紡錘體形成。

染色體異常疾病

疾病	發生率	檢出率	陽性預測率
唐氏症（21 號 3 條染色體症） Down syndrome	約 1/600~1/1,000 	>99%	>99%
愛德華氏症（18 號 3 條染色體症） Edwards syndrome	約 1/3,000~1/6,000 	>99%	>98 %
巴陶氏症（13 號 3 條染色體症） Patau syndrome	約 1/5,000-10,000 	>99%	>99%

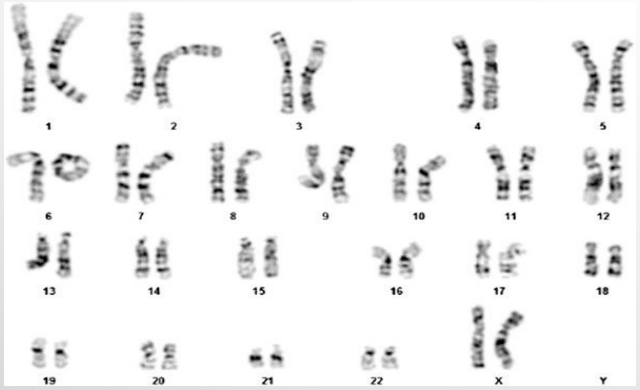
http://musoonclinic.blogspot.com/2015/04/blog-post_22.html

染色體條帶技術

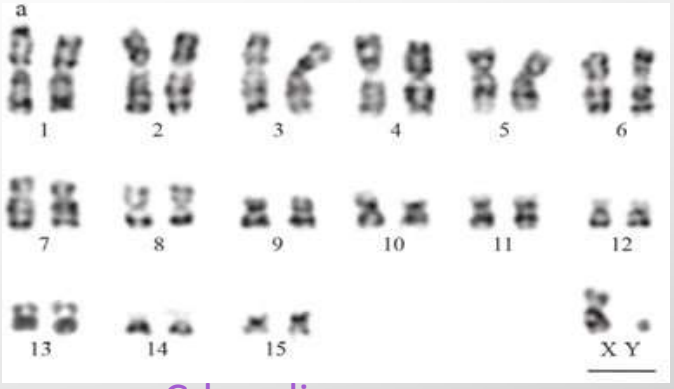
- 利用染劑使染色體呈現各自獨特條帶形態，藉以區別染色體的不同

技術	方法	亮帶	暗帶
G 帶	胰蛋白酶+ Giemsa	GC rich	AT rich
R 帶	熱磷酸鹽處理+ Giemsa	AT rich	GC rich
Q 帶	Quinacrine(螢光染料)	GC rich	AT rich
C 帶	熱鹼$Ba(OH)_2$ + Giemsa	著絲點以外	著絲點

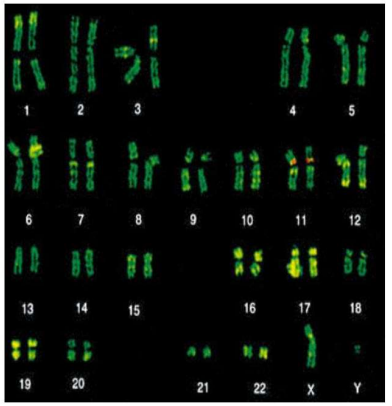
G-banding



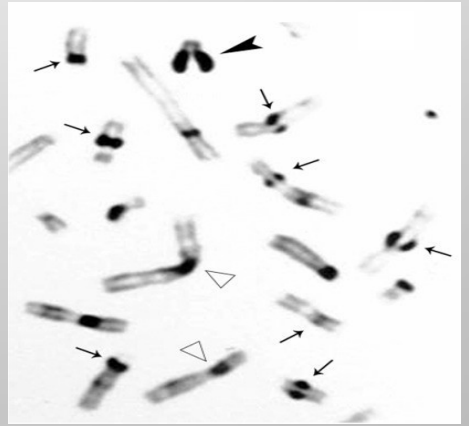
R-banding



Q-banding

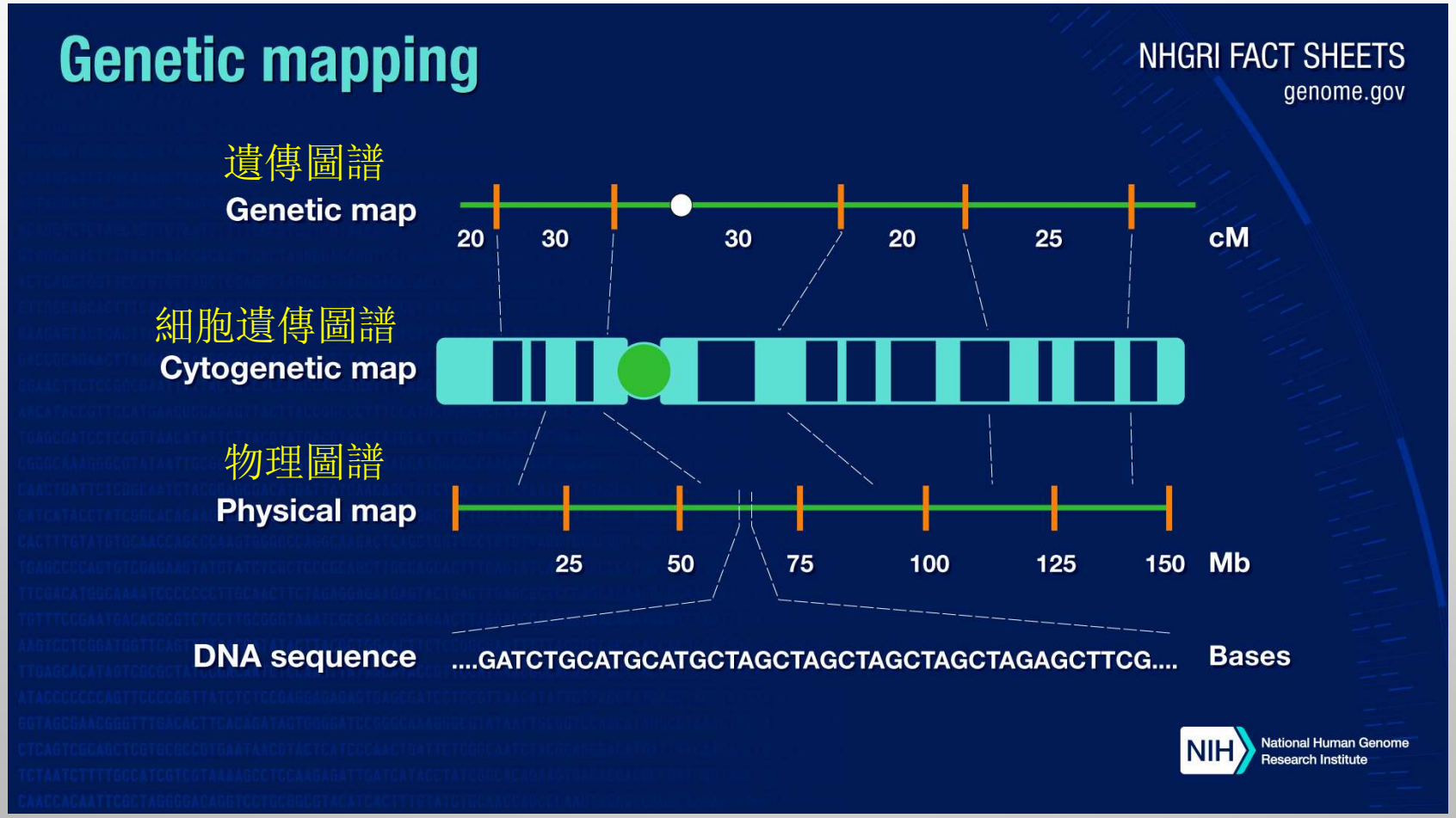


C-banding



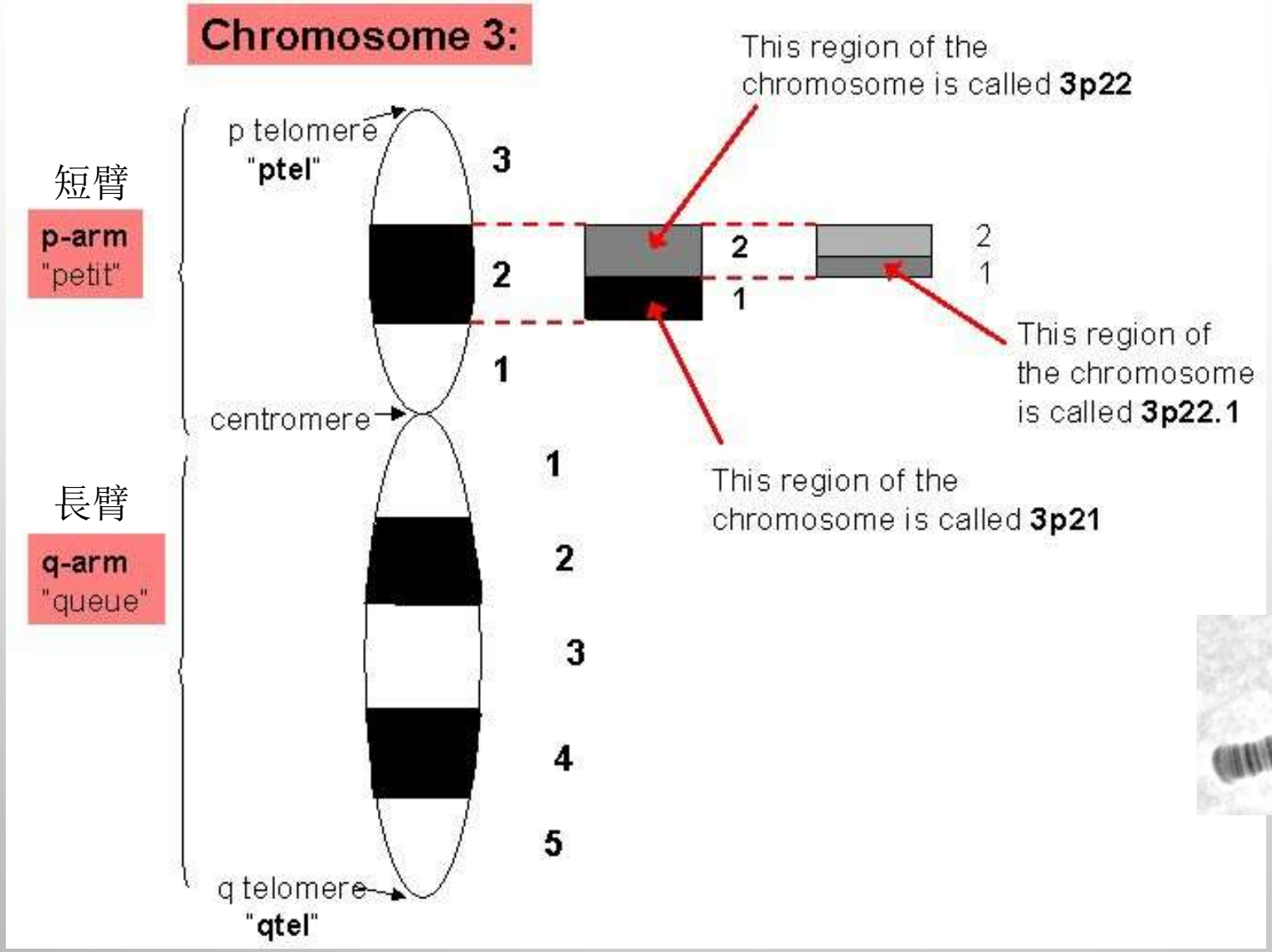
基因體圖譜 (Genome map)

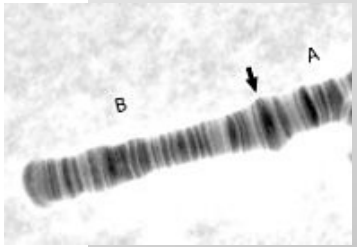
Cytogenetic map (細胞遺傳圖譜): 由染色體染色而來，沒有單位，以區域劃分
Genetic map (遺傳圖譜): 由互換率計算而來，單位cM (centimorgan)
Physical map (物理圖譜)：由序列定序而來, 單位bp



https://www.genome.gov/about-genomics/fact-sheets/Genetic-Mapping-Fact-Sheet

細胞遺傳圖譜 (Cytogenetic map)

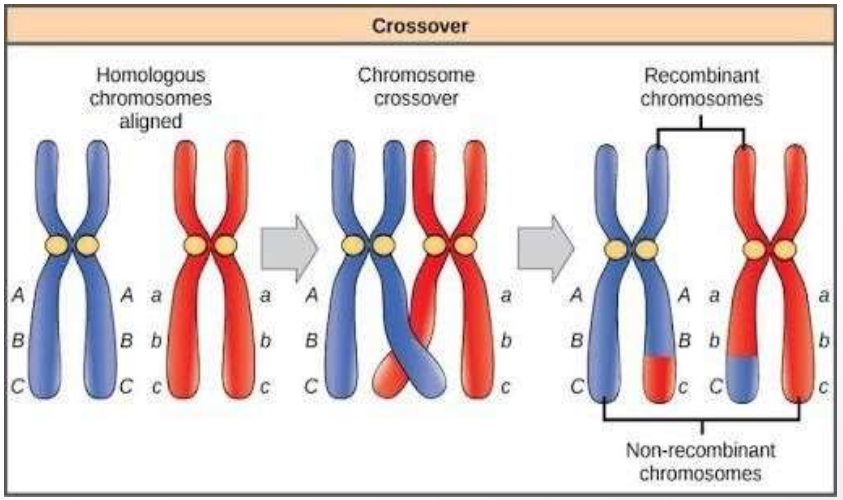


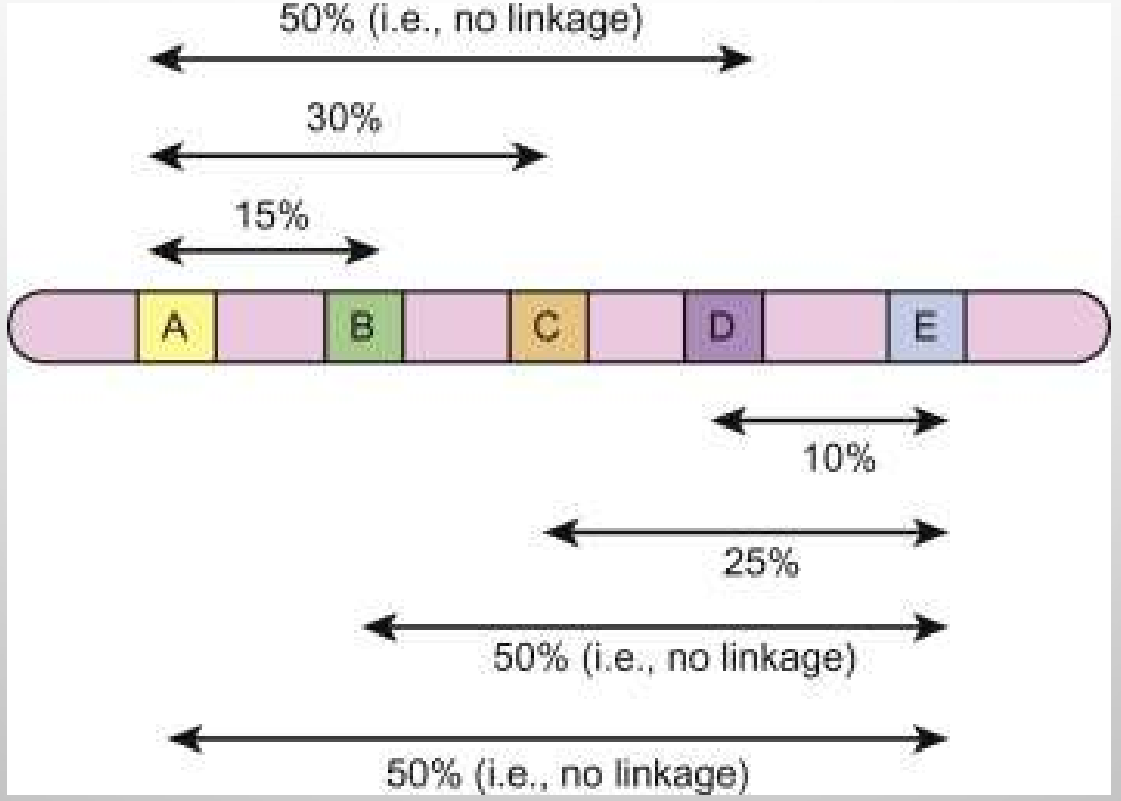


Cytogenetic mapping techniques:. An approach to genome analysis RRBS, 7(6) 2013

遺傳圖譜 (Genetic map)

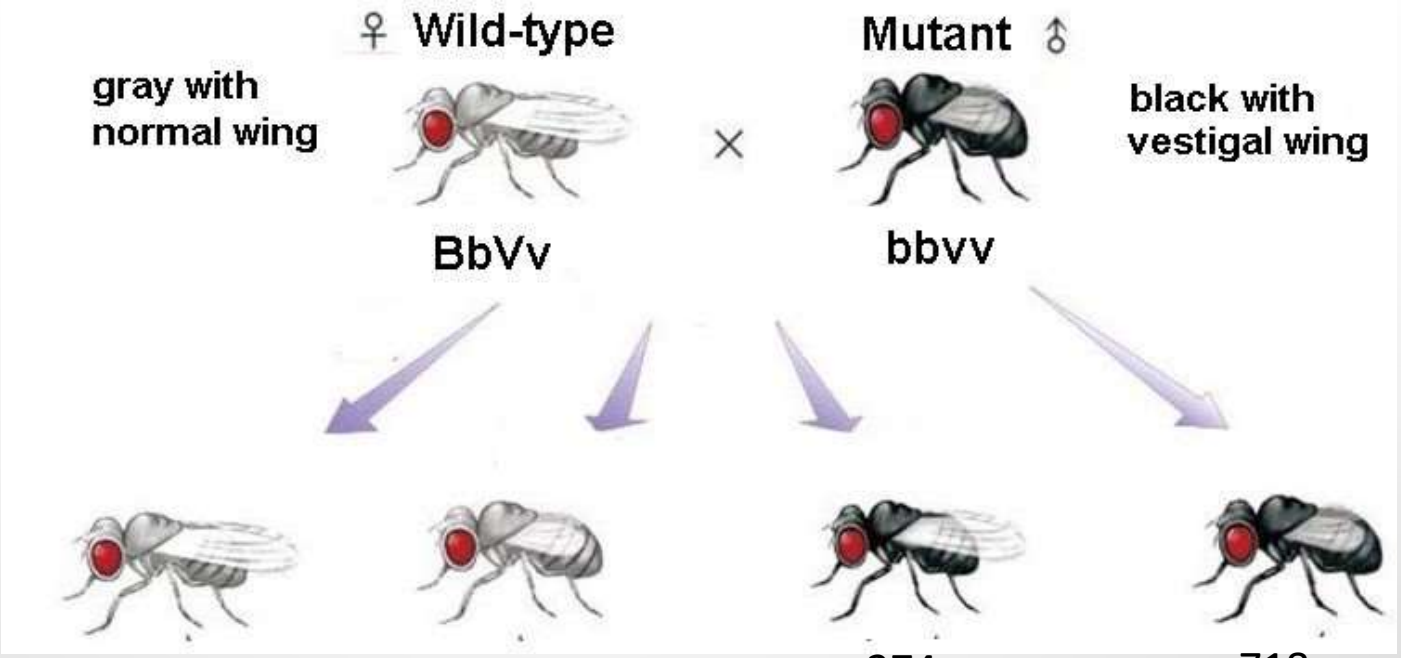
- 1% 重組率 = 1 cm (centimorgan)
- 重組率 < 50% --> " 連鎖"
- 重組率 = 0% --> "完全連鎖"
- 在人類細胞中 1 cM 約1 Mb





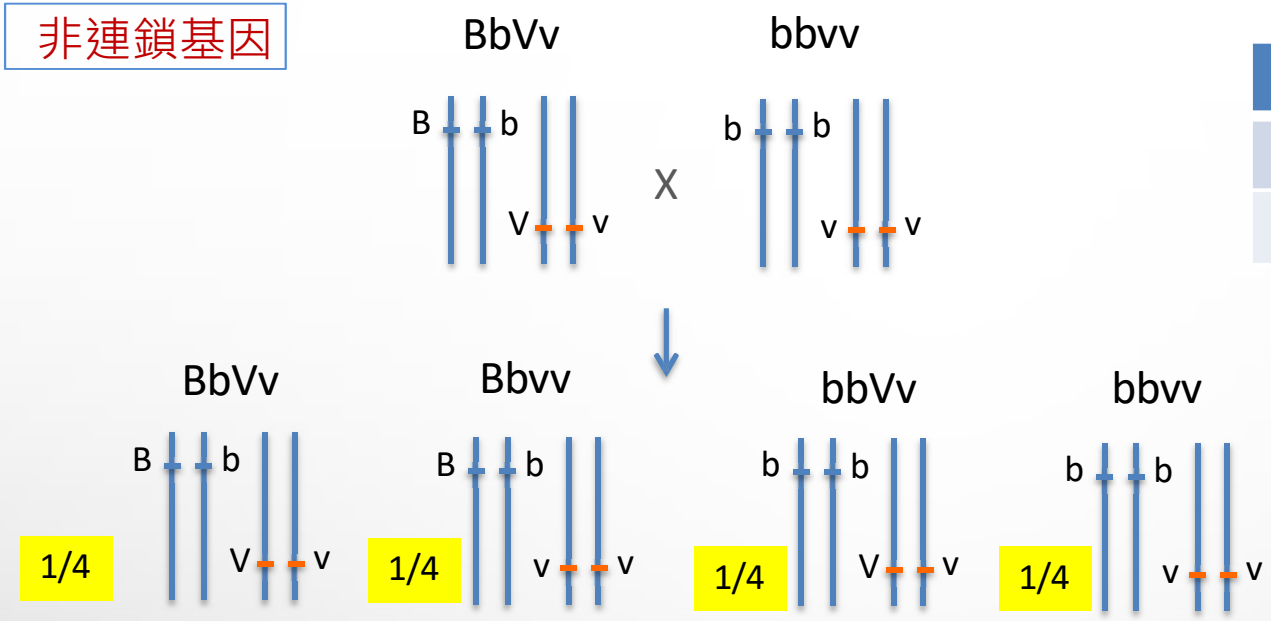
https://www.sciencedirect.com/topics/immunology-and-microbiology/genetic-linkage

染色體重組 (CHROMOSOME RECOMBINATION)



	742	266	274	718
				重組型
B, V 非連鎖基因之期望值	500 1/4	500 1/4	500 1/4	500 1/4
B, V 為完全連鎖之期望值	1000 1/2	0	0	1000 1/2

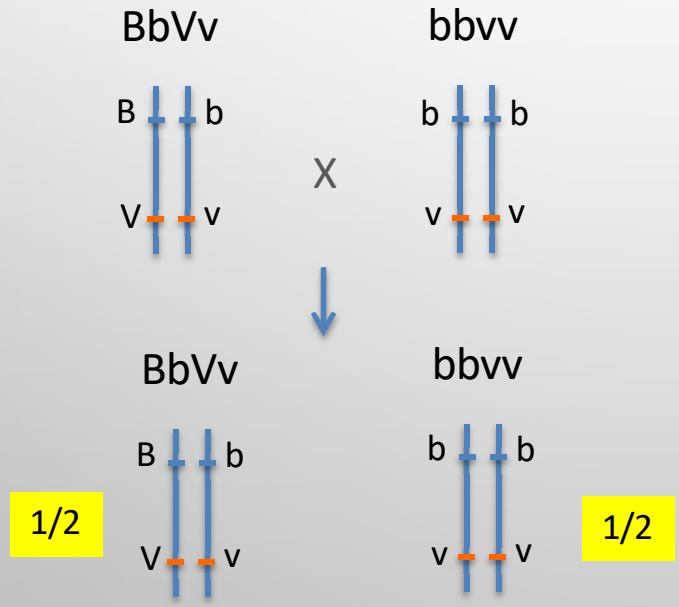
非連鎖基因



	b	b
B	Bb	Bb
b	bb	bb

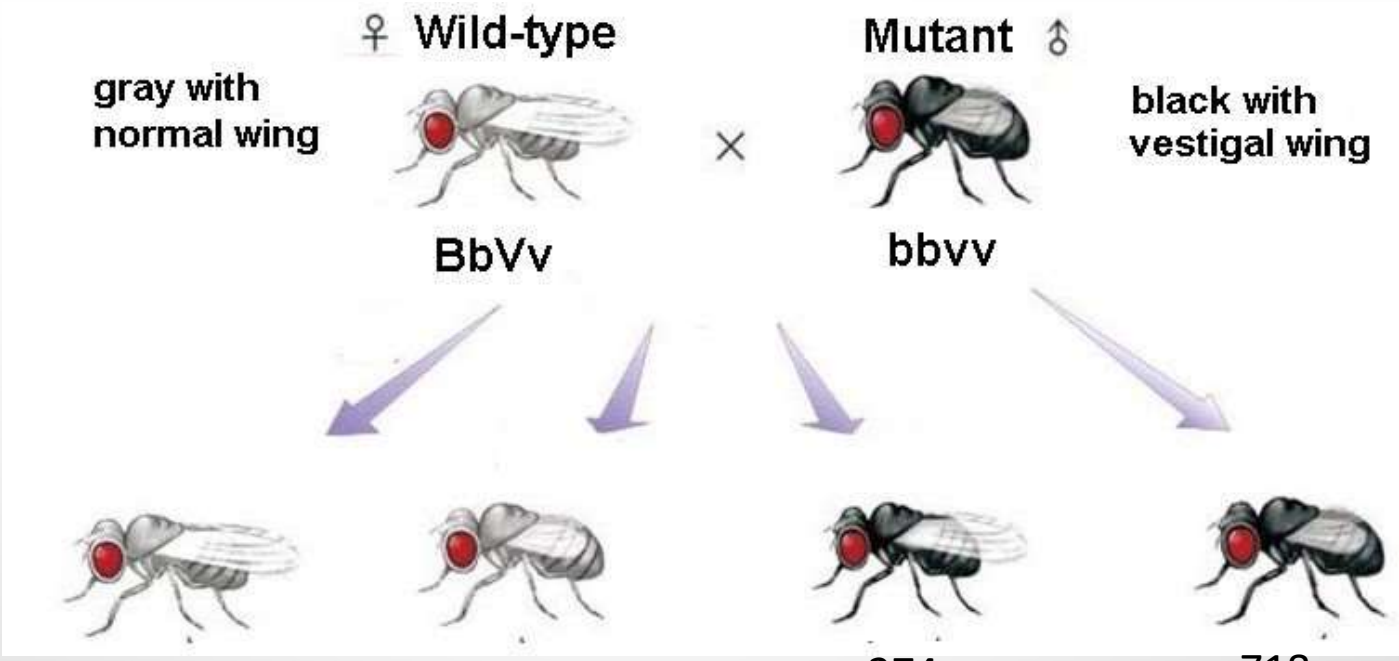
	v	v
V	Vv	Vv
v	vv	vv

基因連鎖



	BV	bv
bv	BVbv	bvbv
bv	BVbv	bvbv

RECOMBINATION RATE (重組率)



重組率= 重組型/ 所有子代

= (266 + 274) / 2000 = 0.27

兩基因距離為 **27 centimorgan (cM)**

*重組率最高為50%

*重組率的精準度和樣品數量成正相關

https://www.macmillanhighered.com

DNA 載體 (DNA Vector)

載體	承載量	宿主細胞
人類人造染體(HAC) 發表於1997年	6000 - 10000 Kb	human cell
酵母人造染色體(YAC) 於1983年建立	100 - 3000 kb	Yeast
細菌人造染色體(BAC)	150 ~ 350 kb	E. Coli
噬菌體載體(PAC)	100- 300 kb	E. Coli
黏質體(Cosmid, 噬菌體載體/質體之複合體)	35-45 kb	E. Coli
質體(plasmid)	<= 15kb	E. Coli

如果以一倍的覆蓋率計算，人類基因組(3,200,000 kb) 需要

320 HAC
1,066 YAC
9,142 BAC
10,666 PAC
71,111 Cosmid
213,333 plasmid

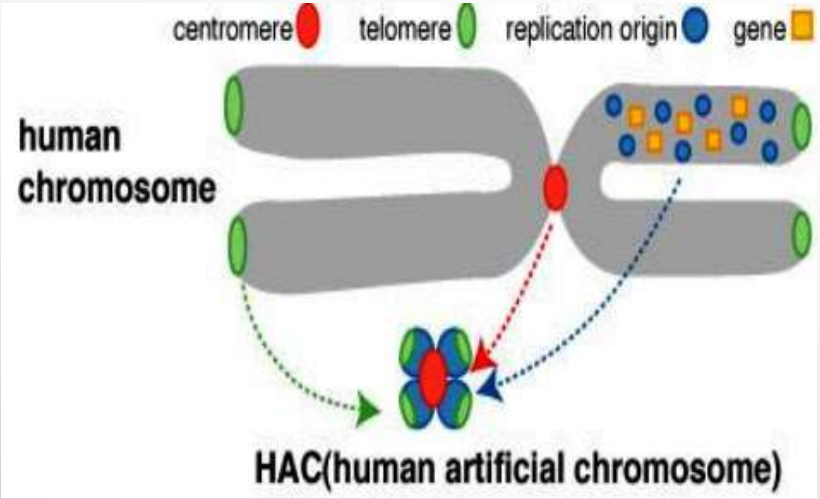
* 通常一個基因組庫的要求為6倍覆蓋率以上

https://www.youtube.com/watch?v=buAIHv6YOHQ

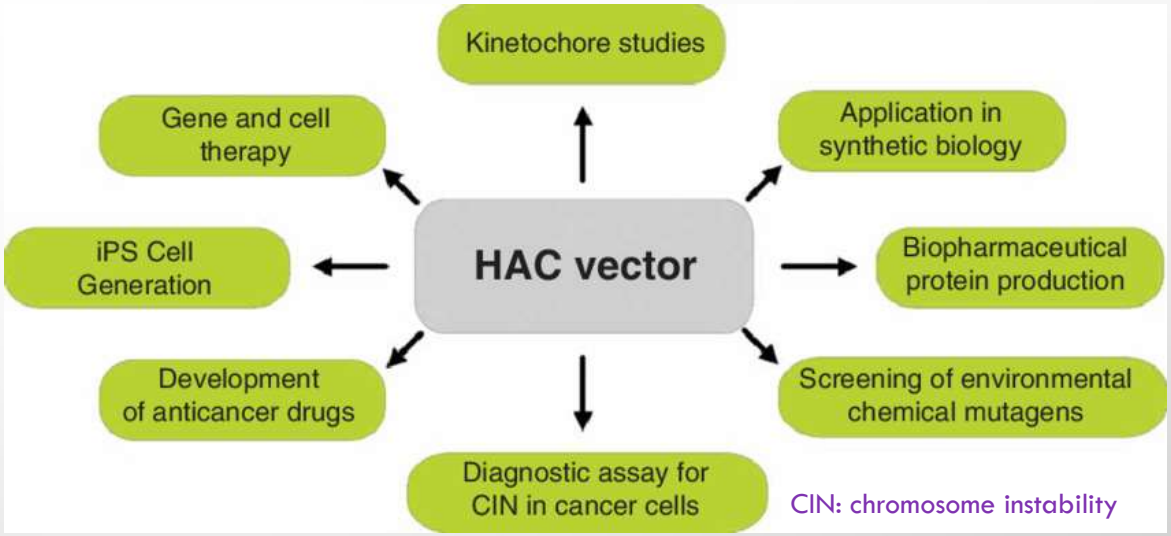
https://en.wikipedia.org/wiki/Cosmid

人類人造染體(Human Artificial Chromosome,HAC)

人類人工染色體是一種小型染色體，可作為載體搭載一些基因，並作為人類細胞中額外的染色體，使這些基因表現於人類體內。



https://plantbreeding2010.blogspot.com/2022/11/a-brief-account-of-yac-bac-hac.html

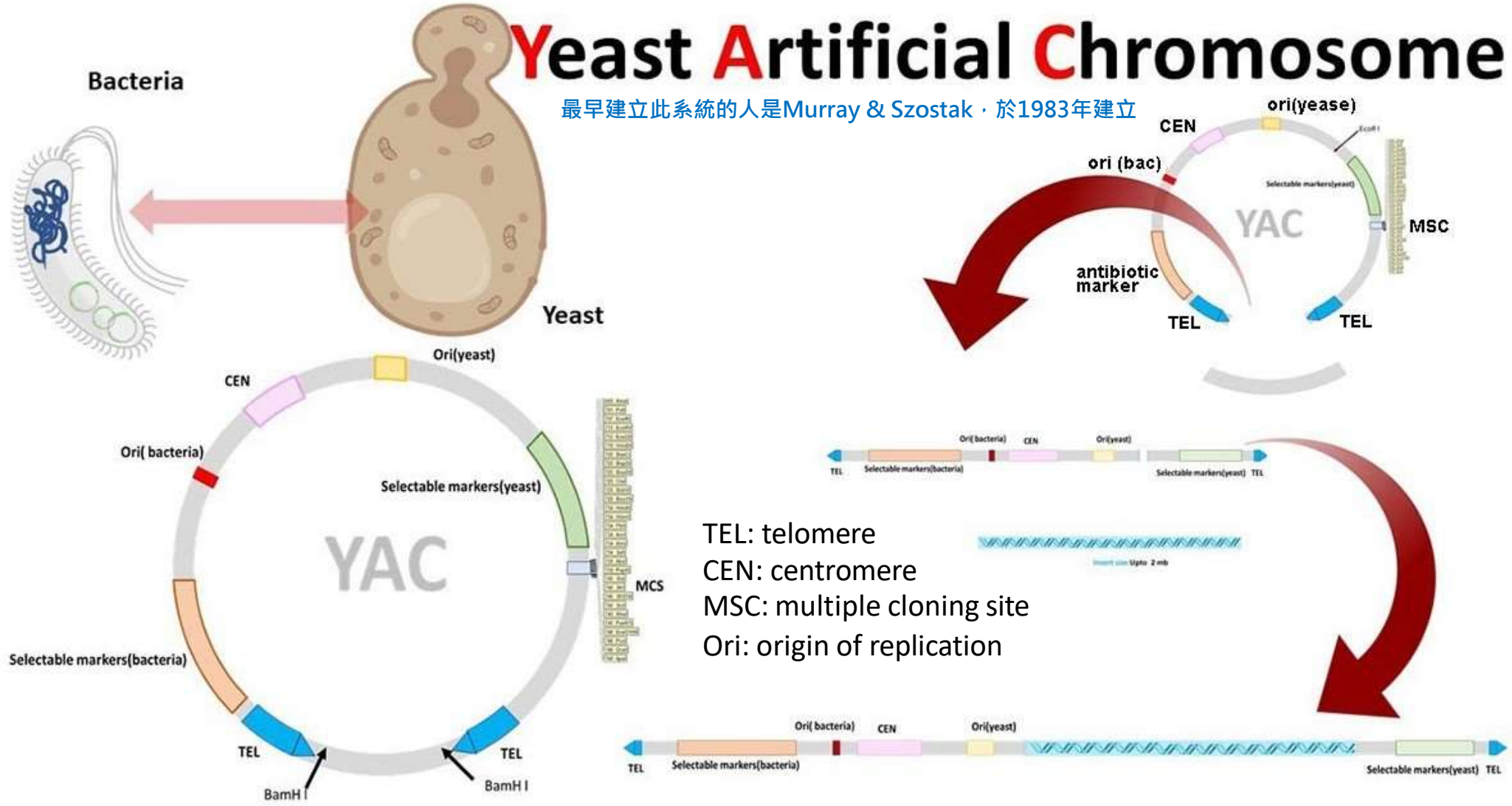


•DOI: 10.1517/17425247.2014.882314

HACs have been used in the development of approaches to the reprogramming of cells into iPSCs, creation of transgenic animals, and the generation of experimental models for the treatment of genetic diseases. HACs have been also extensively used to study chromosome functions and chromosomal instability.

人類人工染色體 成功植入老鼠
https://www.youtube.com/watch?v=tNQdTzOQtel

酵母人造染色體(Yeast Artificial Chromosome,YAC)



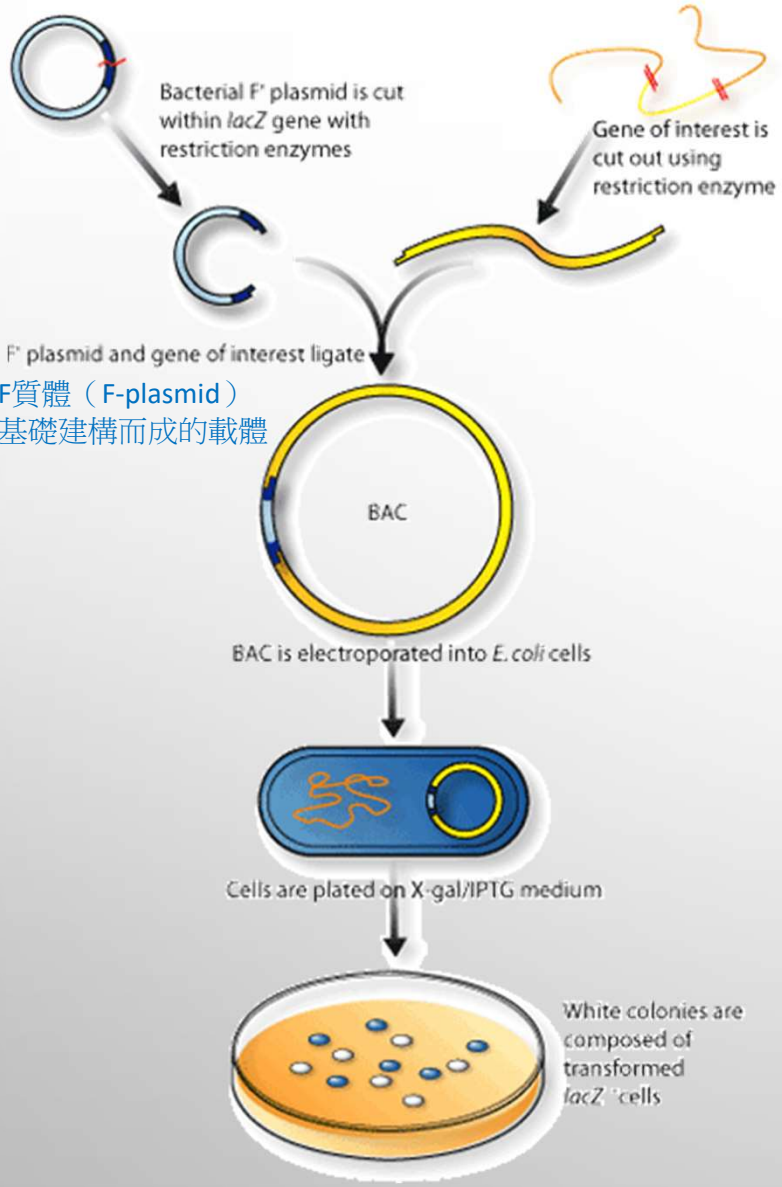
Capacity: 100 -3,000 kb

https://www.youtube.com/watch?v=ZvLPTNYpQX8

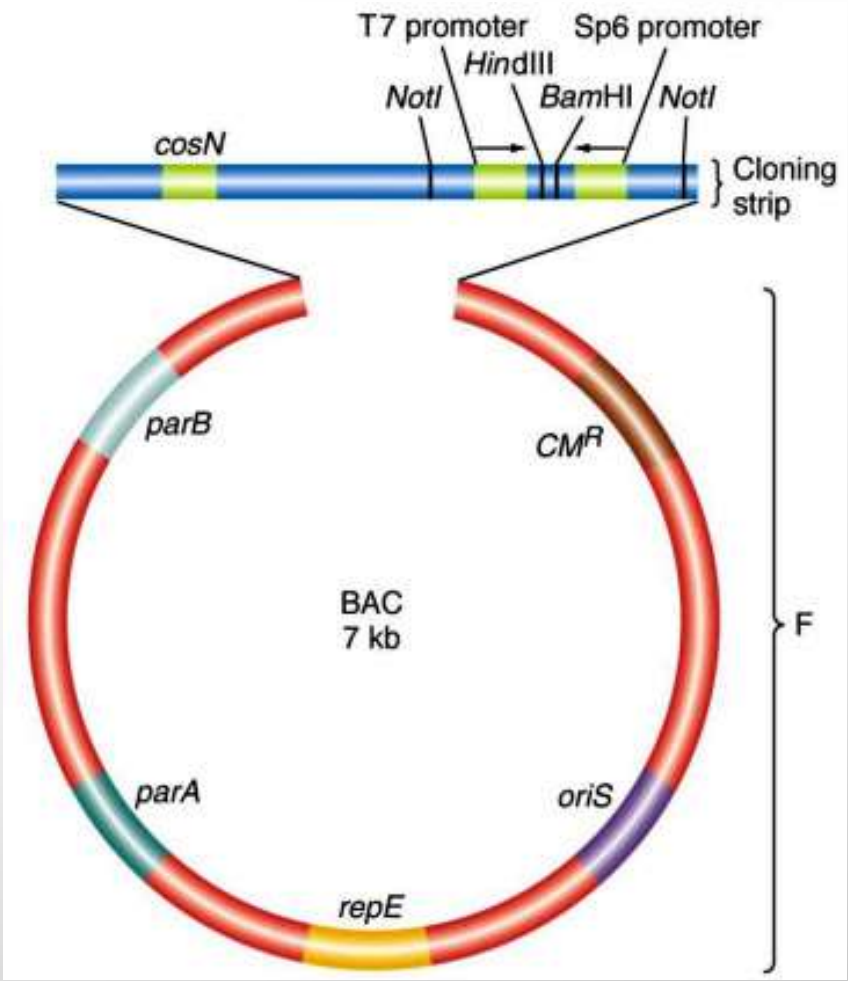
缺點: 其大小與內源的染色體的大小相近，就很難從中分離出來，不利於進一步分析。

優點:為高容量選殖載體。酵母細胞比大腸桿菌對不穩定的、重複的和極端的 DNA 有更強的容忍性。由於高等真核生物的基因大多數是多外顯子結構並且有長的內含子，大型基因組片段可通過 YAC 載體轉移到動物或動物細胞系中，進行功能研究。

細菌人造染色體(Bacterial Artificial Chromosome, BAC)



以F質體（F-plasmid）為基礎建構而成的載體

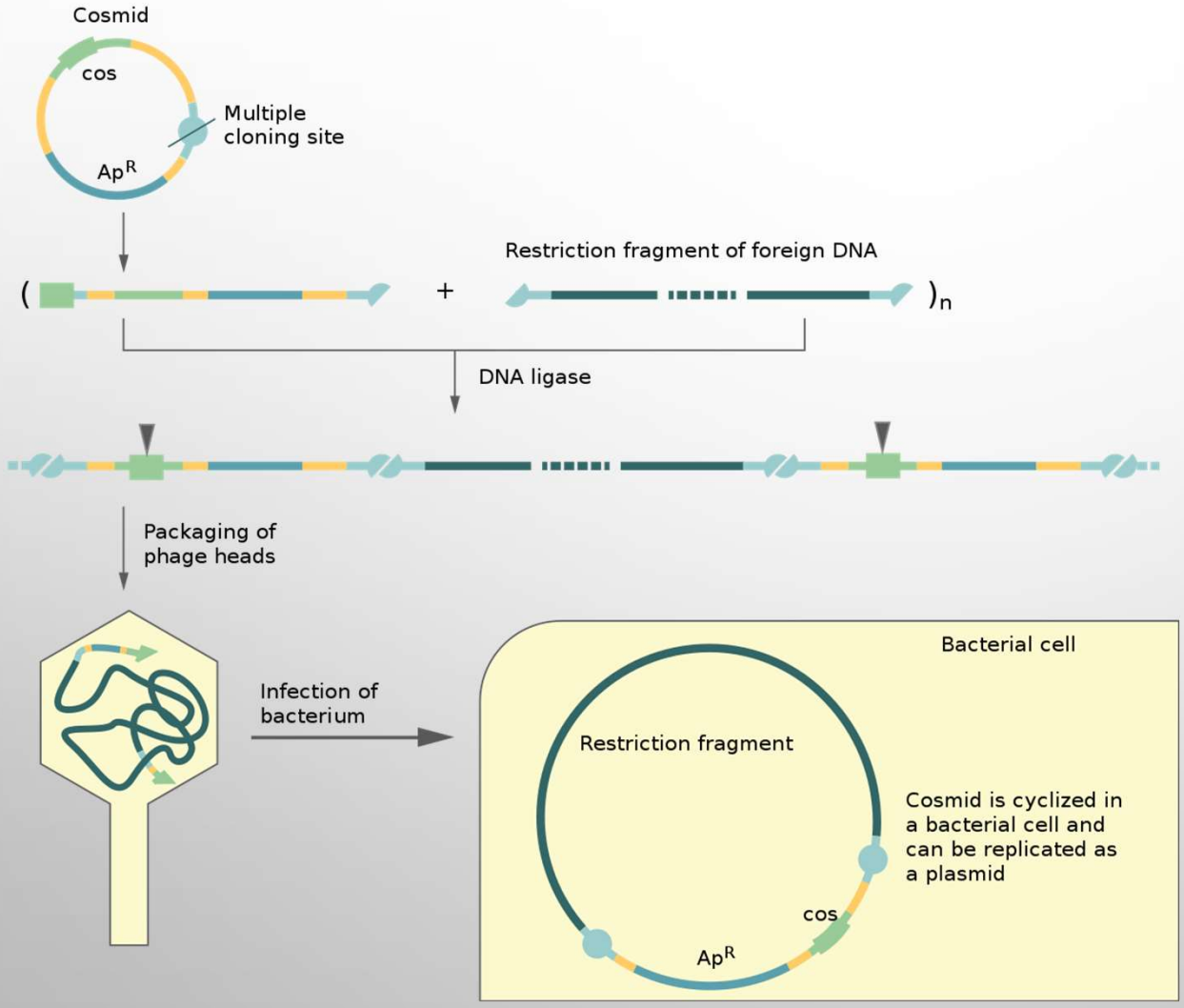


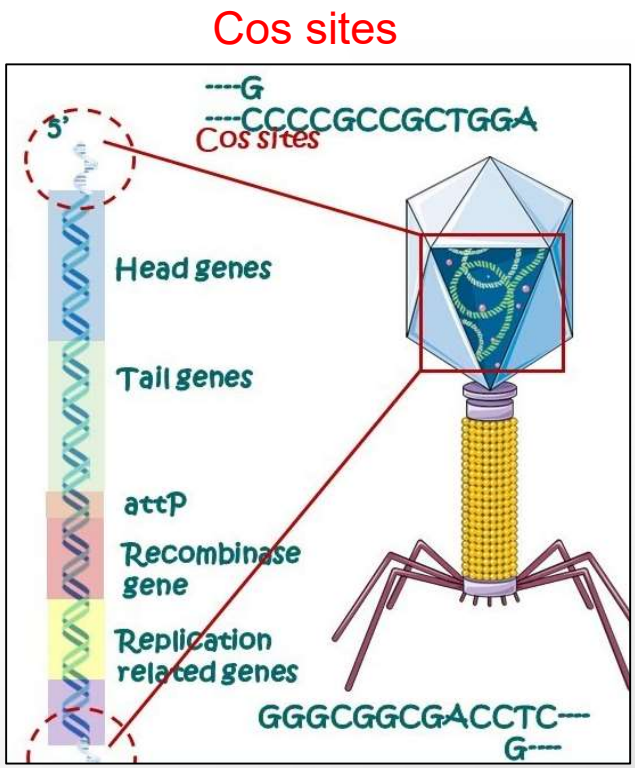
https://plantbreeding2010.blogspot.com/2022/11/a-brief-account-of-yac-bac-hac.html

https://www.youtube.com/watch?v=Qqsw6ytGsyE

黏質體(Cosmid)

黏質體是一種以λ噬菌體中的cos sequences所建構而成的質體，是常用的選殖載體之一，可用於建構基因組庫。

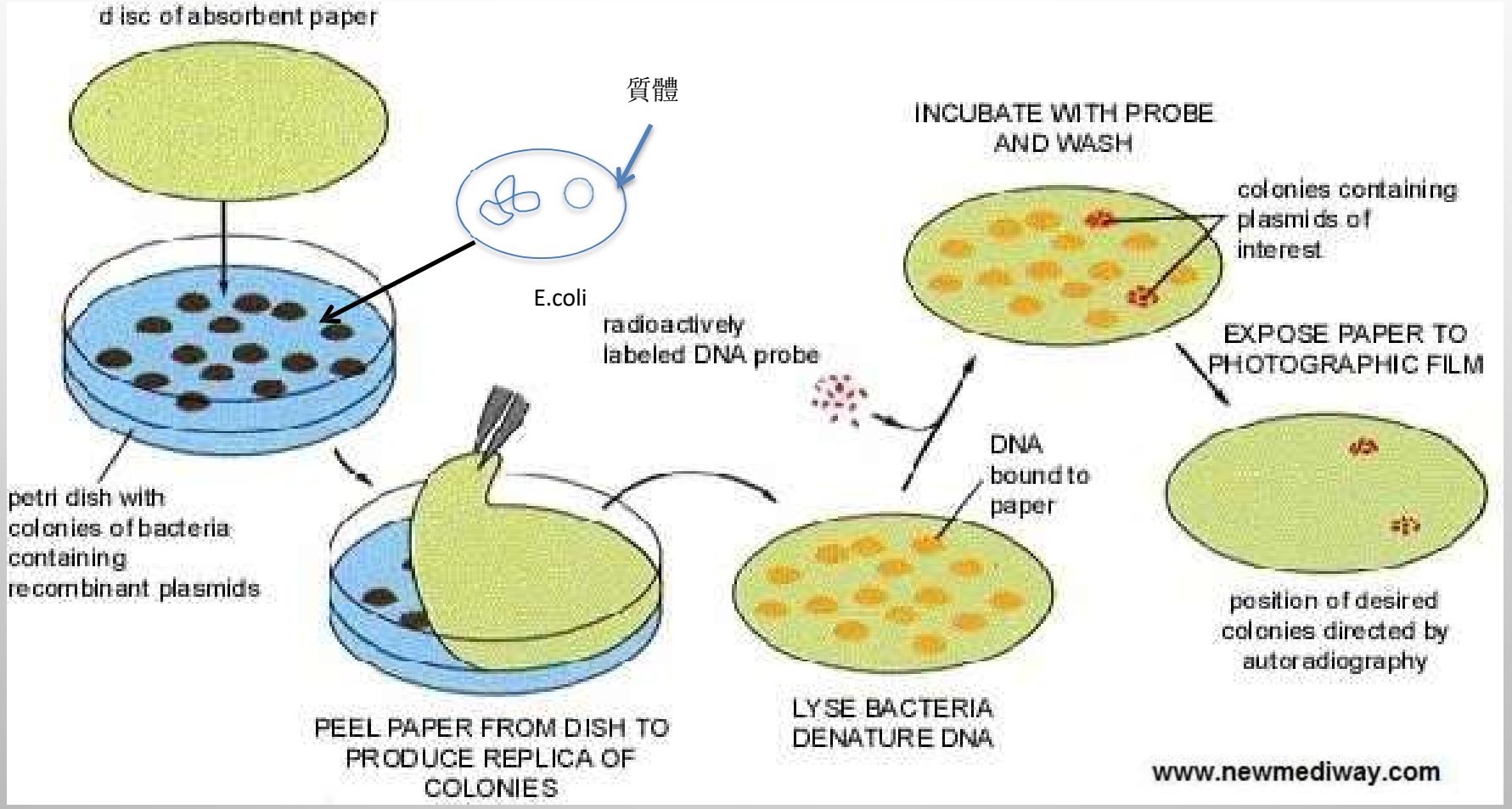




https://www.youtube.com/watch?v=1lqQn3_PvMs

菌落雜交(Colony Hybridization)

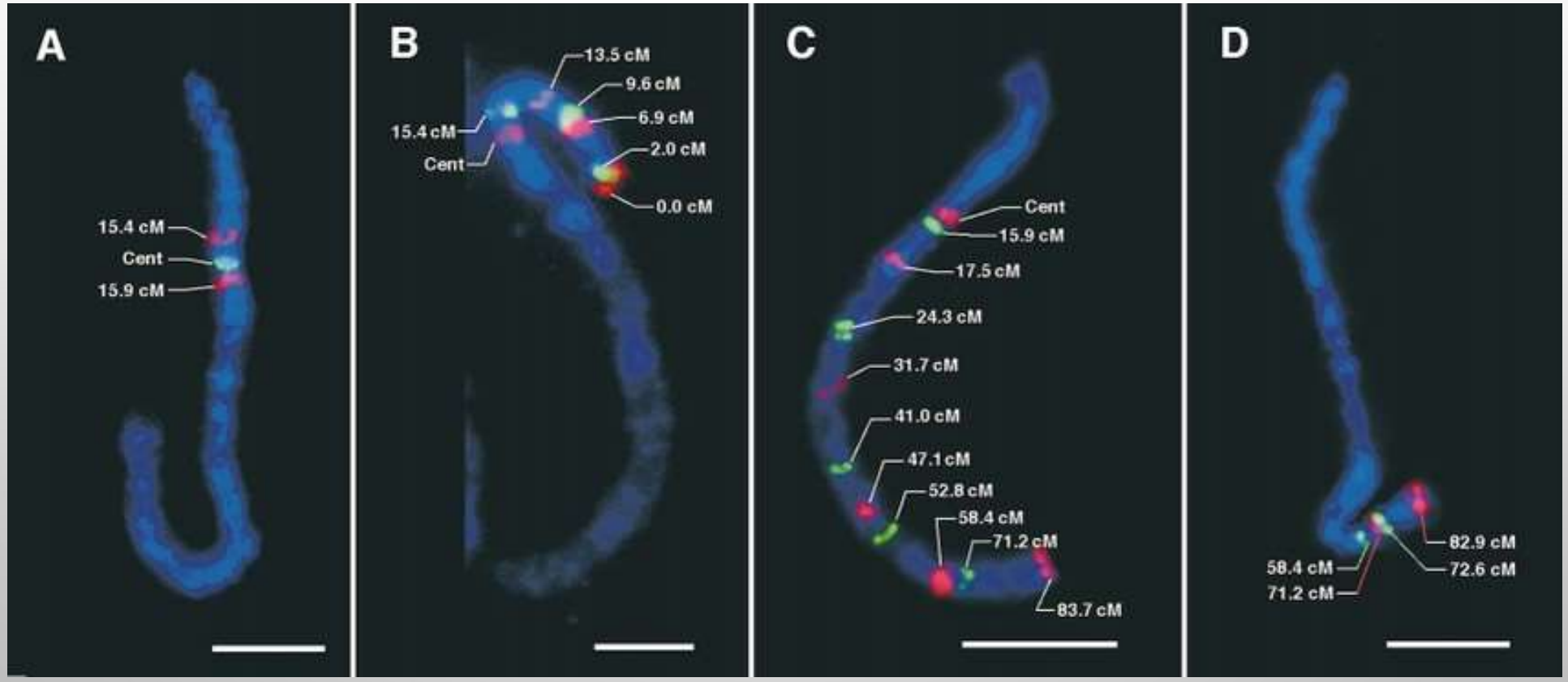
- 主要原理為單股DNA可和放射性探針(probe,單股DNA)結合
- 目的為尋找含有特定序列的菌落
- 培養菌落→ 拓印至硝化纖維膜→ 鹼破壞打破細胞並使DNA變 性→放射性探針雜 合-> 訊號偵測



https://www.youtube.com/watch?v=FtNreXY7poA

螢光原位雜合FISH
(Fluorescence in situ hybridization)

- 主要原理為單股DNA可和螢光標定探針(probe,單股DNA)結合
- 目的為確認目標序列在染色體上的位置
- 細胞固定於玻片→ 以formamide將染色體變性→螢光標定探針雜合→螢光顯微鏡觀察



BMC Genomics2011**12**:639

https://www.youtube.com/watch?v=LiRJoTi44TA

History of DNA sequencing

History of DNA sequencing

1869 – Discovery of DNA

1909 – Chemical characterization

1953 – Structure of DNA solved

1977 – Sanger sequencing invented
- Sequencing by degradation (Maxam-Gilbert)
- First genome sequenced – ΦX174 (5 kb)
  Enterobacteria phage phiX174

1986 – First automated sequencing machine

1990 – Human Genome Project started

1992 – First "sequencing factory" at

The Institute for Genomic Research (TIGR)

1995 – First bacterial genome – *H. influenzae* (1.8 Mb)

1996 – Pyrosequencing

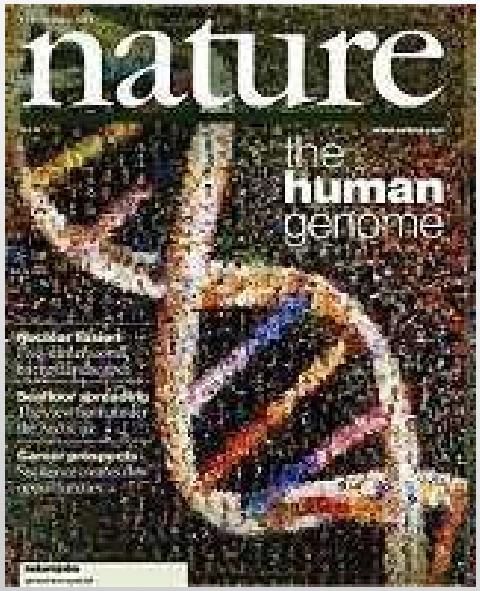
1998 – First animal genome – *C. elegans* (97 Mb)

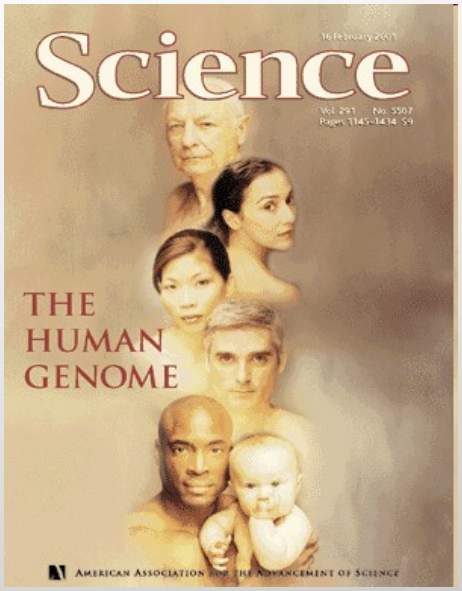
2003 – Completion of Human Genome Project (3 Gb)

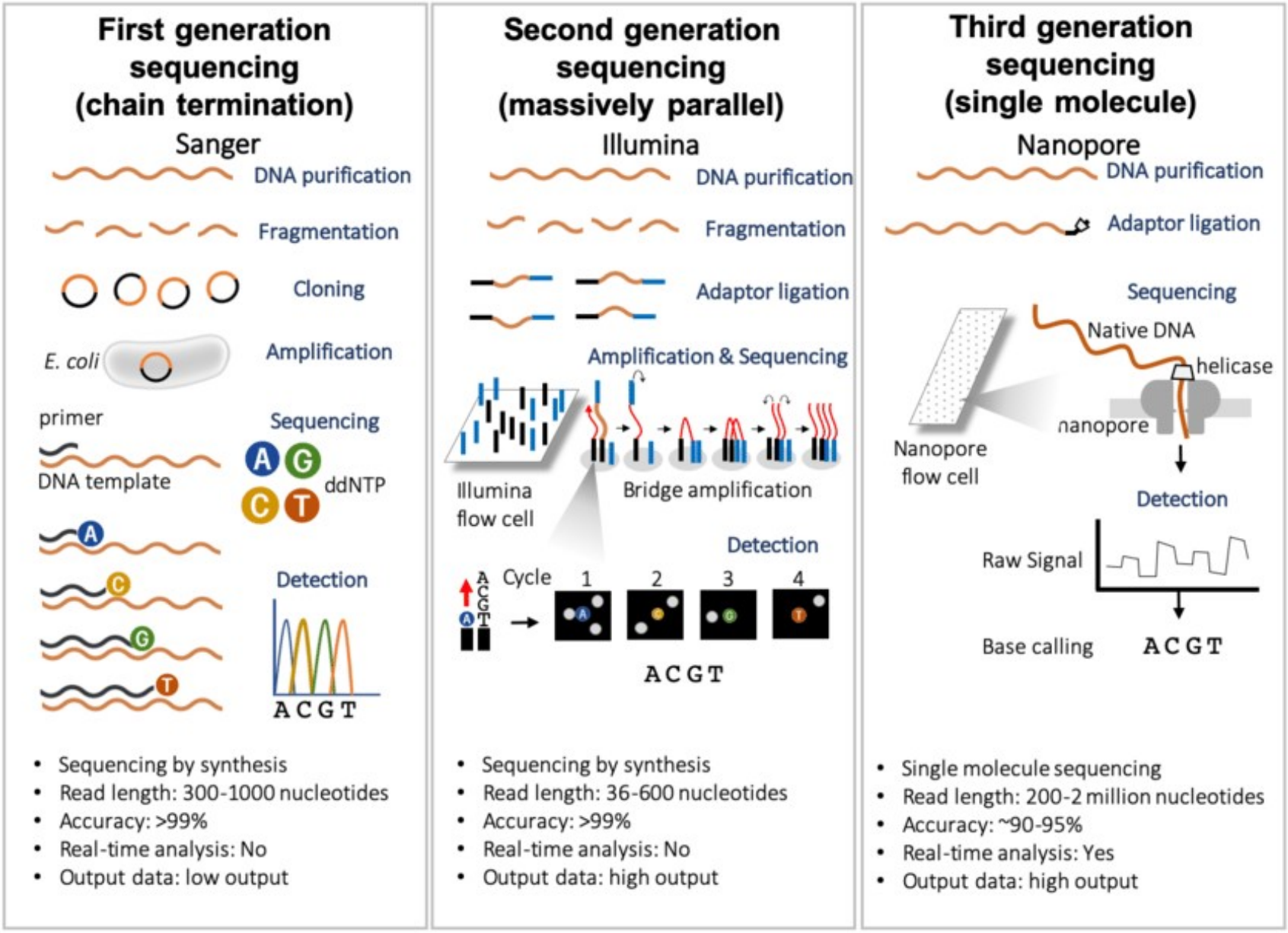
2005 – First "next-generation" sequencing instrument

2011-The "third generation" sequencer

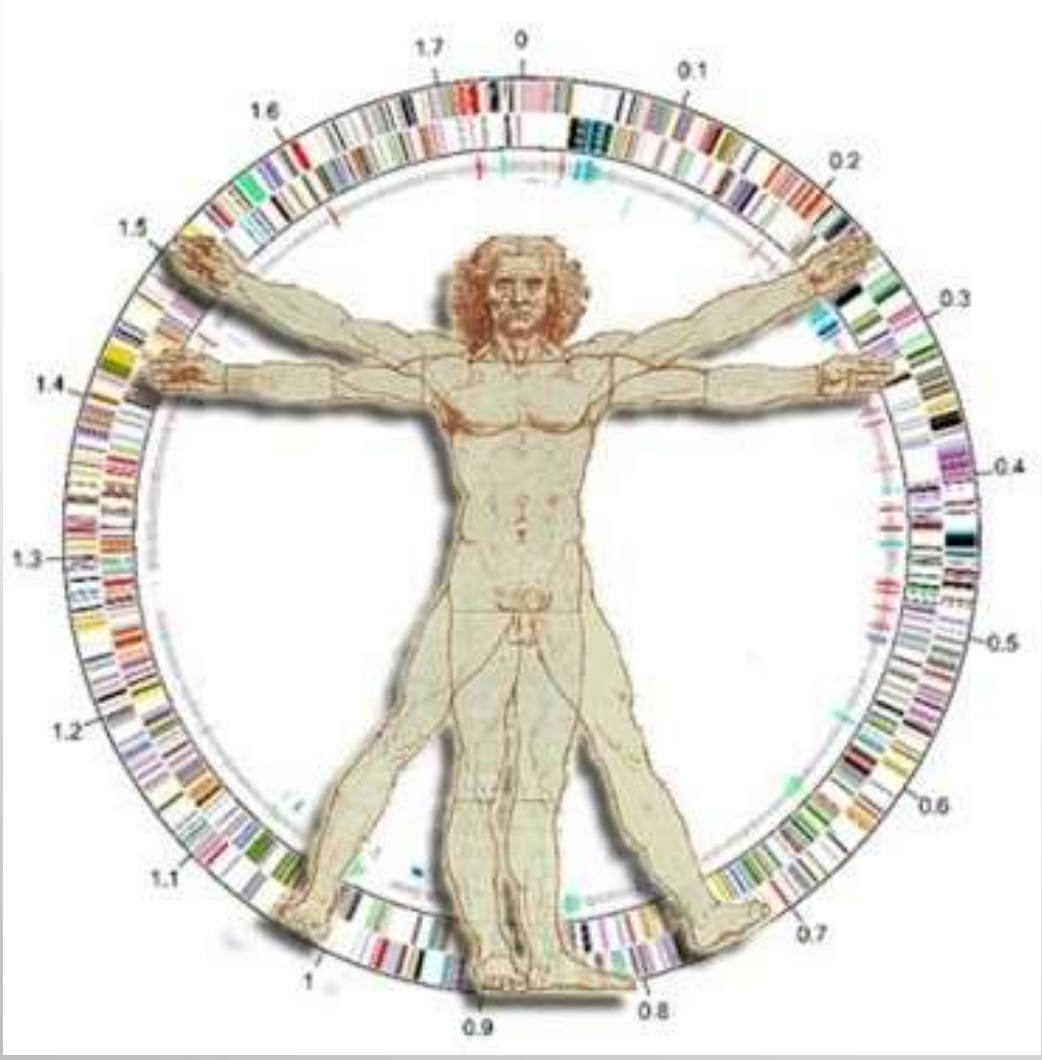
2013– >10,000 genome sequences in NCBI database






First generation sequencing (chain termination)
Sanger
DNA purification
Fragmentation
Cloning
E. coli
Amplification
primer
DNA template
Sequencing
ddNTP
Detection
A C G T
Sequencing by synthesis
Read length: 300-1000 nucleotides
Accuracy: >99%
Real-time analysis: No
Output data: low output
Second generation sequencing (massively parallel)
Illumina
DNA purification
Fragmentation
Adaptor ligation
Amplification & Sequencing
Illumina flow cell
Bridge amplification
Detection
Cycle 1 2 3 4
A C G T
Sequencing by synthesis
Read length: 36-600 nucleotides
Accuracy: >99%
Real-time analysis: No
Output data: high output
Third generation sequencing (single molecule)
Nanopore
DNA purification
Adaptor ligation
Sequencing
Native DNA
helicase
nanopore
Nanopore flow cell
Detection
Raw Signal
Base calling
A C G T
Single molecule sequencing
Read length: 200-2 million nucleotides
Accuracy: ~90-95%
Real-time analysis: Yes
Output data: high output

Human genome project (人類基因體計劃)



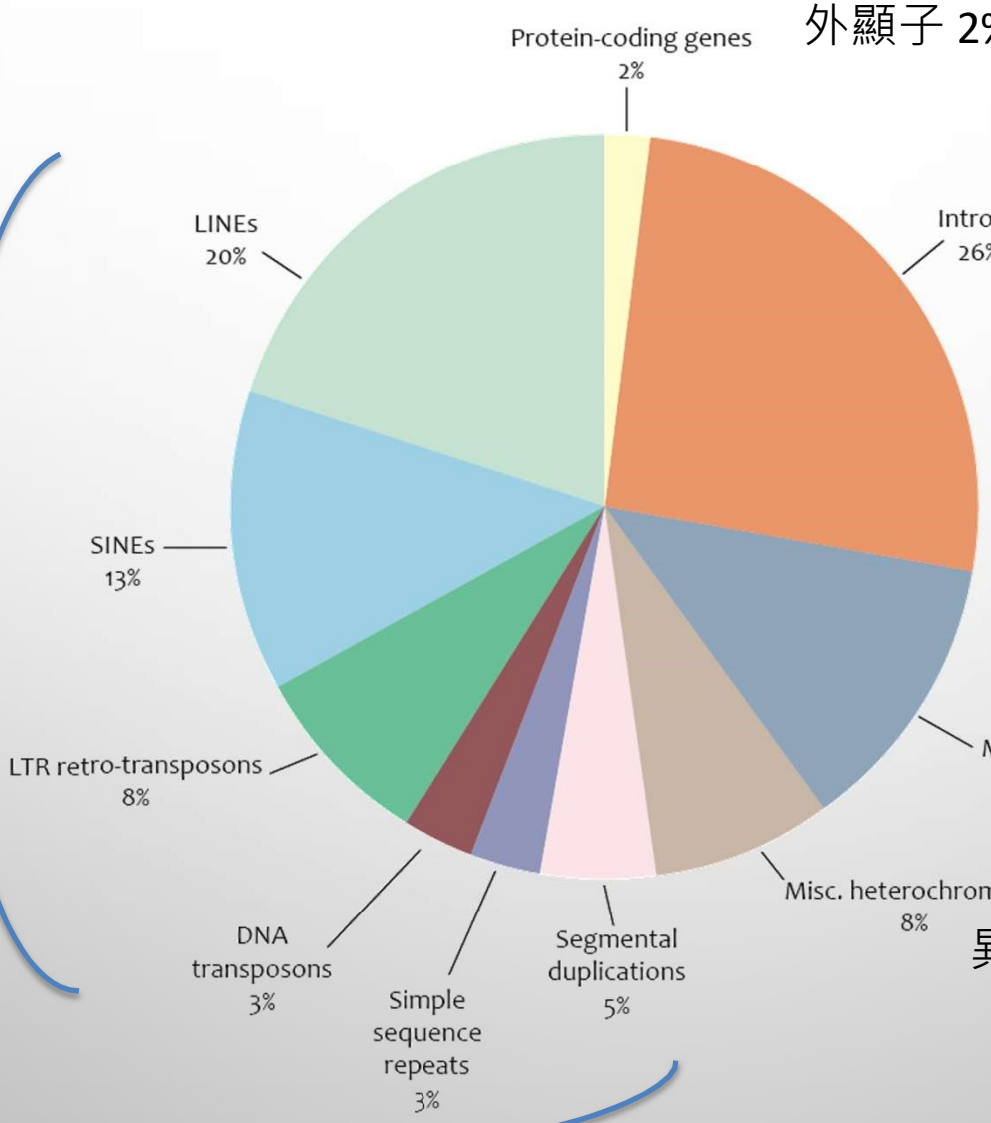
- 目的：將人類基因組序列完全定序並註解所有基因
- 1990計劃啟動
- 2003公佈草圖
- 總經費$3,000,000,000美元
- 共18個國家參與

https://goa.pscnotes.com

- 1984 – 科學家於美國能源部會議提出構想
- 1986 – 與會科學家再次強調該計劃重要性並討論
- 1988 – 與會科學家一致同意該計劃重要性並準備著手進行
- 1990 - 提出初步構想(為期15年, 經費美金$3,000,000,000,採用階層式定序法)
- 1992 - 發布低解析度基因組草圖(genome map)
- 1998 – Celera公司宣佈將以霰彈槍定序法於五年內完成基因組定序，經費$300,000,000，完成後將註冊所有基因
- 1999 – 第一條染色體公布(chromosome 22)
- 2000 - Celera公司宣佈已完成~97%
- 2003 - 人類基因組計劃完成(99%)
- 2022 - 人類基因組計劃完成(100%) Science 376.6588 (2022): 44-53.

https://en.wikipedia.org/wiki/Human_Genome_Project

人類基因體序列分析



外顯子 2%

蛋白質譯碼基因

內含子 26%

跳躍子 44%

非蛋白質譯碼基因

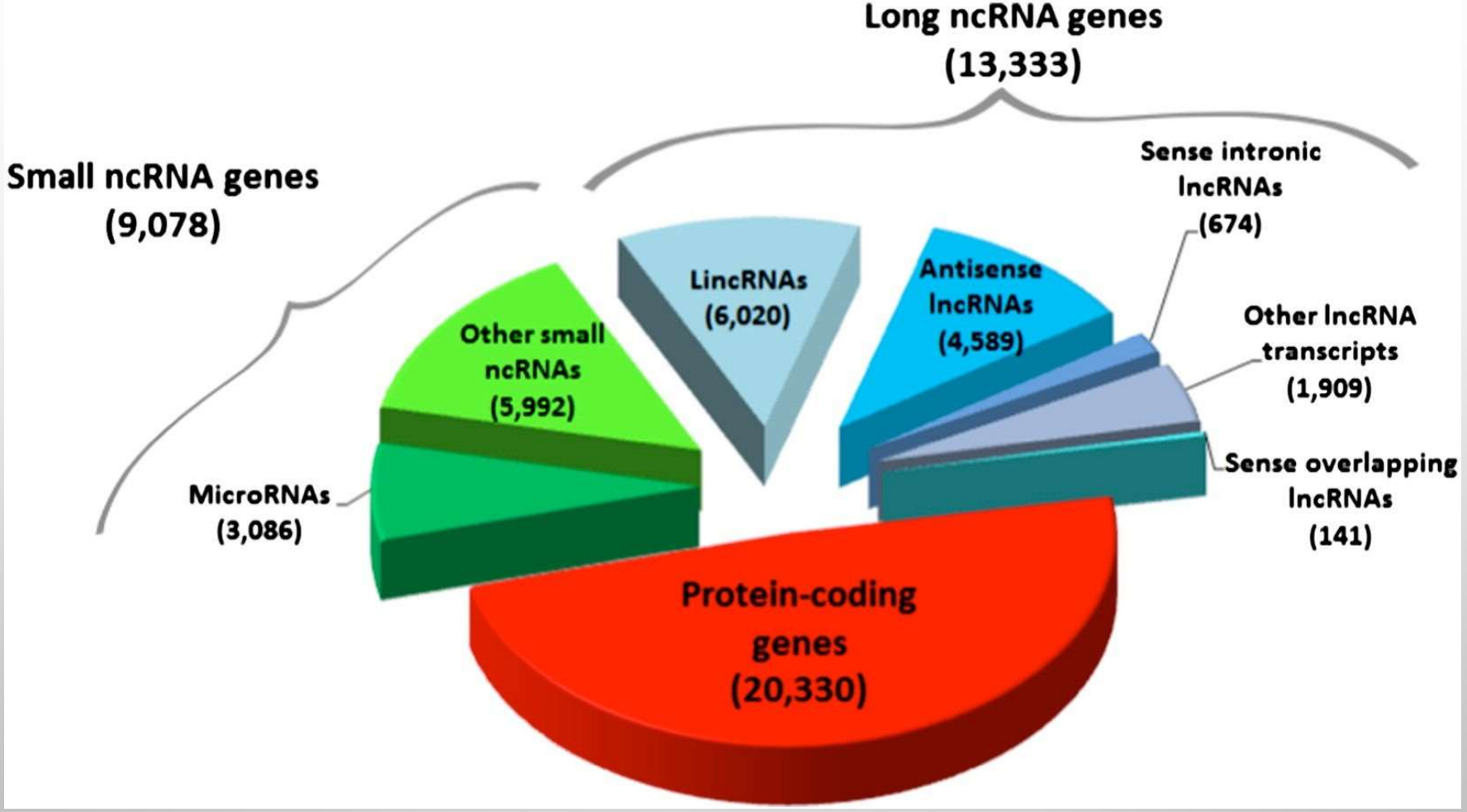
異染色質

重覆序列

https://link.springer.com/chapter/10.1007/978-3-030-73151-9_1

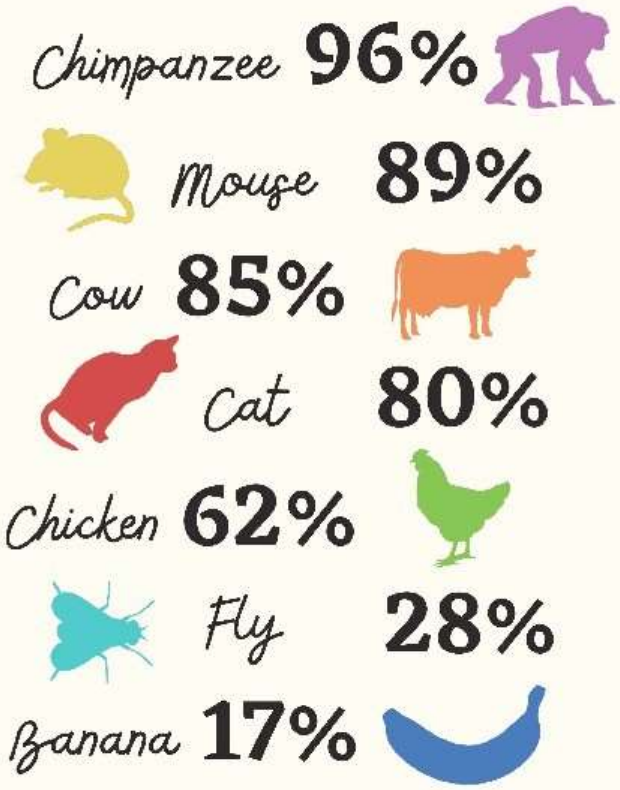
人類基因體基因註解

ncRNAs (ncRNAs genes) versus mRNAs (protein coding genes)



https://www.researchgate.net/publication/351788403

What portion of genes do humans share with...?



- the portion of the human genome that contains a unique ortholog in the comparison species, considering only protein coding genes
- calculated using omadb, a python wrapper for the OMA database
- inspired by Natasha Glover's blog post "The Banana Conjecture"

Posted in r/genetics

reddit

全基因組定序(Whole Genome Sequencing)

SEQUENCING
OUTSMART YOUR GENES®

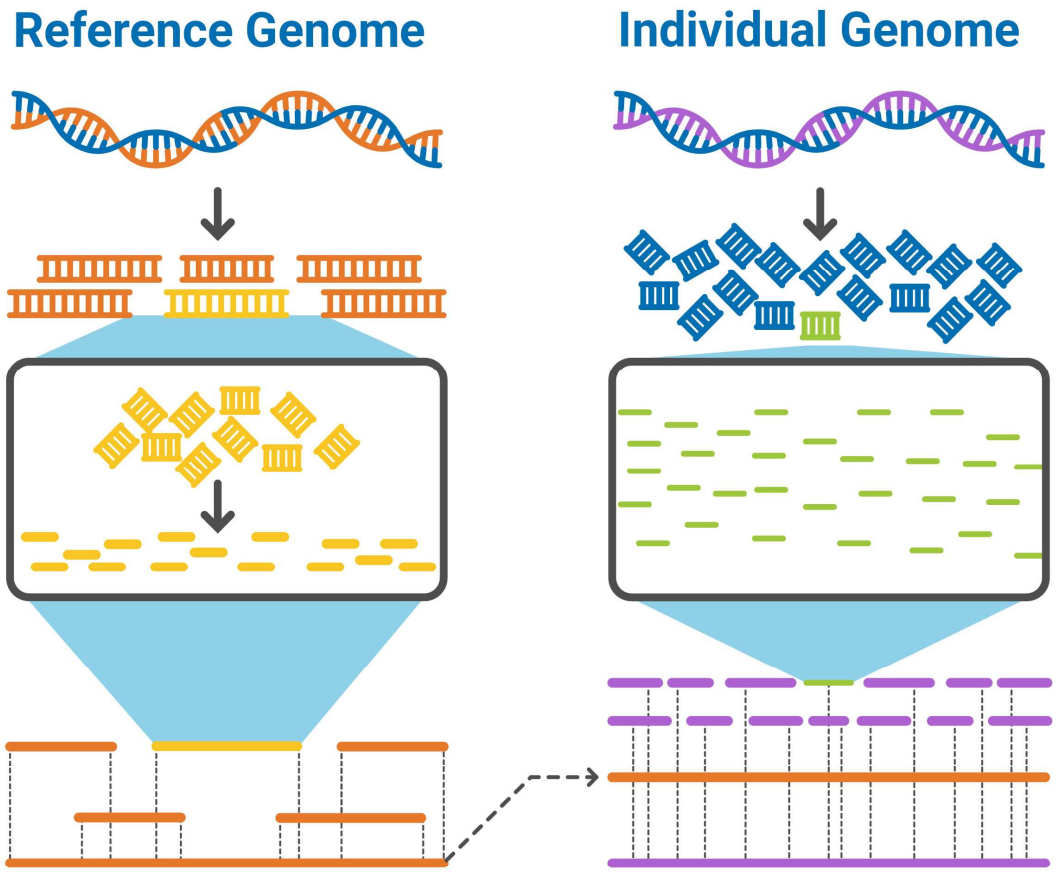
1 Break genome into large fragments and clone

2 Break individual clone into small fragments

3 Generate thousands of sequence reads

4 Assemble sequence reads for each clone

Reference genome



1 Break genome into small fragments

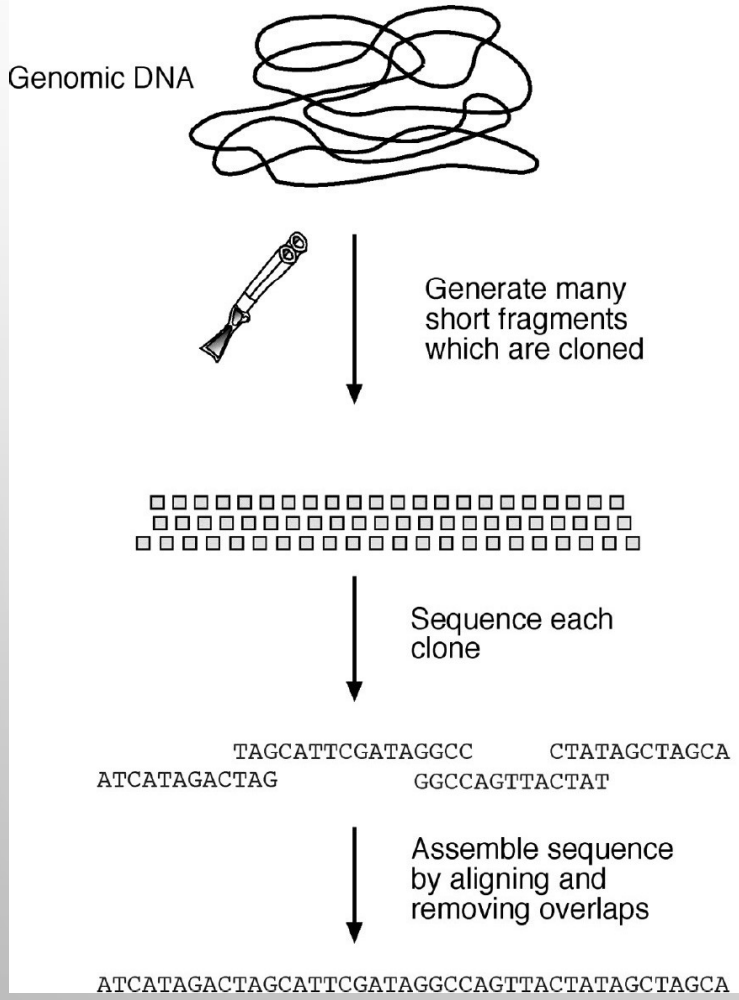
2 Generate millions of sequence reads

3 Align sequence reads into a reference genome

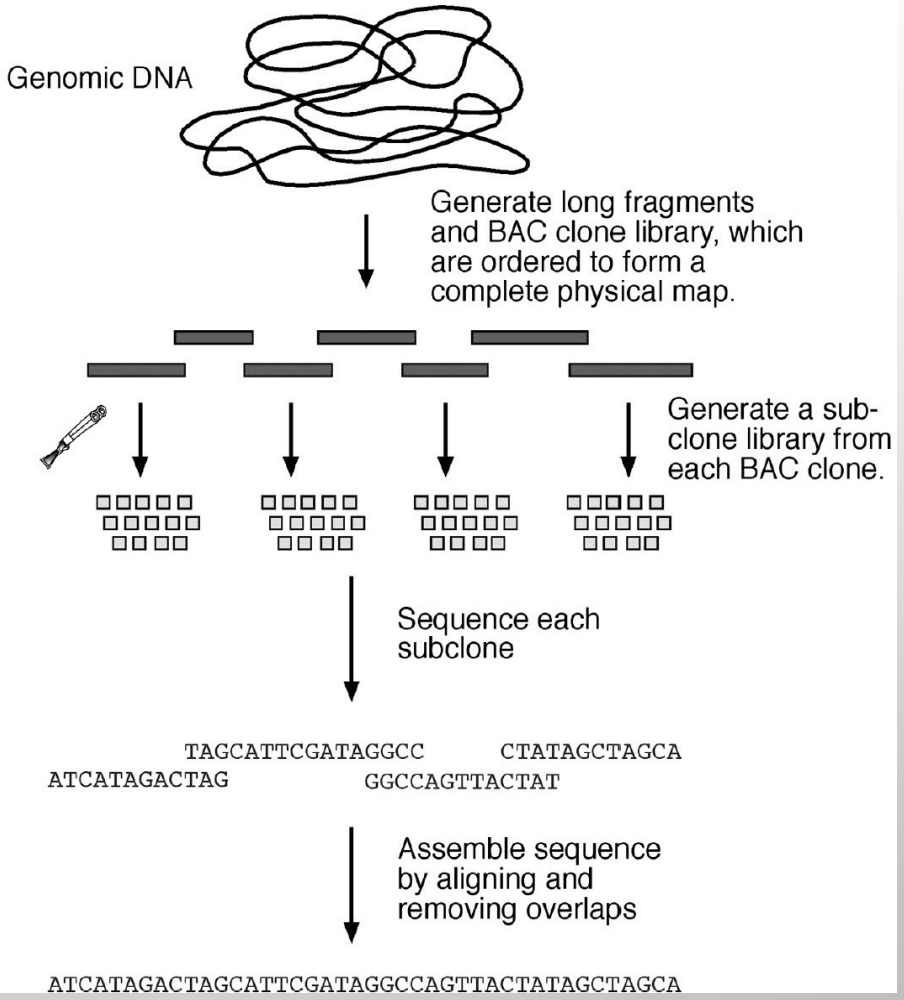
Individual genome

基因組定序策略(Genome sequencing strategies)

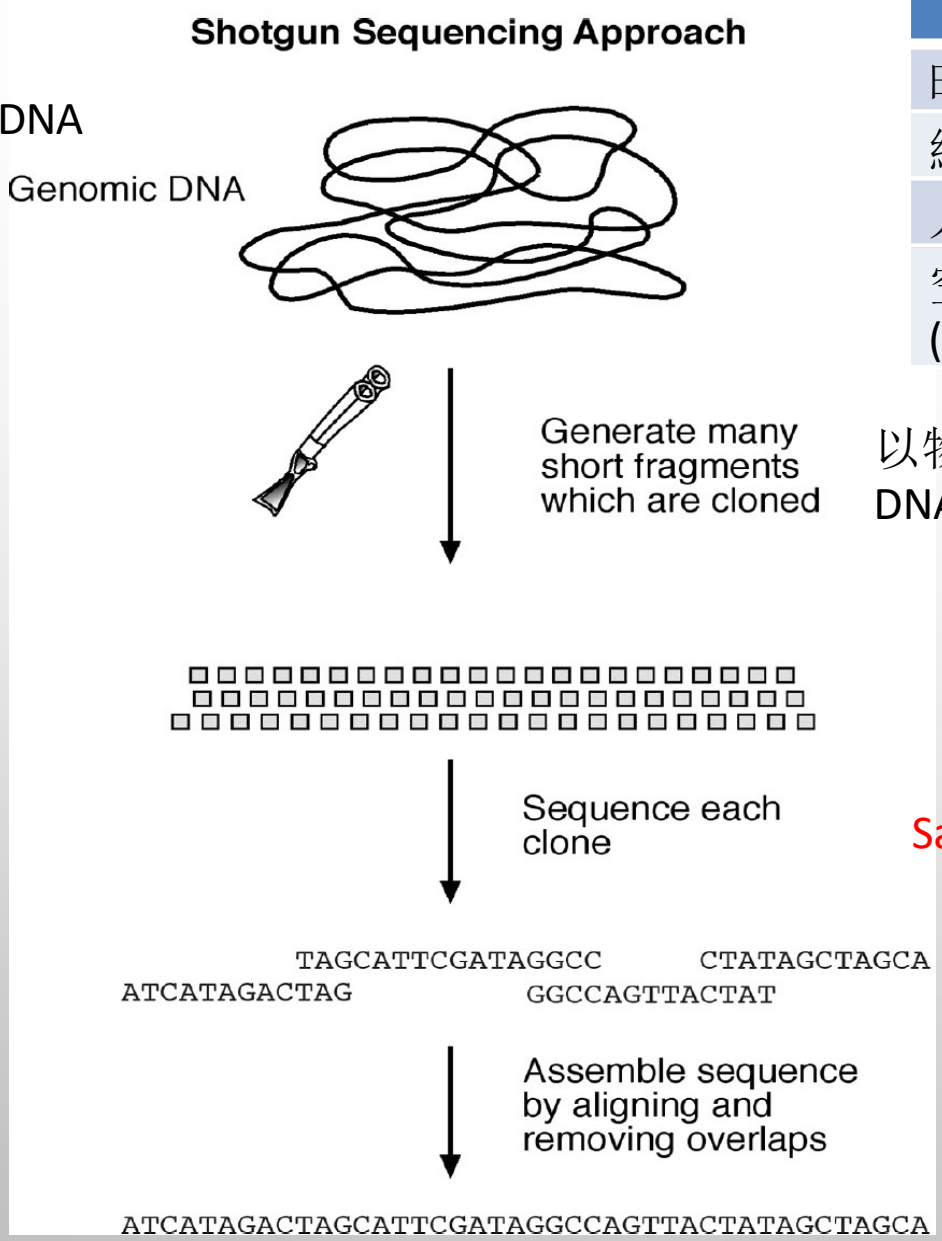
Shotgun sequencing approach (霰彈槍定序法)



Hierarchical sequencing approach (階層式定序法)



Shotgun sequencing approach (霰彈槍定序法)



	霰彈式定序
時間	短
經費	少
人力	少
空缺區域 (gap)	大

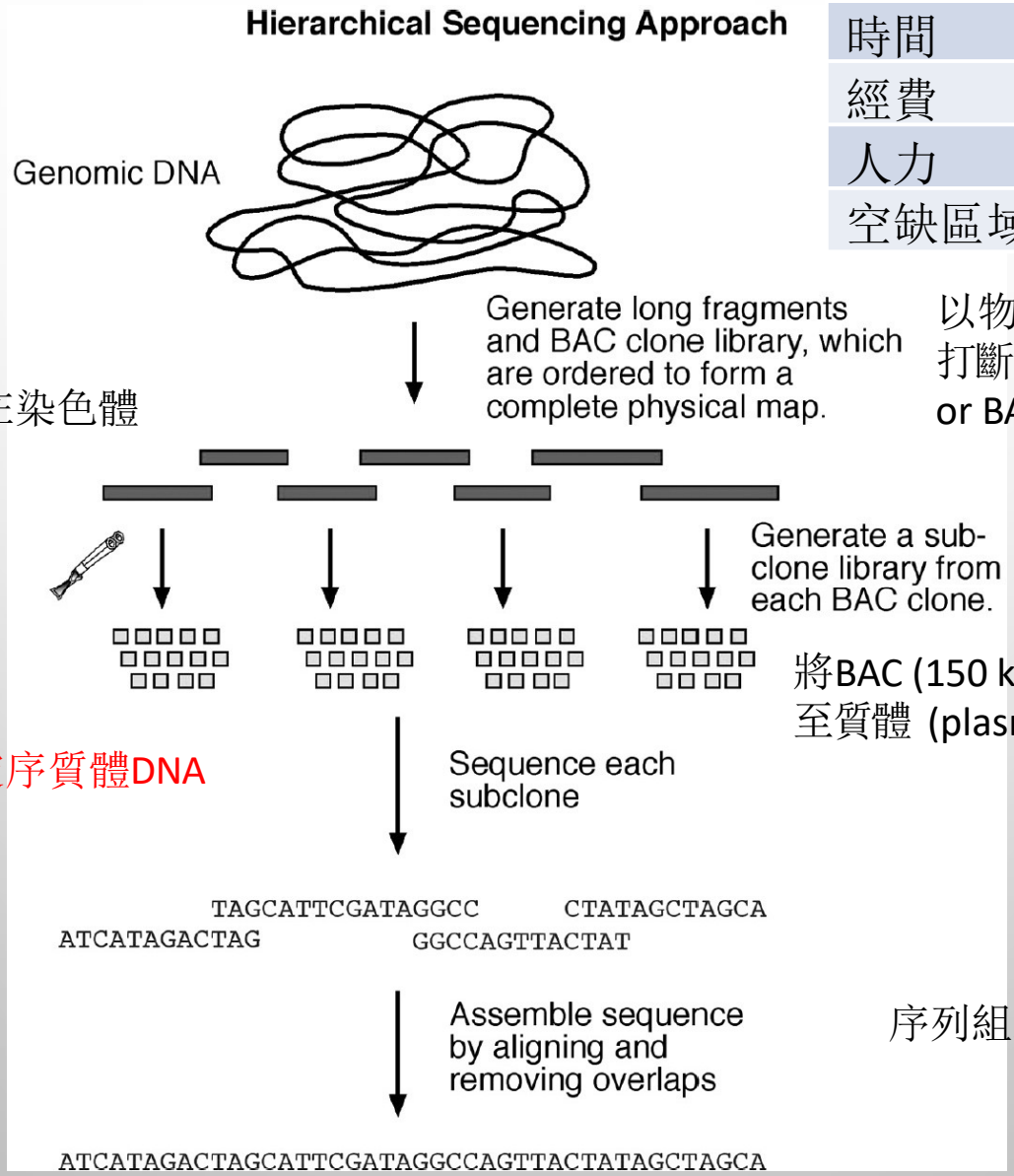
以物理、化學或酵素法將DNA打斷~500 bp

Sanger 或次世代定序

基因組序列組裝

Hierarchical sequencing approach (階層式定序法)

	階層式定序
時間	長
經費	多
人力	多
空缺區域(gap)	小



以物理、化學況酵素法將DNA打斷 ~ 150-300 Kb, 並接在 YAC or BAC

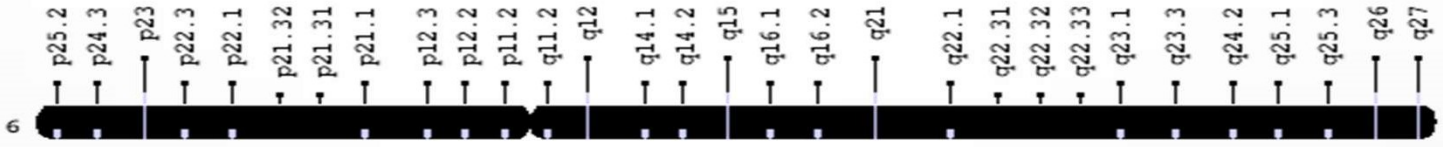
將BAC以FISH定位在染色體

將BAC (150 kb) 片段亞克隆至質體 (plasmid, 2 kb)

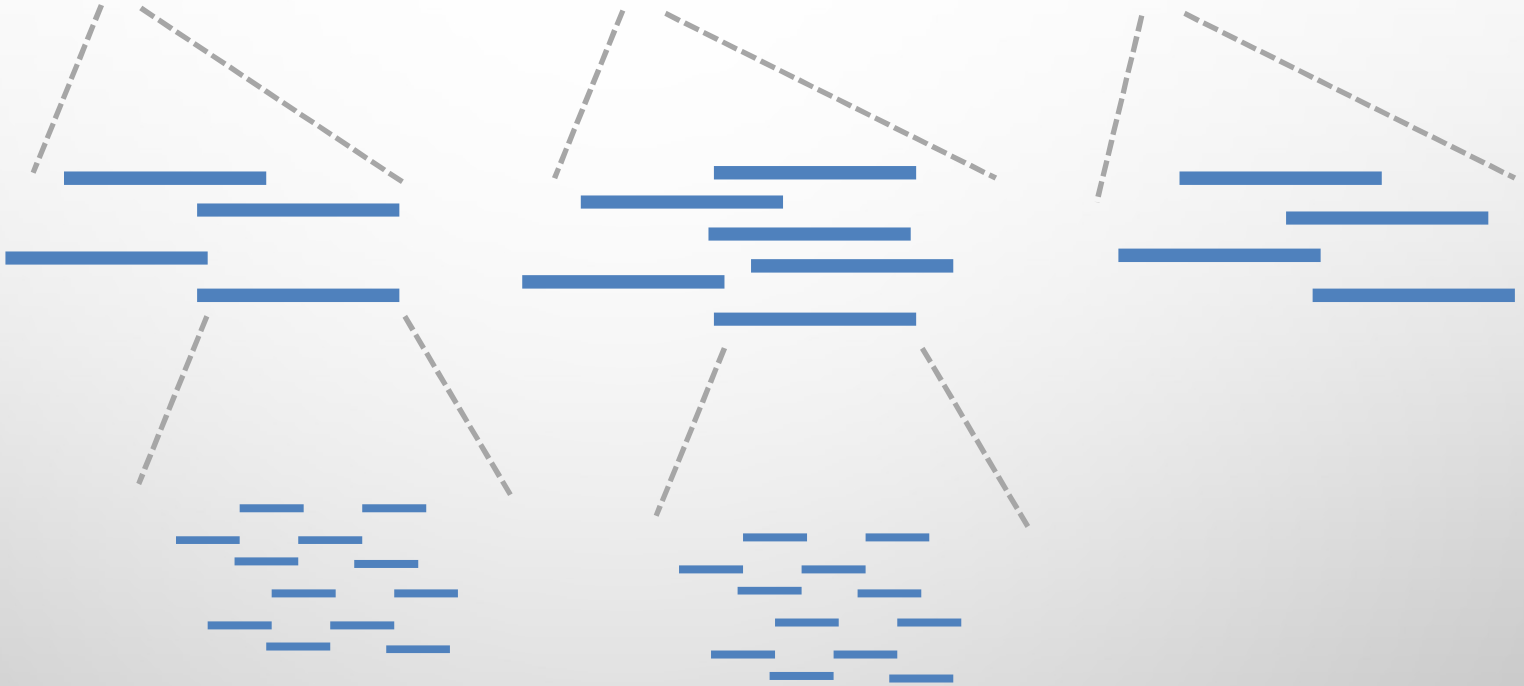
以sanger定序質體DNA

序列組裝

基因組
DNA

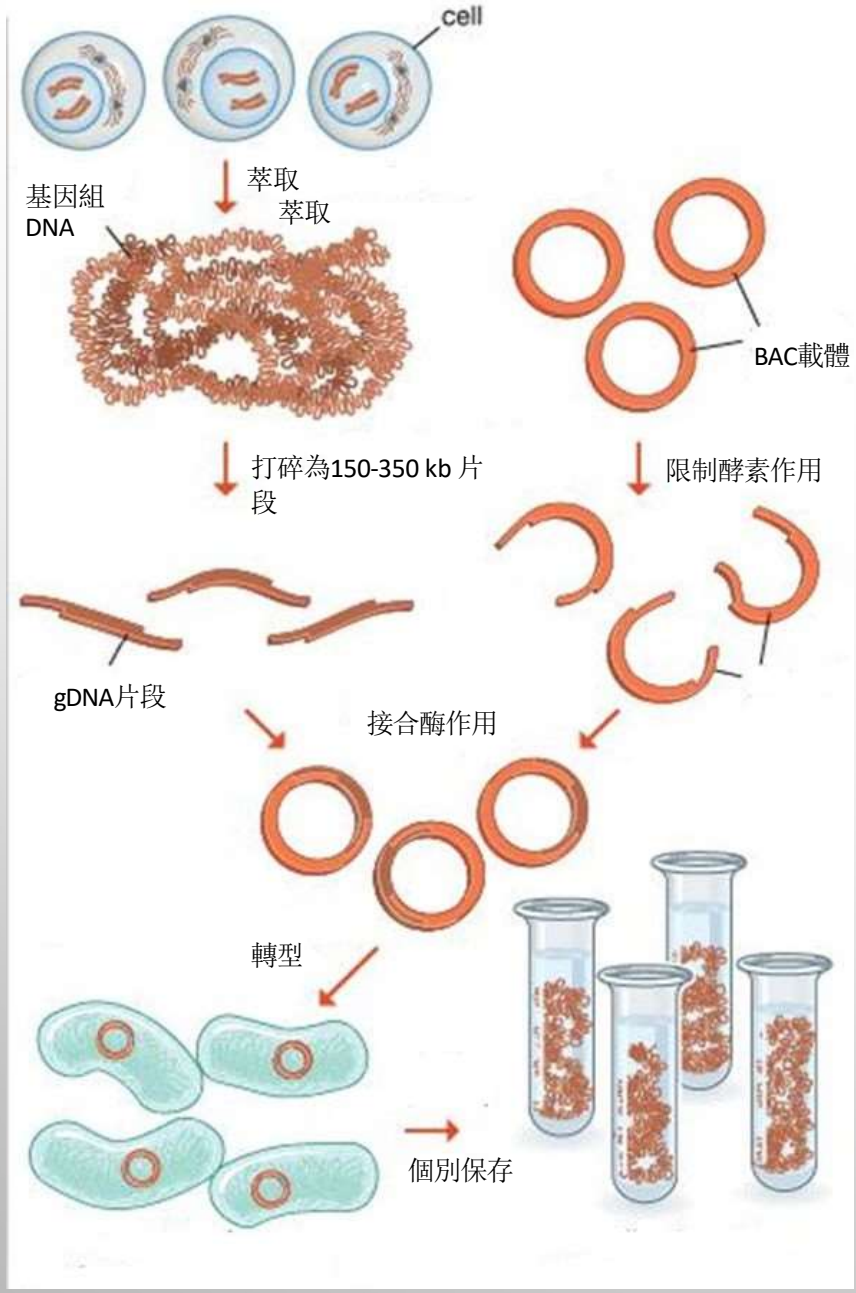


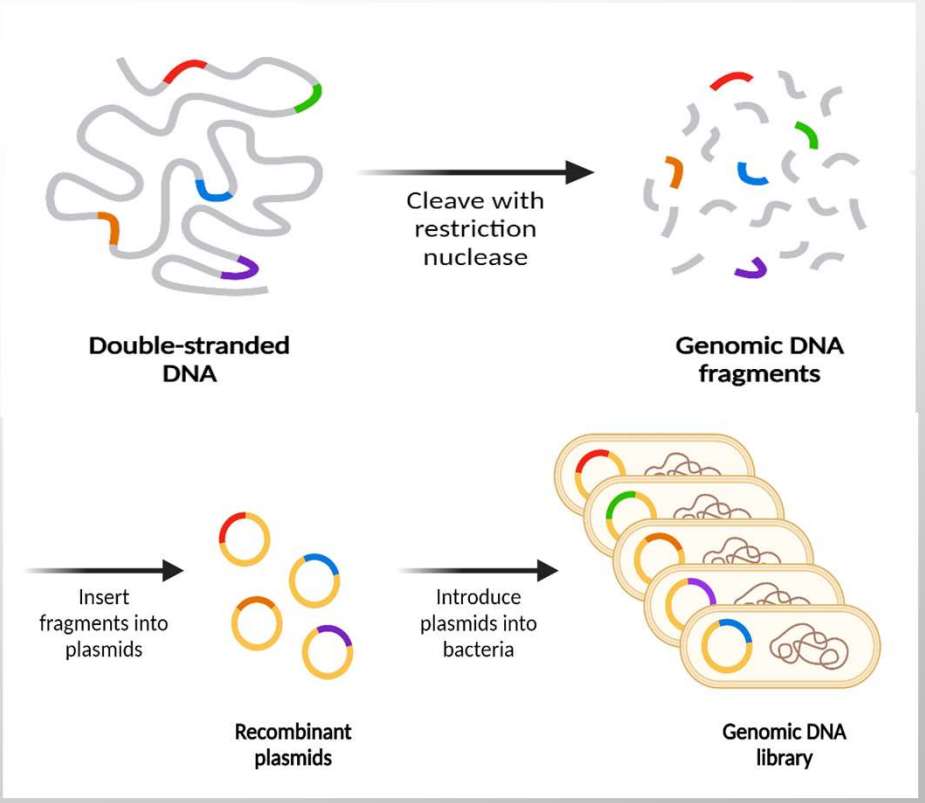
BAC | YAC
基因庫



質體
基因庫

基因組庫建構 (Genomic DNA library construction)



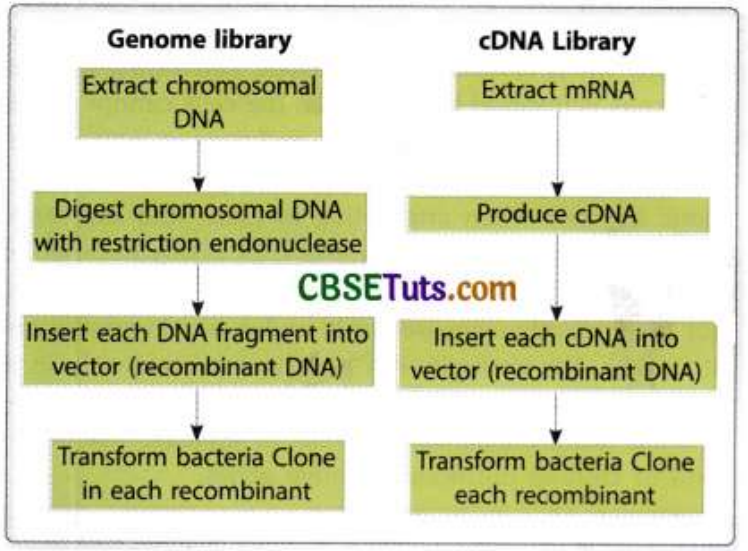


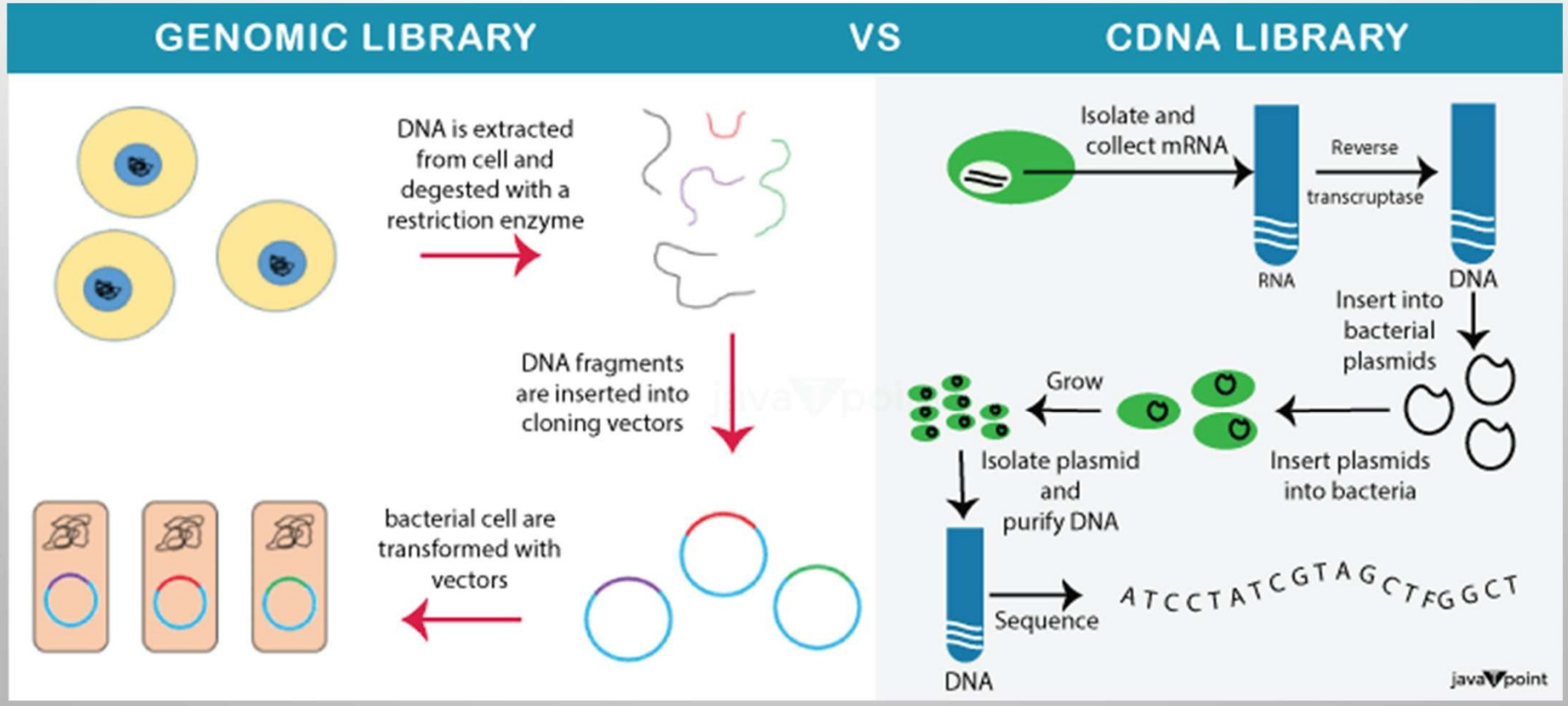
https://www.pinterest.com

Genomic DNA library

vs

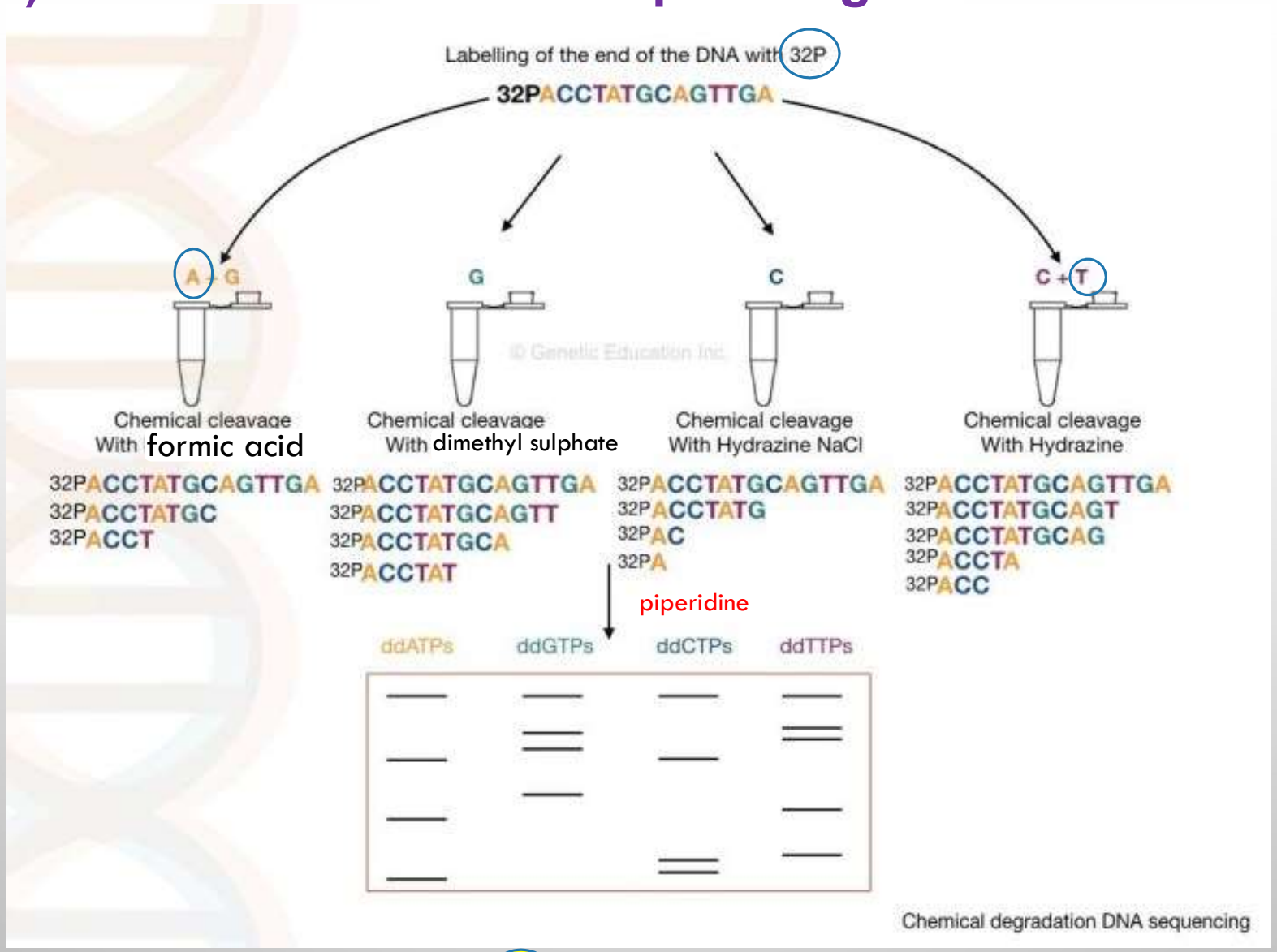
cDNA library





The first-generation DNA sequencing

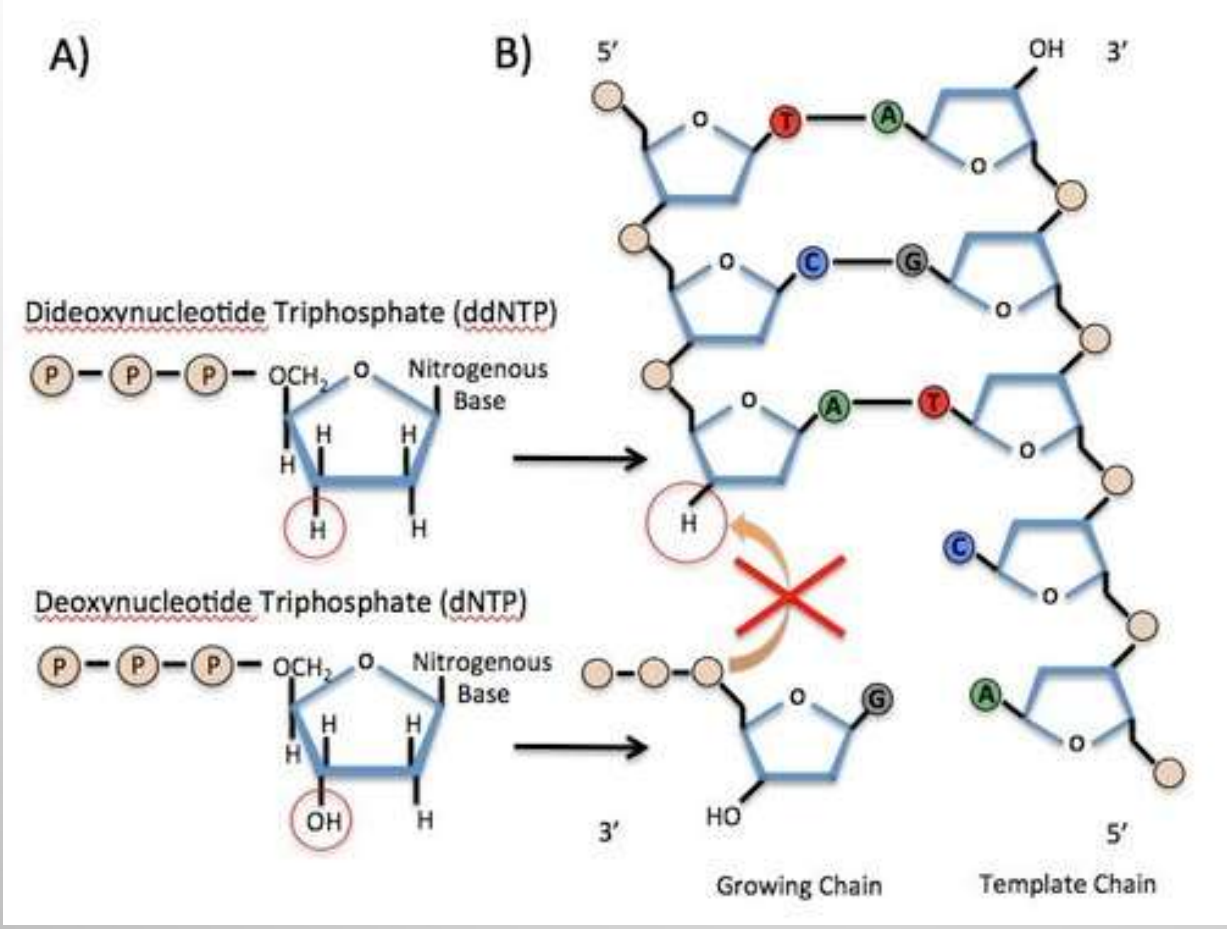
(1) Maxam Gilbert DNA sequencing



https://www.youtube.com/watch?v=_B5Dj8PL4E0

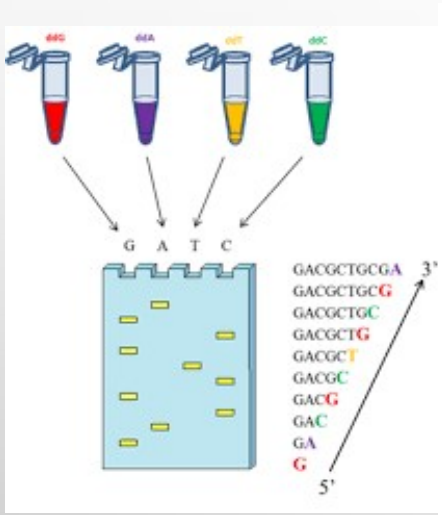
(2) SANGER SEQUENCING

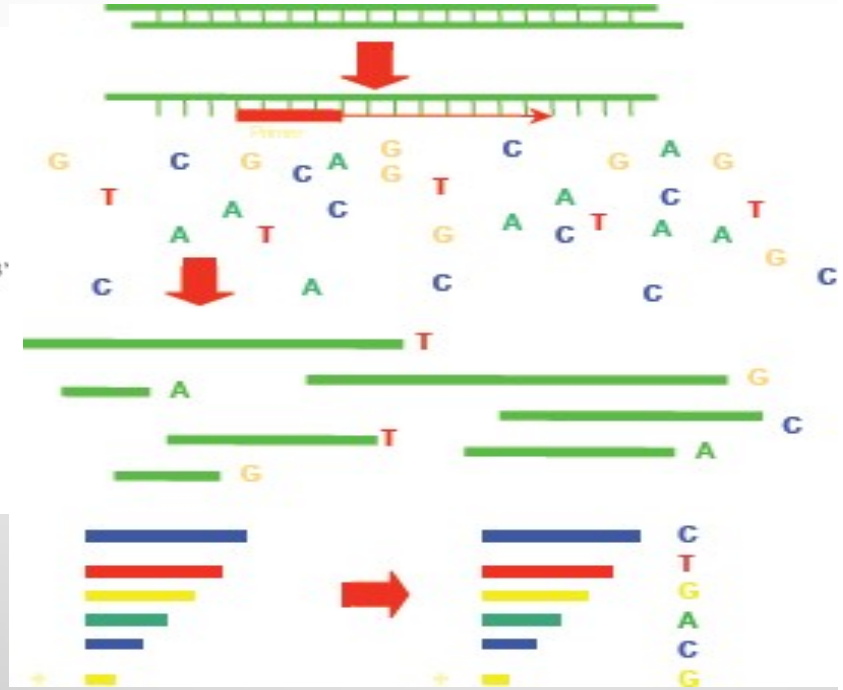
- Uses DNA polymerase
- All four nucleotides, plus one dideoxynucleotide (ddNTP)
- Random termination at specific bases
- Separate by gel electrophoresis

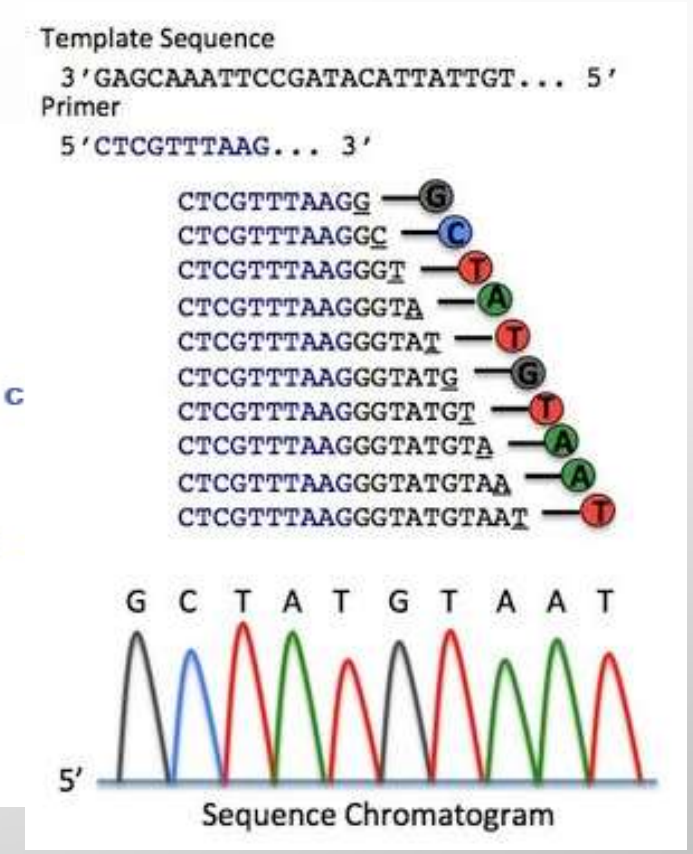


http://cbt.fms.tw/media/8940

1986: 4 Reactions to 1 Lane
fluorescently labelled ddNTPs



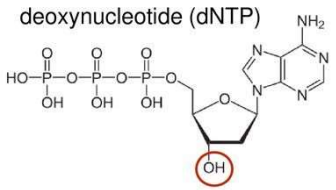


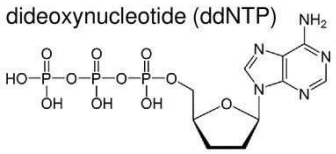


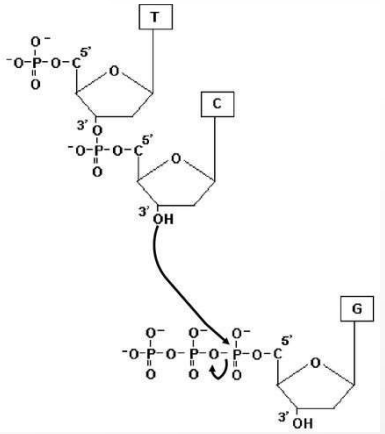
Sequencing Reaction Products

Progression of Sequencing Reaction

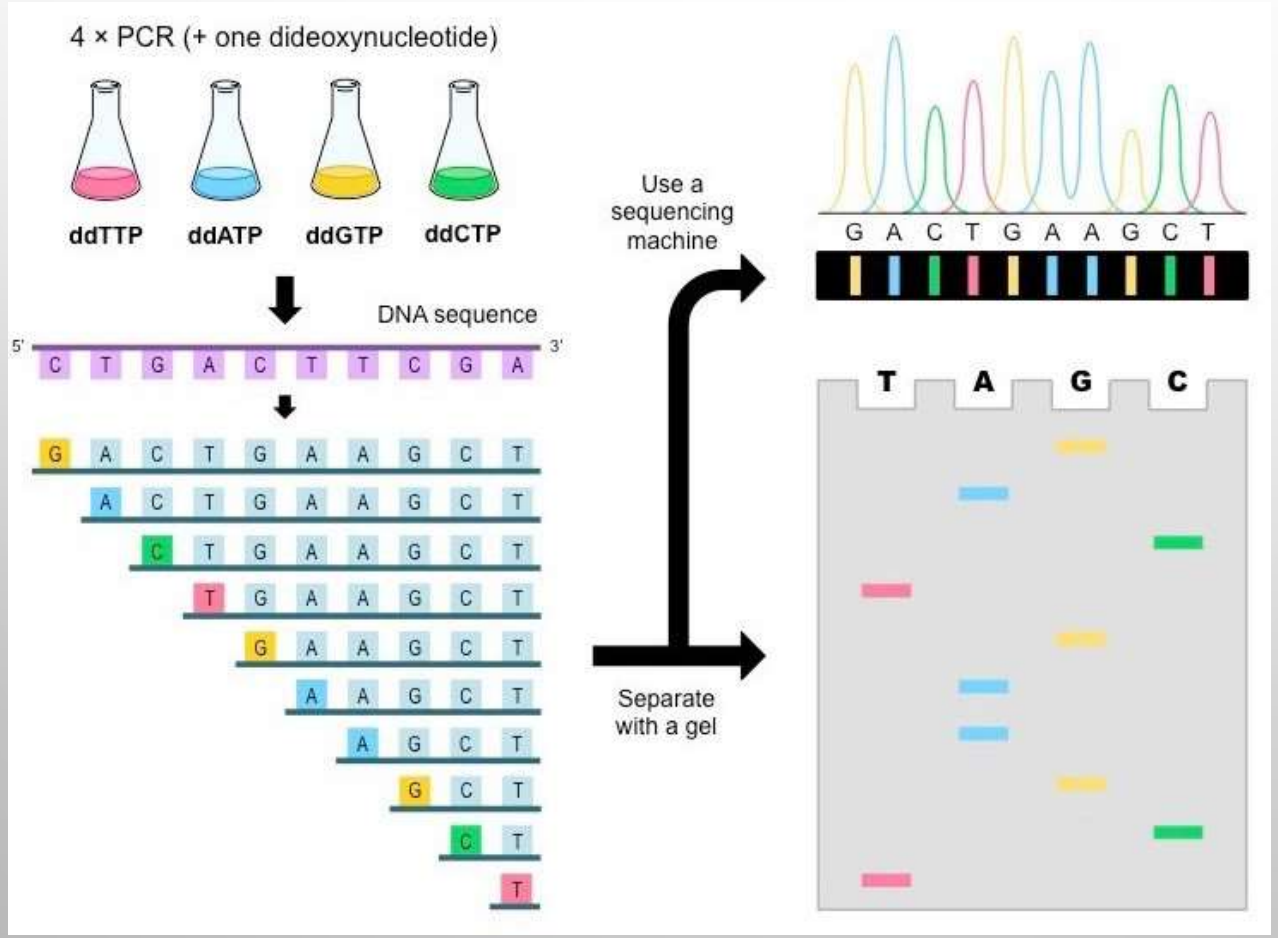
- 藉由螢光標定ddNTP(雙脫氧核苷酸,五碳糖缺乏3端OH基)使PCR反應停止
- 定序長度 500-1200 bp







SANGER SEQ



定序序列

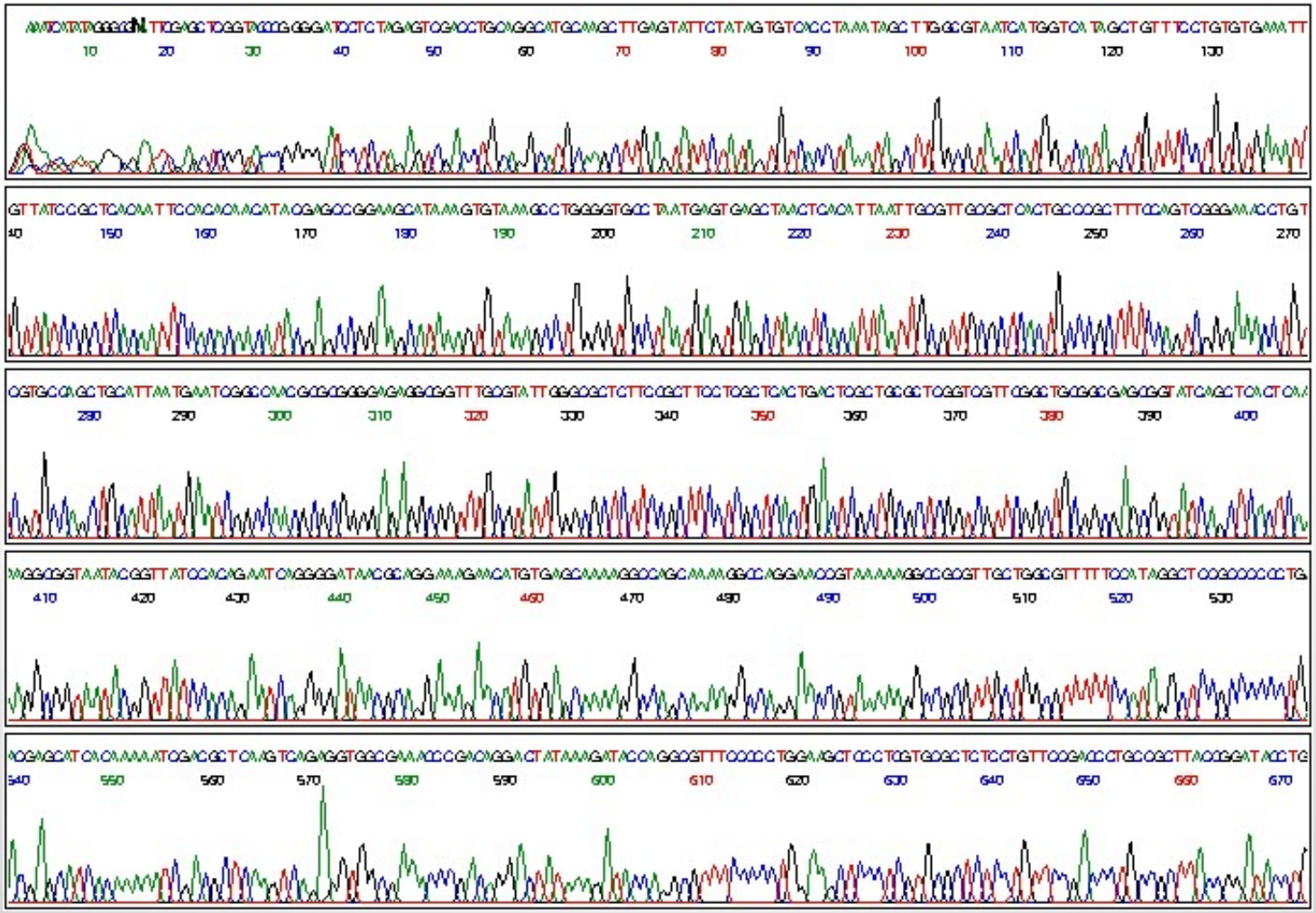
5’ 3’

5’

引子

PCR

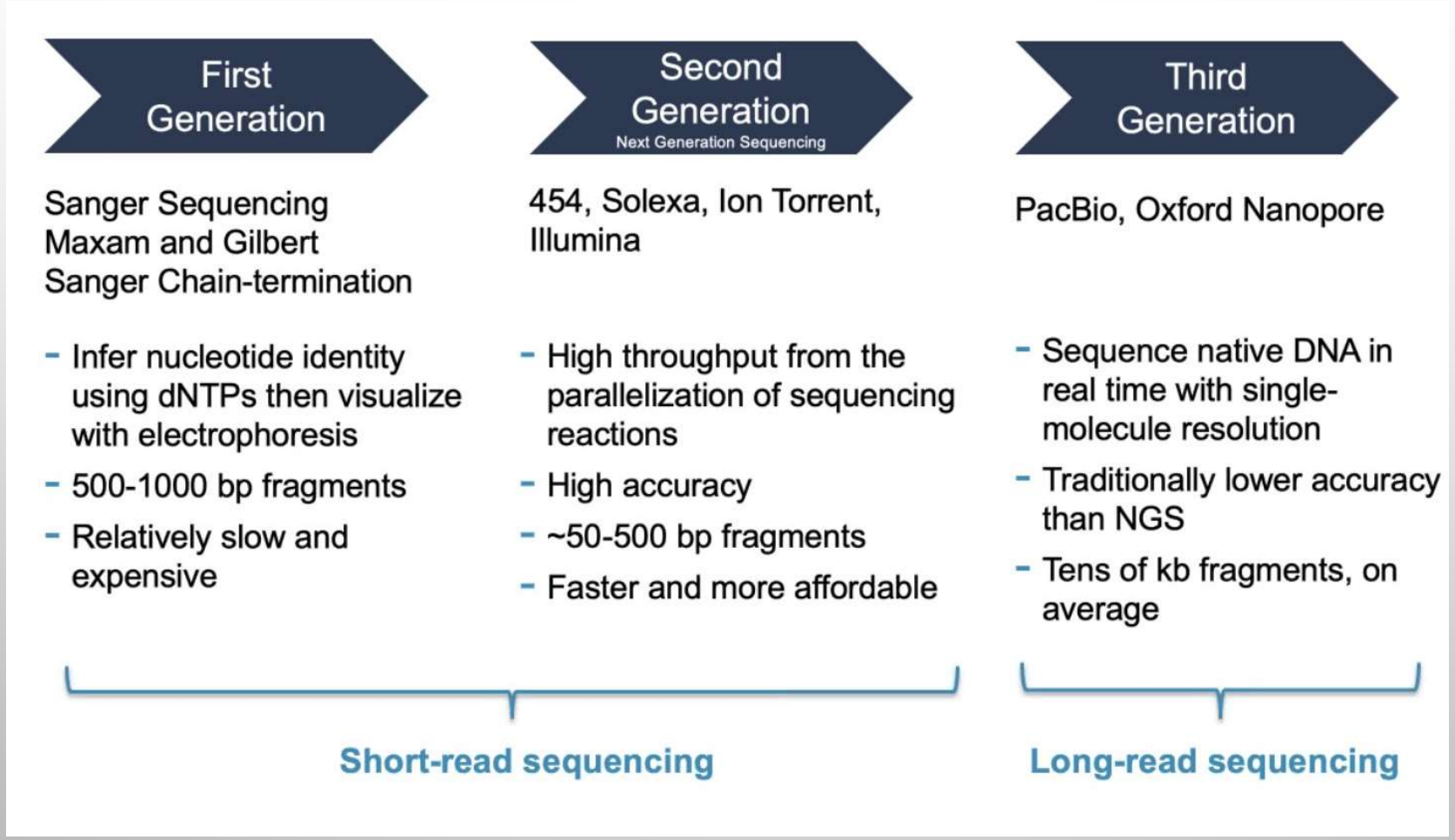
跑膠，螢光分析



https://www.youtube.com/watch?v=FvHRio1yyhQ

https://www.youtube.com/watch?v=iqAmkNSu3oI

Next-generation sequencing (NGS)

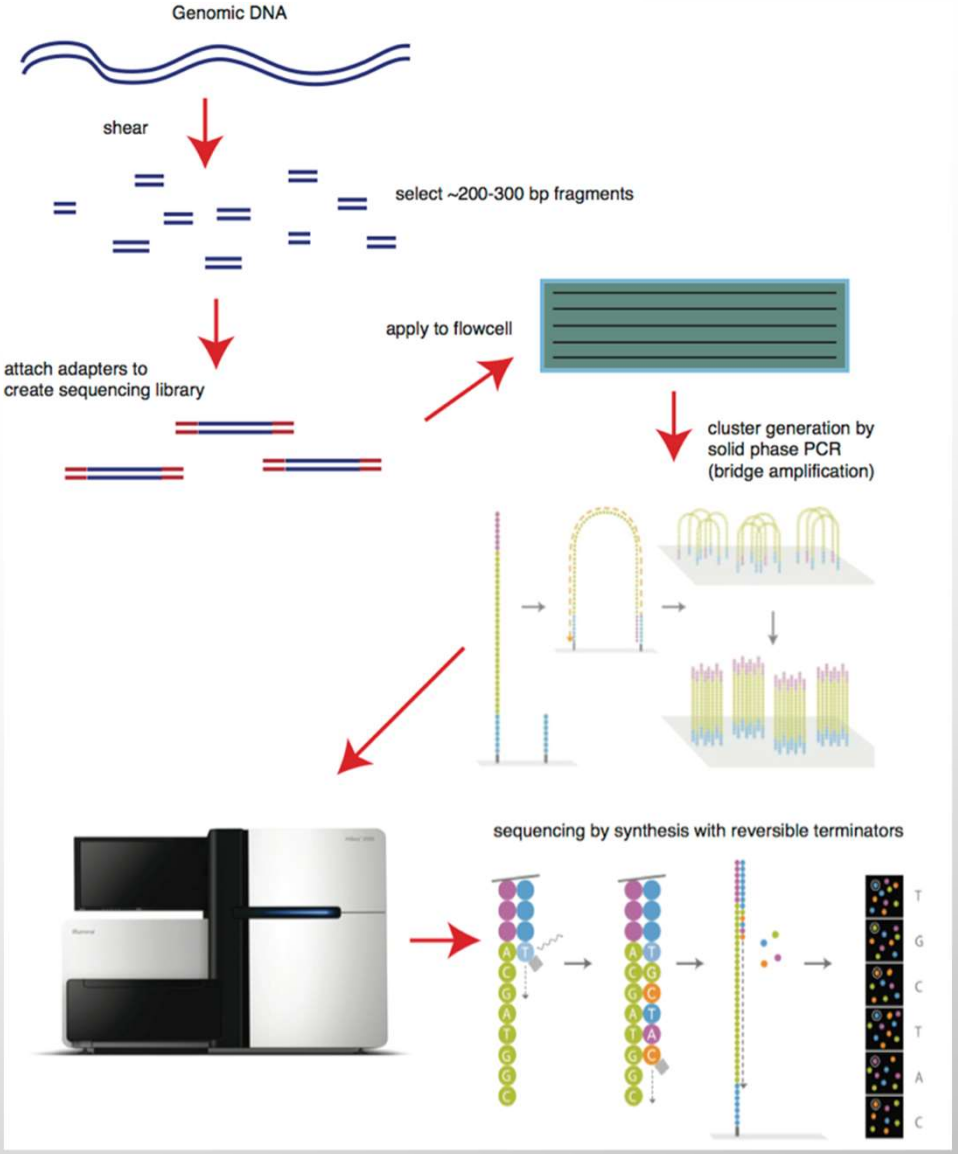


核心方法著重於直接定序單一分子

避免將 DNA 震碎成小片段；若針對 DNA 的表觀遺傳修飾，也能省略建立定序基因庫（library）的階段，故能夠避免 PCR error。

Steps to illumina sequencing

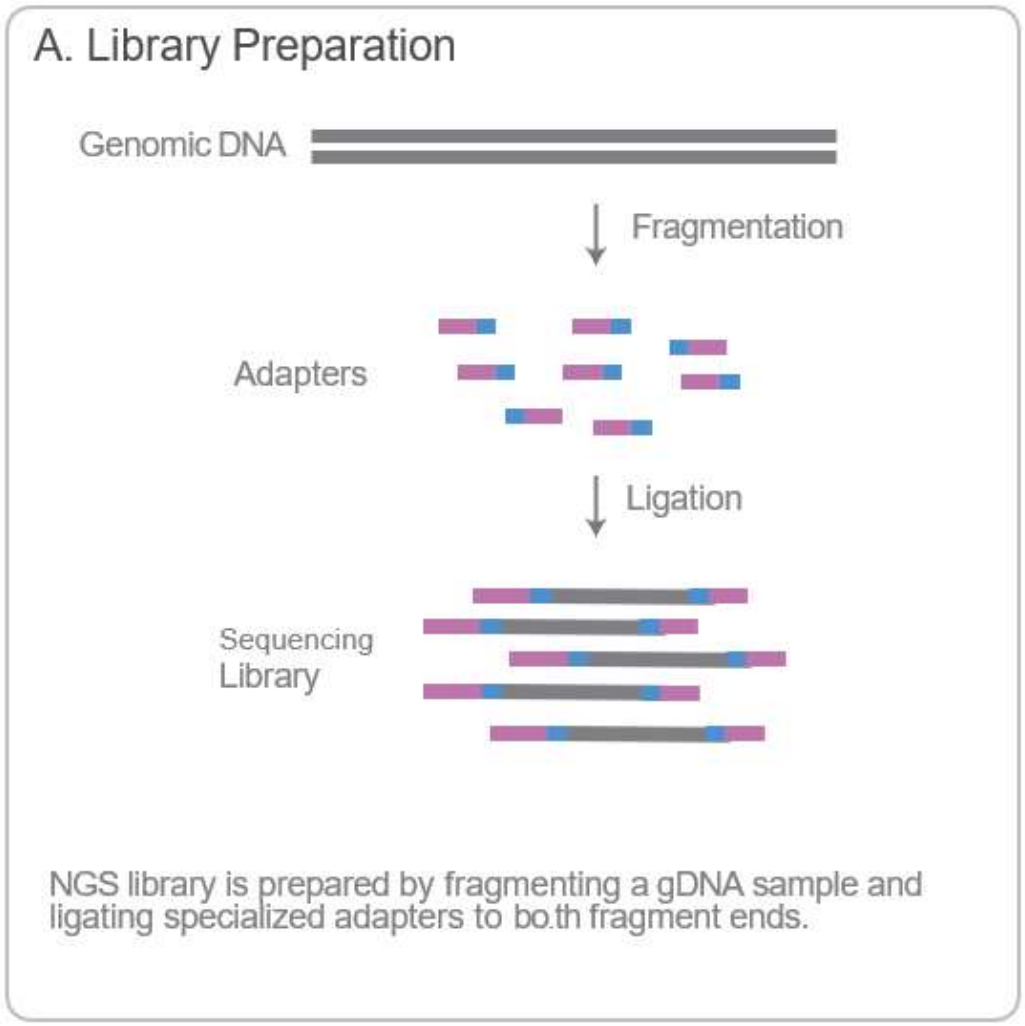
- Library construction
  - - Fragment, attach adapter DNA
- Cluster generation
  - - Add to flow cell
  - - Bridge amplification
- Sequencing
  - - Single base at a time, imaging
- Data analysis
  - - Images transformed into basecalls and 'reads'



https://www.youtube.com/watch?v=womKfikWIxM

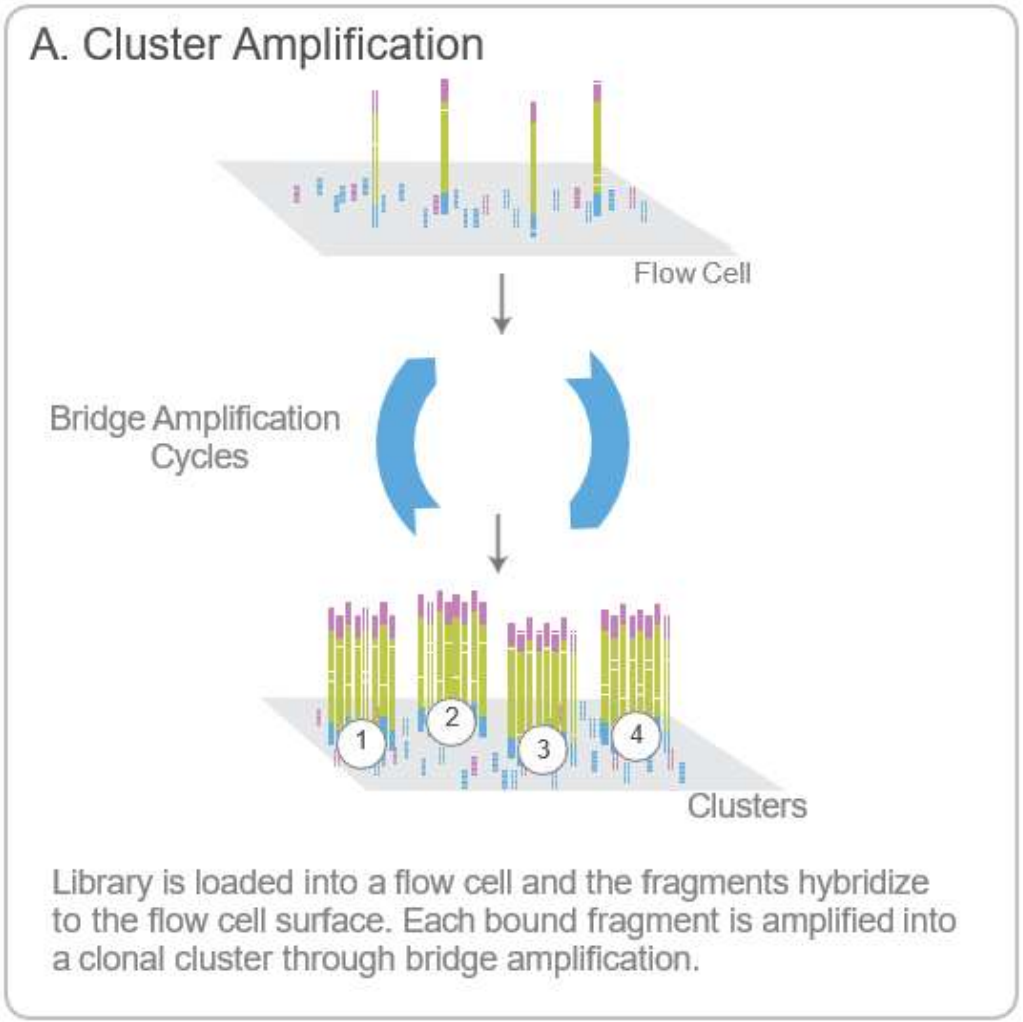
https://www.youtube.com/watch?v=fCd6B5HRaZ8

1. Library preparation—the sequencing library is prepared by random fragmentation of the DNA or cdna sample, followed by 5' and 3' adapter ligation. Alternatively, "tagmentation" combines the fragmentation and ligation reactions into a single step that greatly increases the efficiency of the library preparation process.[9] adapter-ligated fragments are then PCR amplified and gel purified.



1. 利用超聲波將DNA打斷成200-500 bp的片段大小，然後於片段兩端接上接頭(adapter)

2. Cluster generation—for cluster generation, the library is loaded into a flow cell where fragments are captured on a lawn of surface-bound oligos complementary to the library adapters. Each fragment is then amplified into distinct, clonal clusters through bridge amplification. When cluster generation is complete, the templates are ready for sequencing.

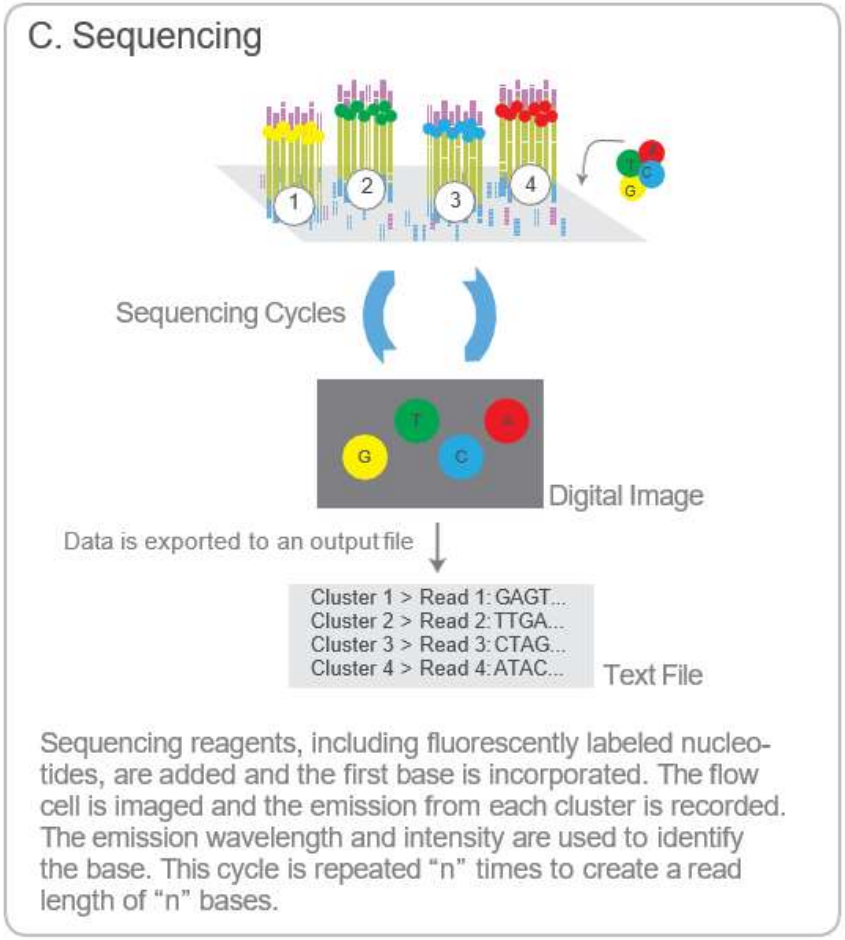


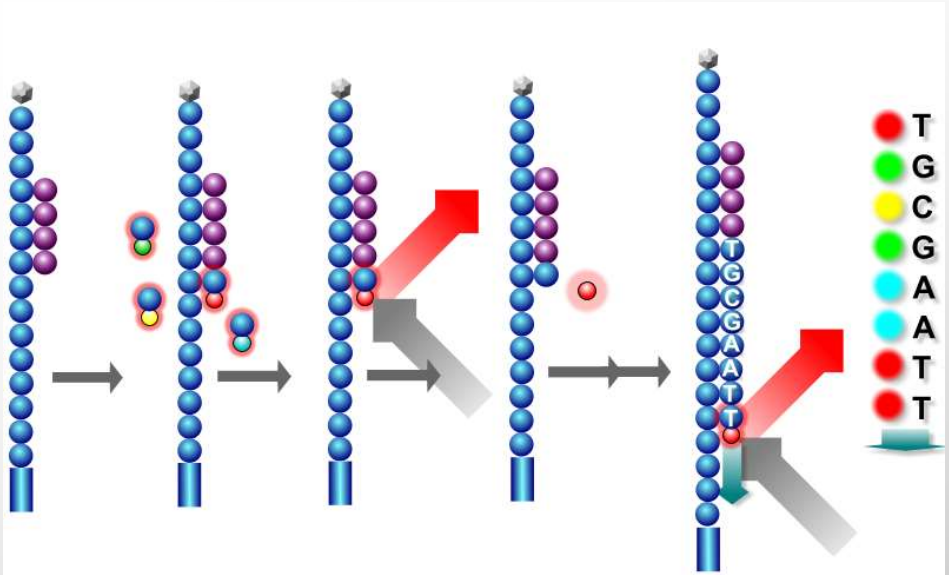
2. 將已接接頭的DNA片段放入到表面帶有互補接頭序列的flowcell，接頭互相配對後讓DNA片段吸附於flowcell上

3. 透過橋式聚合酶連鎖反應進行DNA複製，放大訊號

illumina®

3. Sequencing—illumina SBS technology utilizes a proprietary reversible terminator–based method that detects single bases as they are incorporated into DNA template strands. As all 4 reversible, terminator-bound dntps are present during each sequencing cycle, natural competition minimizes incorporation bias and greatly reduces raw error rates compared to other technologies. The result is highly accurate base-by-base sequencing that virtually eliminates sequence-context-specific errors, even within repetitive sequence regions and homopolymers.



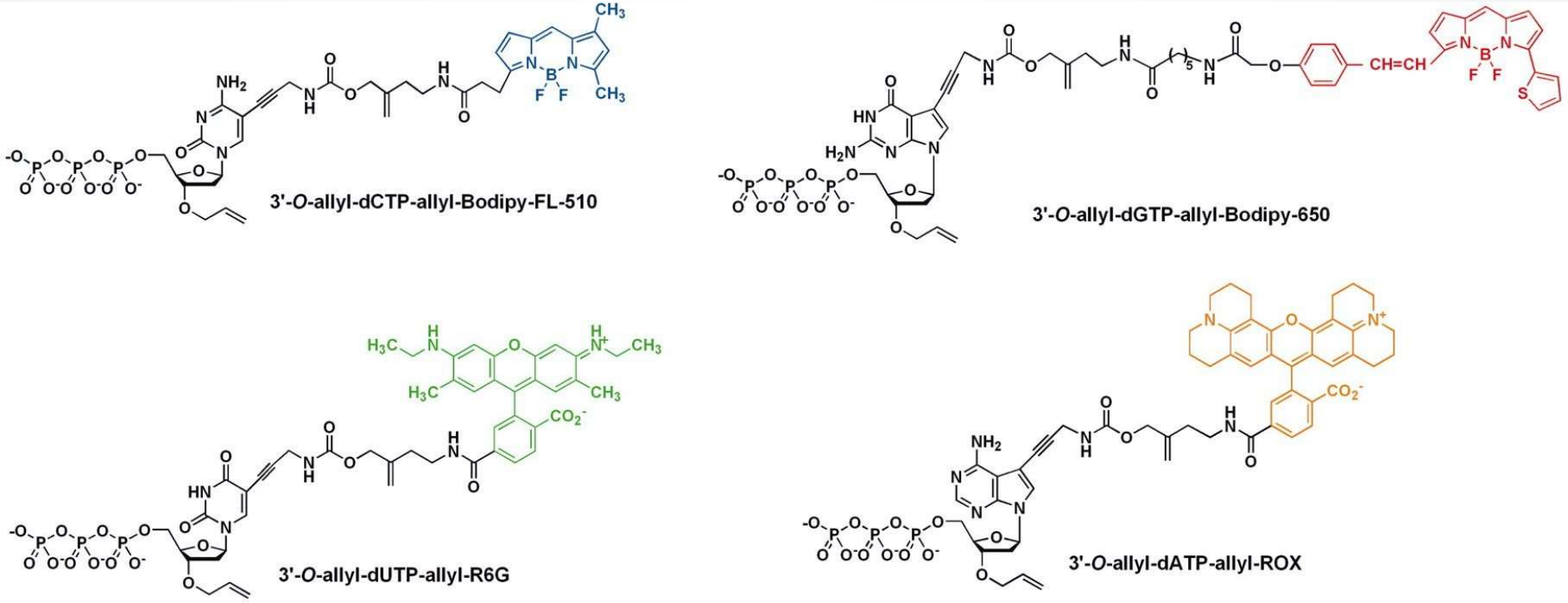


Illumina Sequencing by Synthesis

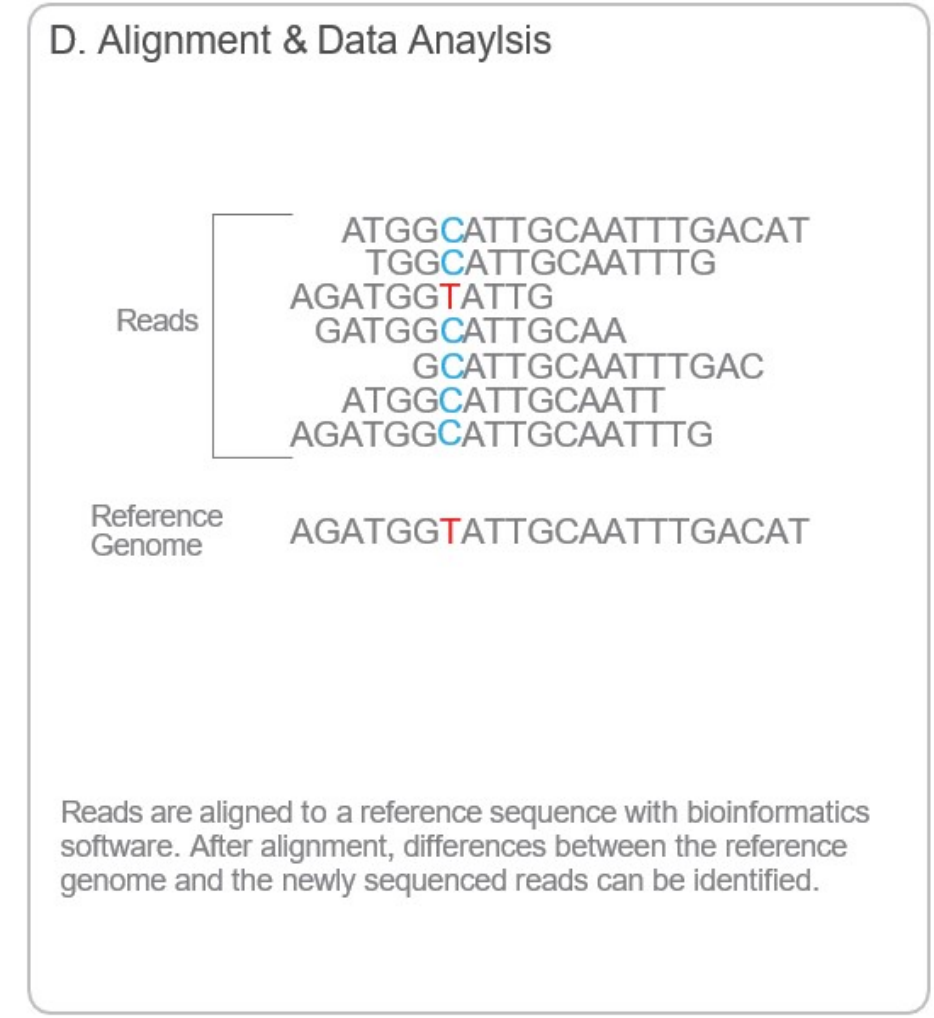
定序方法採用邊合成邊定序（如同Sanger定序法），加入不同已標定特定可移除螢光分子之鹼基(dNTP)及合成反應試劑，反覆讓螢光移除及偵測，達到高速且大量的DNA定序。

SBS 一反應用特殊核苷酸

- SBS所使用的核苷酸為特殊鹼基，上頭帶有發色基，可在激光後發出不同顏色
- 核苷酸五碳糖3端上帶有烯丙基(allyl group), 可使PCR反應無法進行
- 發色基及五碳糖3端的烯丙基可以鈀(pd)反應將其移除，使PCR反應繼續

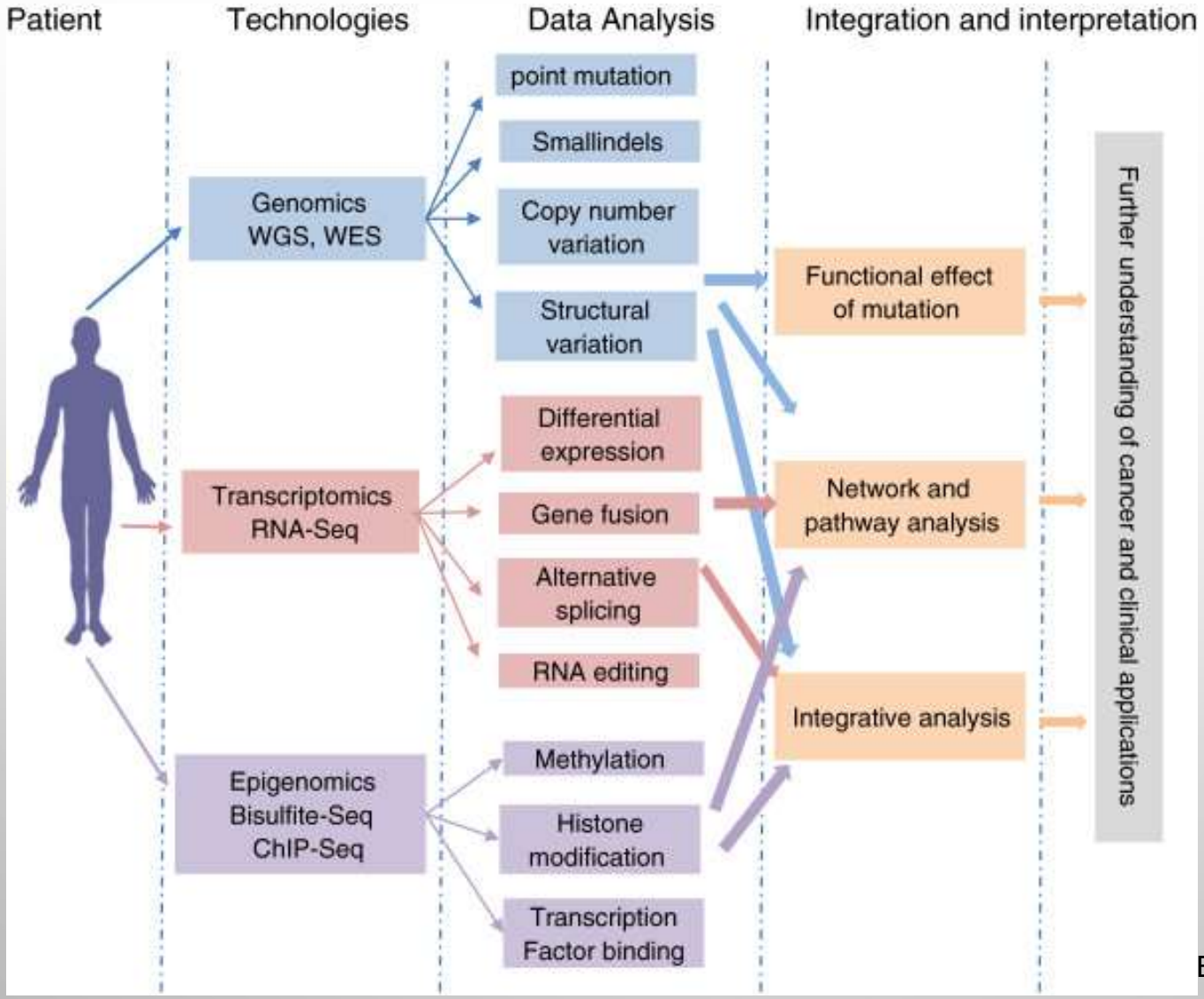


4. Data analysis—during data analysis and alignment, the newly identified sequence reads are then aligned to a reference genome. Following alignment, many variations of analysis are possible such as single nucleotide polymorphism (SNP) or insertion-deletion (indel) identification, read counting for RNA methods, phylogenetic or metagenomic analysis, and more.



https://www.youtube.com/watch?v=l4BAfRekohk

NGS applications



- DNA LEVEL:
  - WHOLE GENOME SEQUENCING
  - GENOME DE NOVO
  - EXOME & TARGET REGION
  - CHIP SEQUENCING
- RNA LEVEL:
  - TRANSCRIPTOME
  - MIRNA EXPRESSION
- SPECIAL PROJECT:
  - EPIGENETICS (METHYLATION)
  - METAGENOMICS (菌相分析)

Biological Procedures Online 15(1):4

whole exome sequencing (WES)

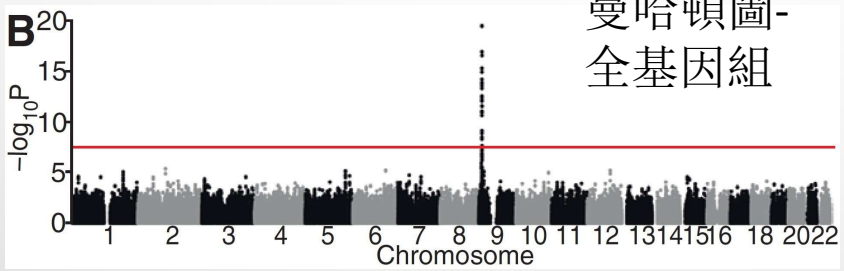
Whole genome sequencing (WGS)

(1) Genome-wide association study, GWAS (全基因組關聯分析)

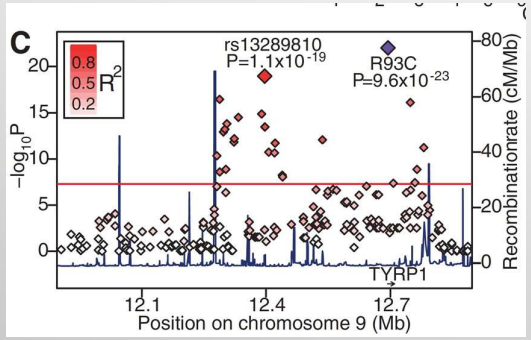
- 由全基因組中找出變異序列或基因，個體間的遺傳背景不宜差異太大
- 金髮個體：42位， 黑髮個體：43位
- 全基因組定序→ GWAS分析→候選基因→ 基因功能分析

42位

43位

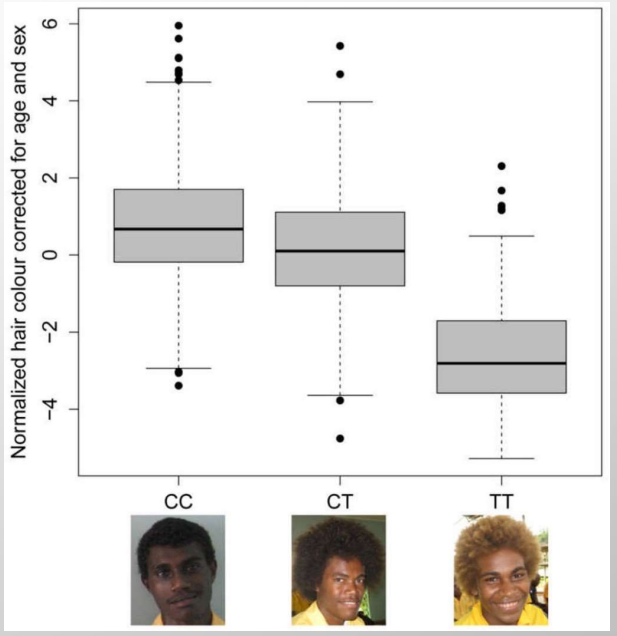


曼哈頓圖-
全基因組



曼哈頓圖-
基因TRYP1
R83C 突變
R93C 突變

R93C DNA突變



19種癌症次世代基因定序納健保 最高補助3萬最快5月上路

（中央社記者陳婕翎2024/2/6 19:25）

包含14種實體瘤及5種血液腫瘤

實體瘤NGS檢測給付14種癌別，包含非小細胞肺癌、三陰性乳癌、卵巢癌／輸卵管癌／原發性腹膜癌、大腸直腸癌、攝護腺癌、泌尿道上皮癌、肝內膽管癌、黑色素瘤、腸胃道間質瘤、甲狀腺癌、甲狀腺髓質癌、胰臟癌、NTRK基因融合實體腫瘤、胃癌。

石崇良表示，實體瘤NGS檢測規劃出3種模式:

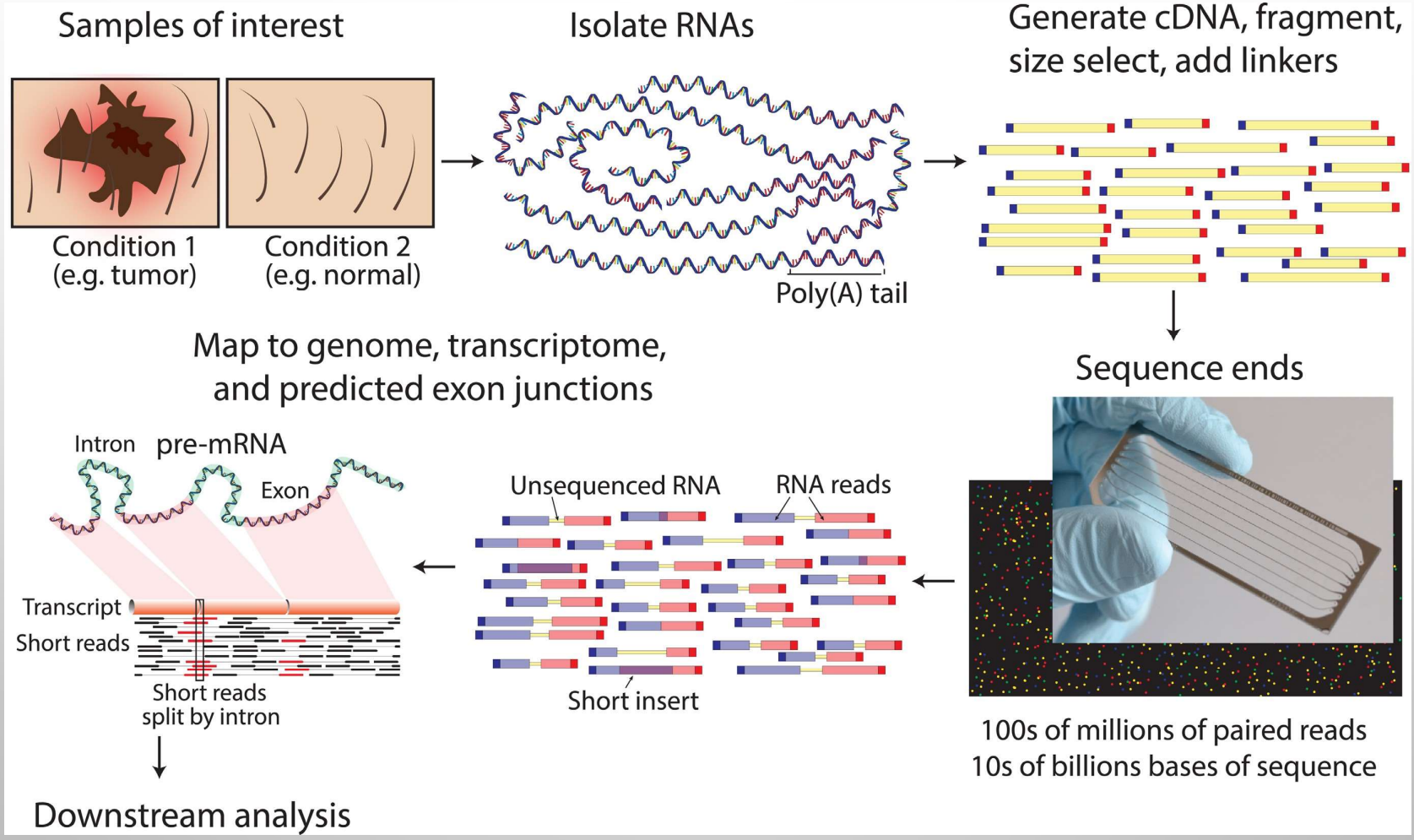
第1種BRCA-1與BRCA-2基因，基因片段較長，採獨立給付，預計給付健保點值1萬點；

第2種是各癌別最基本基因位點小於100個基因的小套組，給付2萬點；

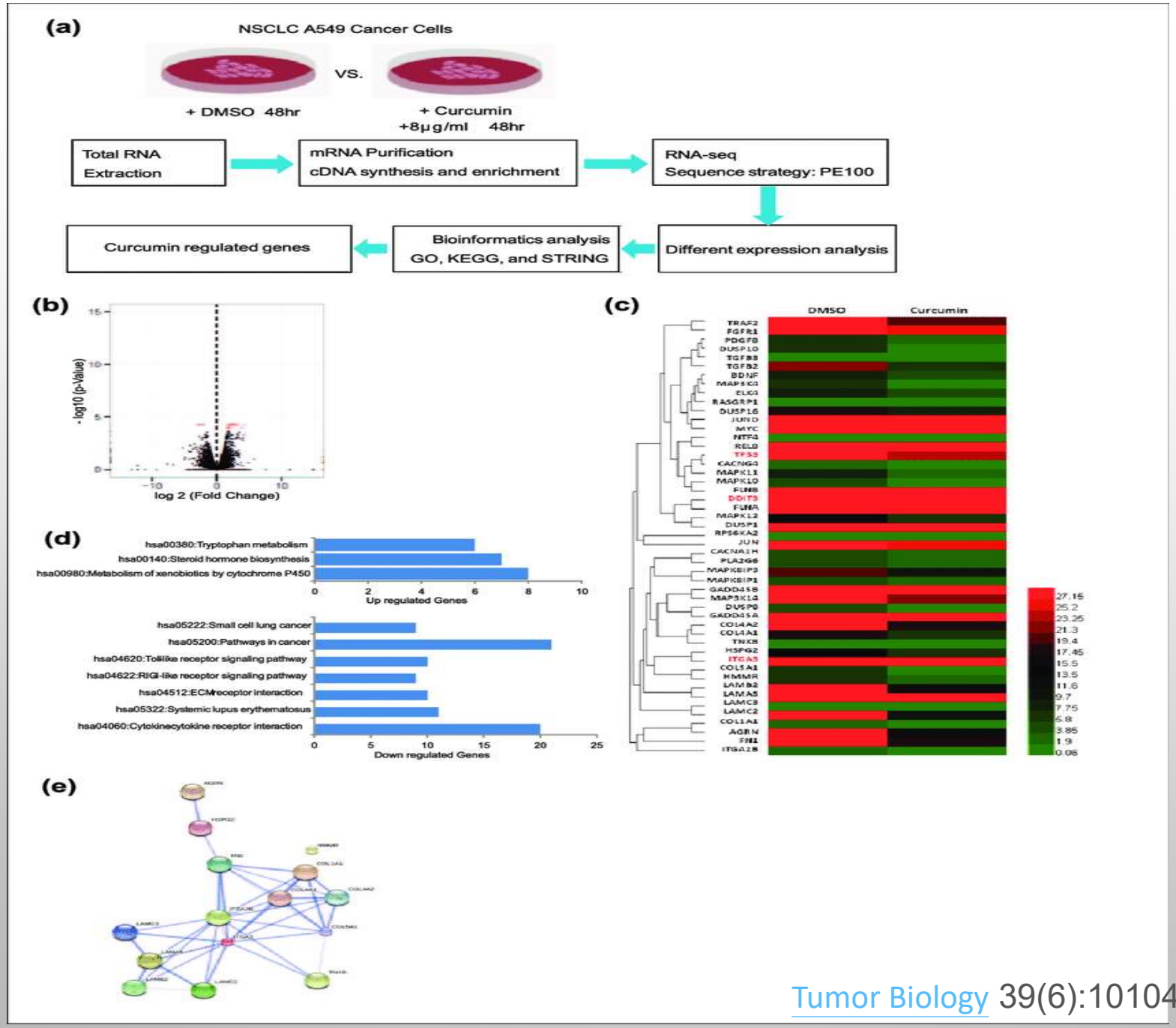
第3種100個以上基因位點大套組，給付3萬點。

(2)轉錄體定序(RNA-seq)

抽取RNA → 轉成雙股DNA並切成小片段→ 霰彈式定序→ 對應到基因組相對位置



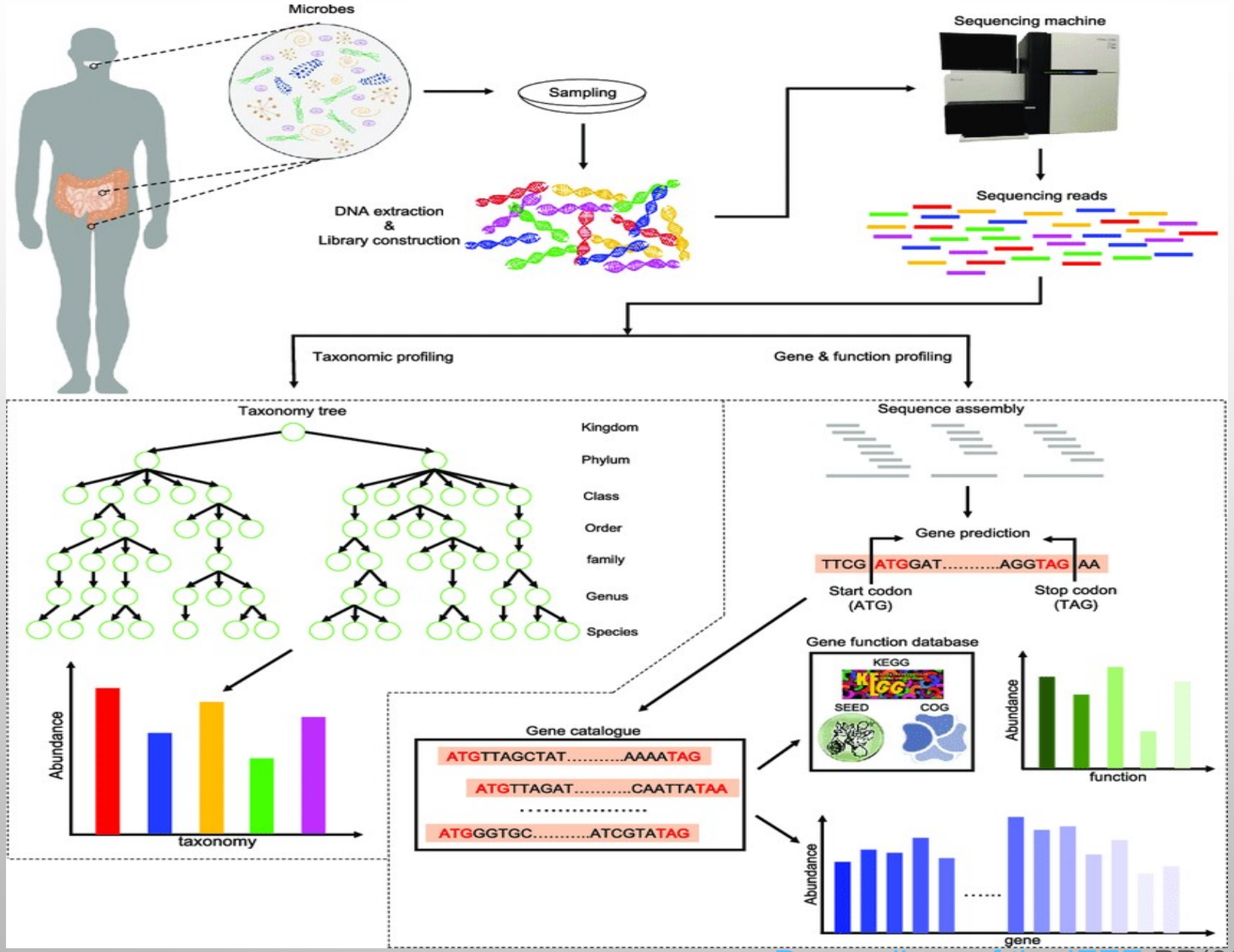
Transcriptome alteration analysis with RNA-seq in A549 NSCLC



Tumor Biology 39(6):1010428317705333

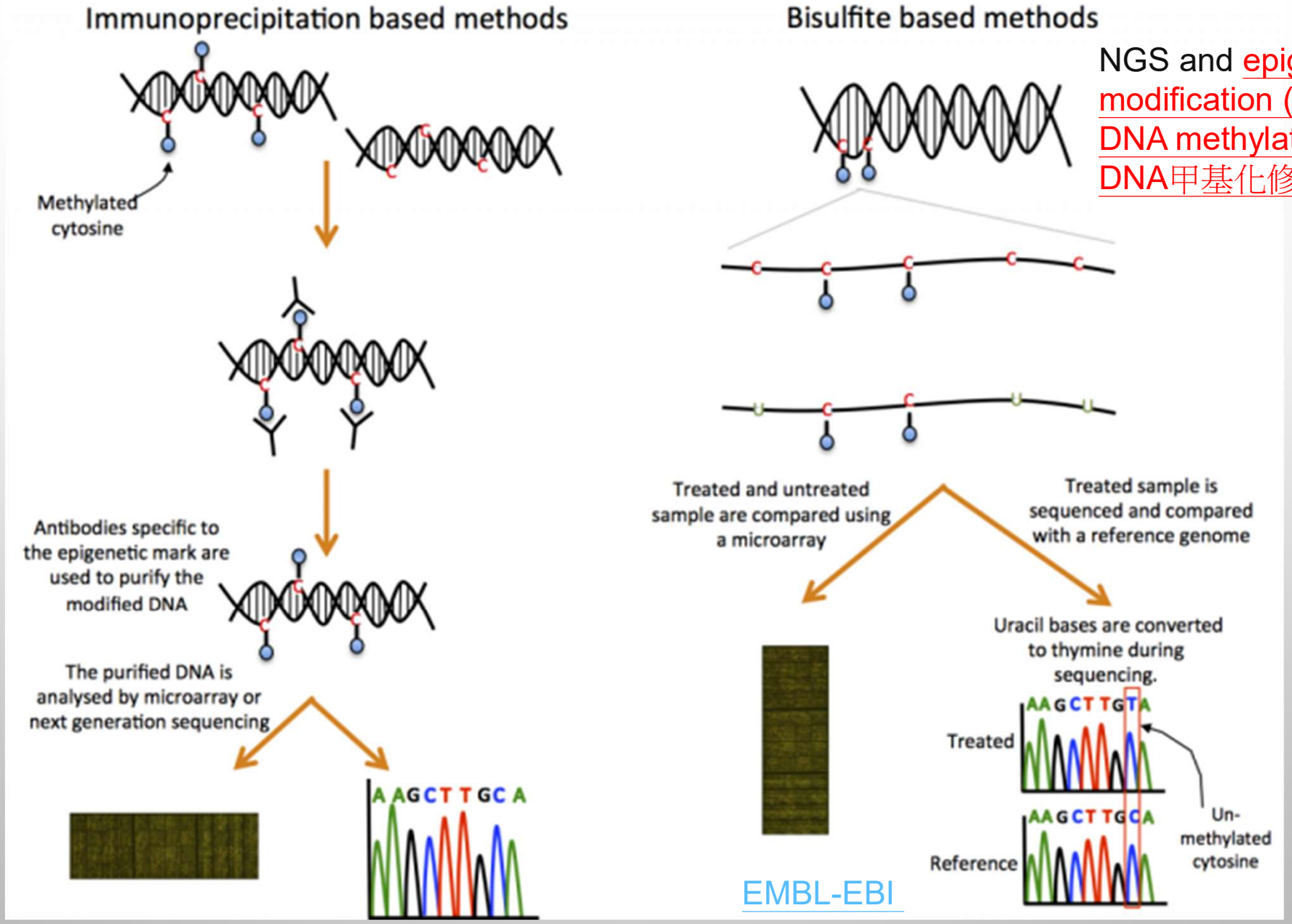
(3)微生物檢定(DNA-seq)

Metagenomic sequencing (Bacterial DNA sequence)



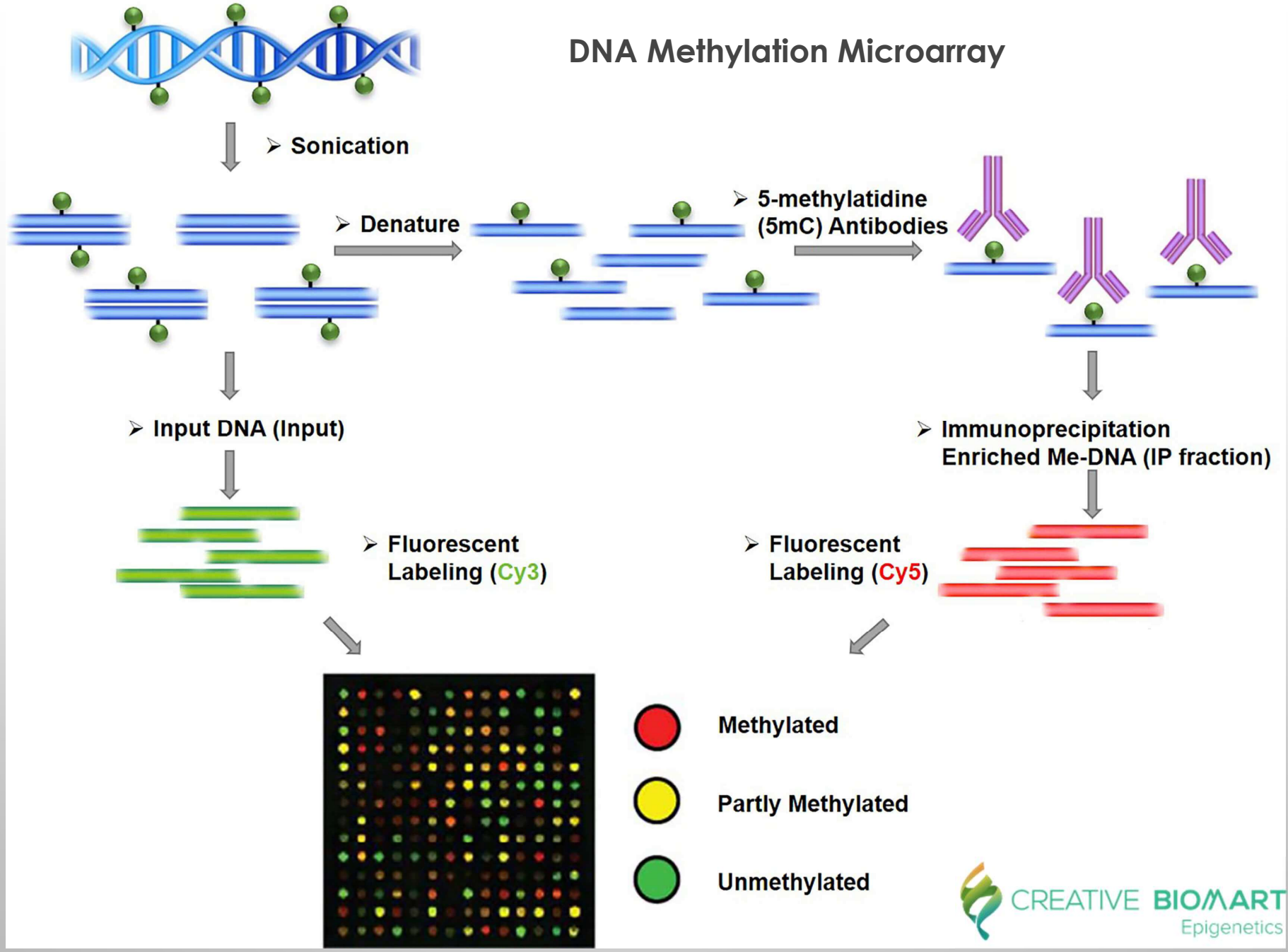
Proceedings of the IEEE PP(99):1-15

(4)後遺傳基因修飾(DNA-seq)



NGS and epigenetic modification (ex. DNA methylation-DNA甲基化修飾)

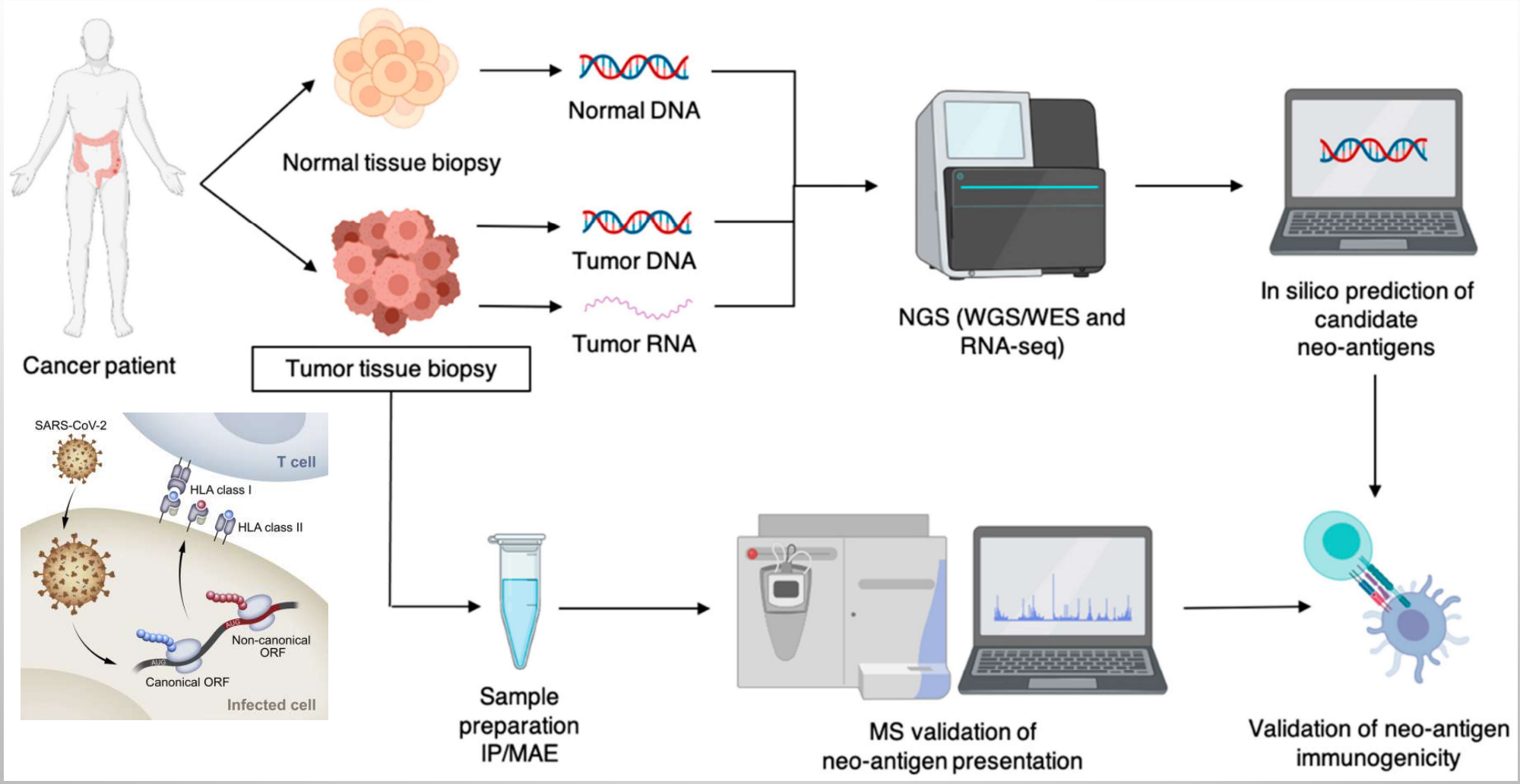
DNA Methylation Microarray



(5)疫苗抗原篩選

Neo-Antigen mRNA Vaccines

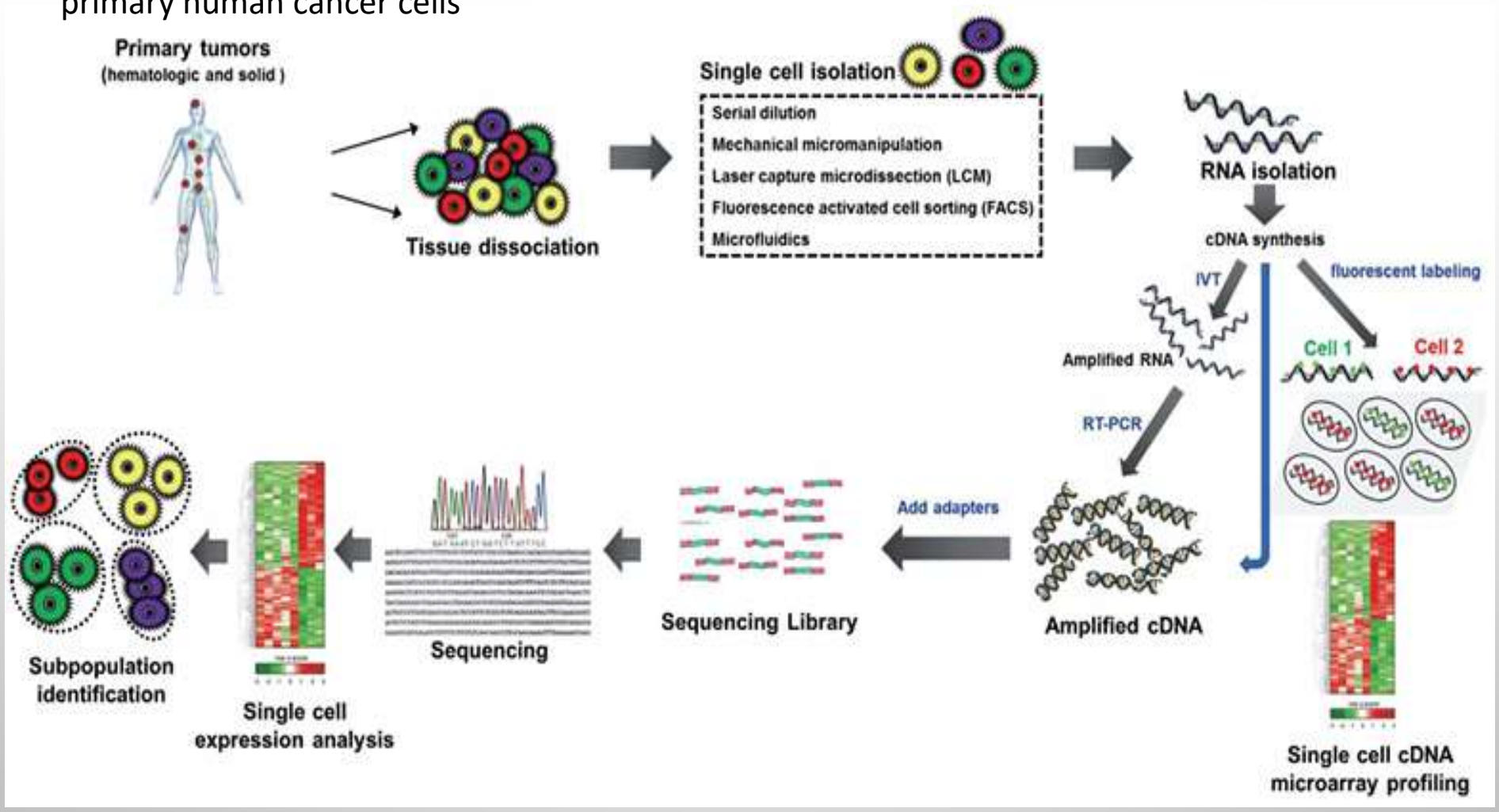
assessment of both human and pathogen genomes, their transcriptomes, as well as examination of host immune responses, such as T and B cell diversity. NGS can be used to assess the quality of vaccine stocks, the diversity of HLA polymorphisms in large populations, and also for detection of new pathogenic strains in mixed samples.



Vaccines **2020**, *8*(4), 776

(6)單細胞定序 (Single cell sequencing)

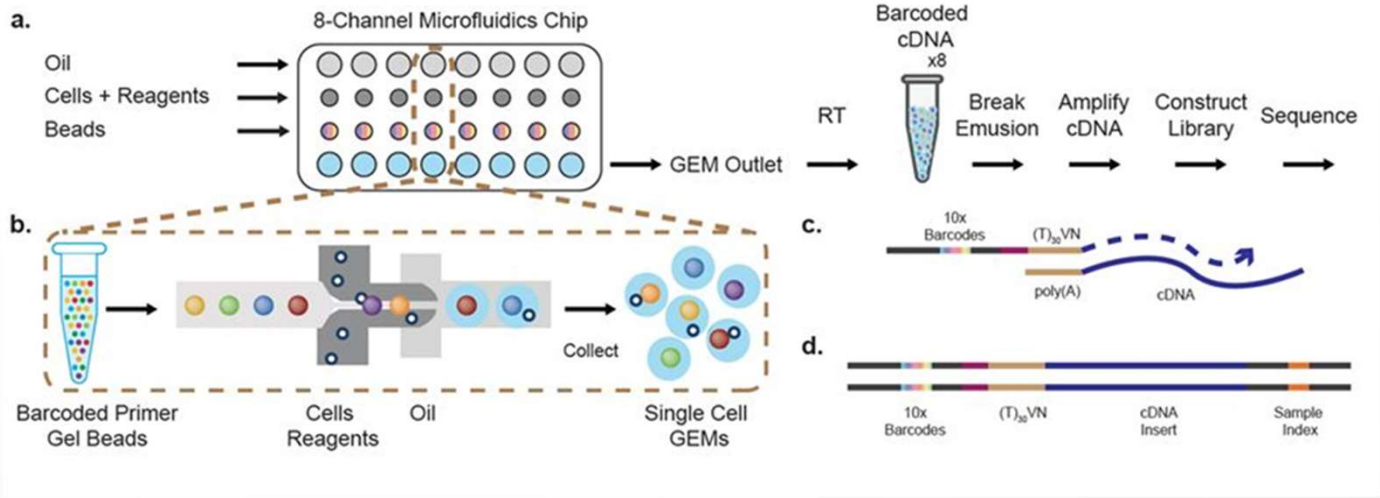
primary human cancer cells

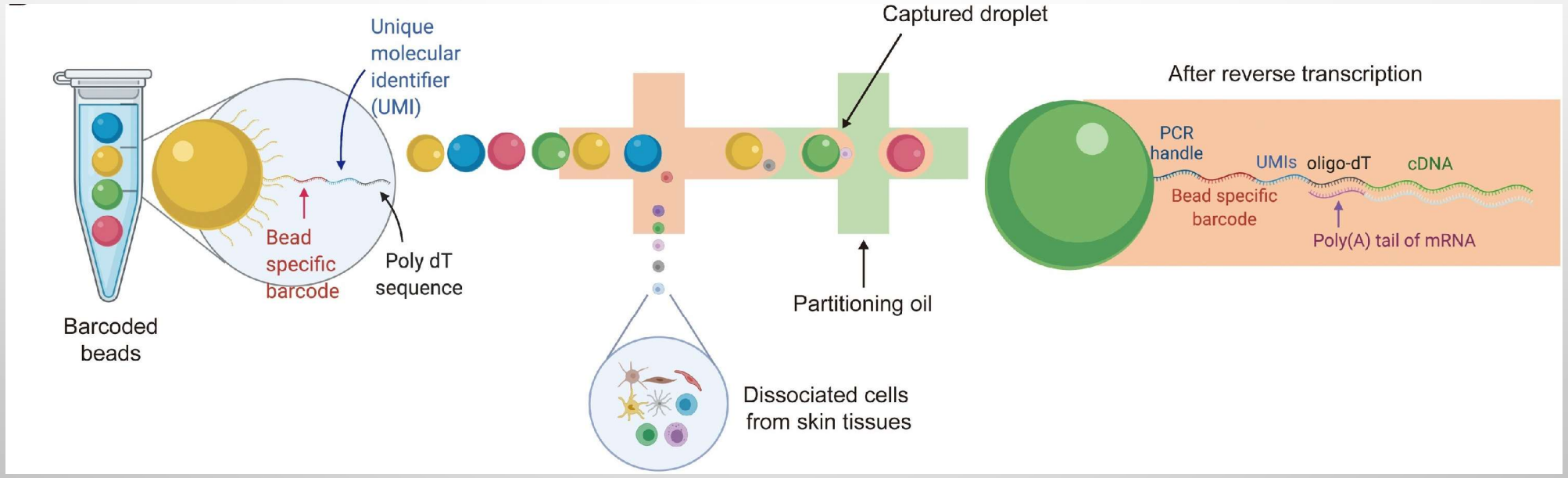


Journal of Liver Cancer 2018;18(1):1-8.

單細胞解析系統原理

Single cell RNA sequencing experiments routinely target the retention of 500 to 10,000 cells which require an input of 800 to 16,000 cells.



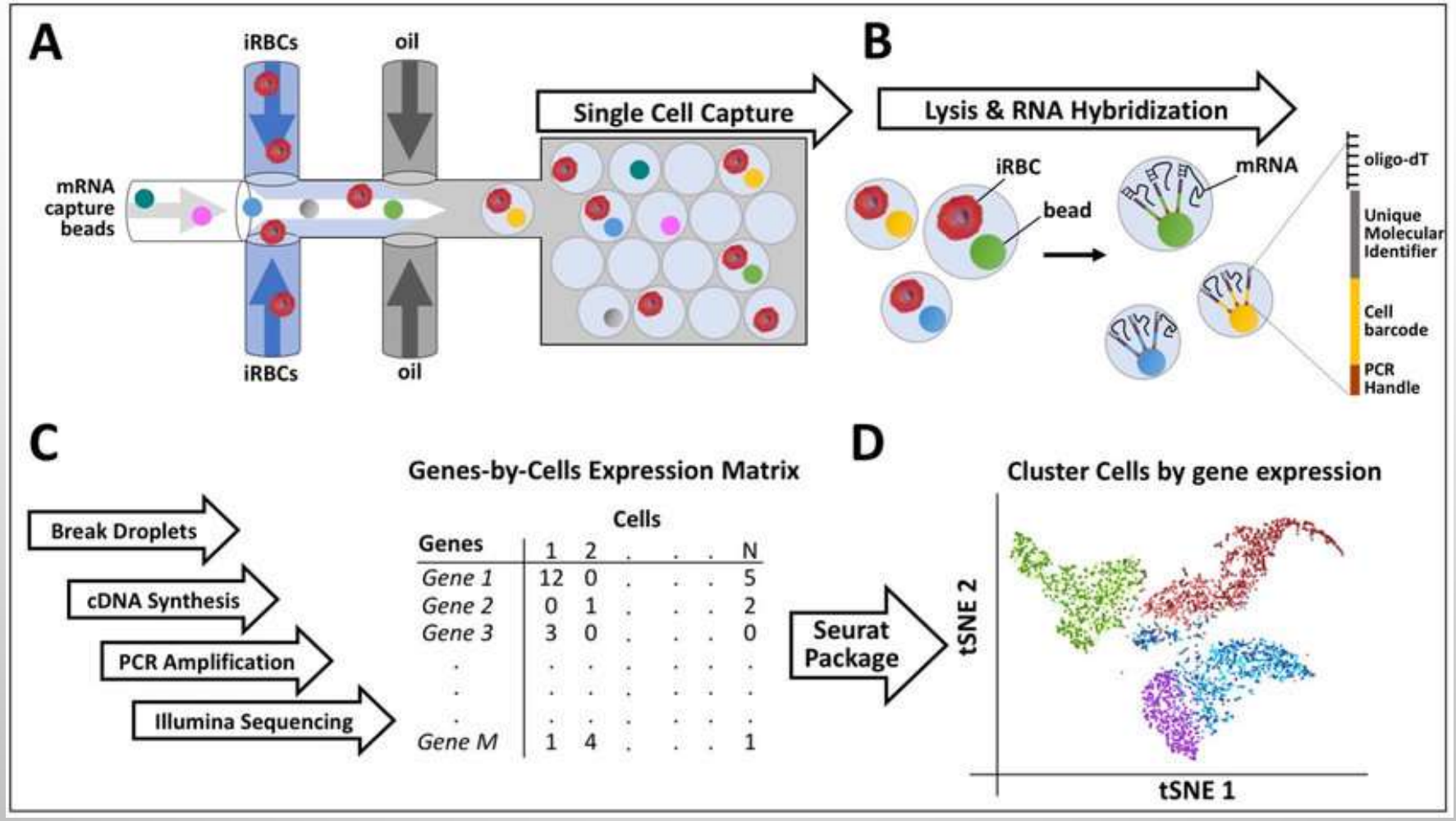


https://youtu.be/Rqz1QrmWcSM

https://youtu.be/Q5pkJITEgxo

https://www.youtube.com/watch?v=dbE1UlpxzHQ

https://doi.org/10.1016/j.jdermsci.2020.06.002



•November 2017 Nature 551(7678)


take a
BREAK

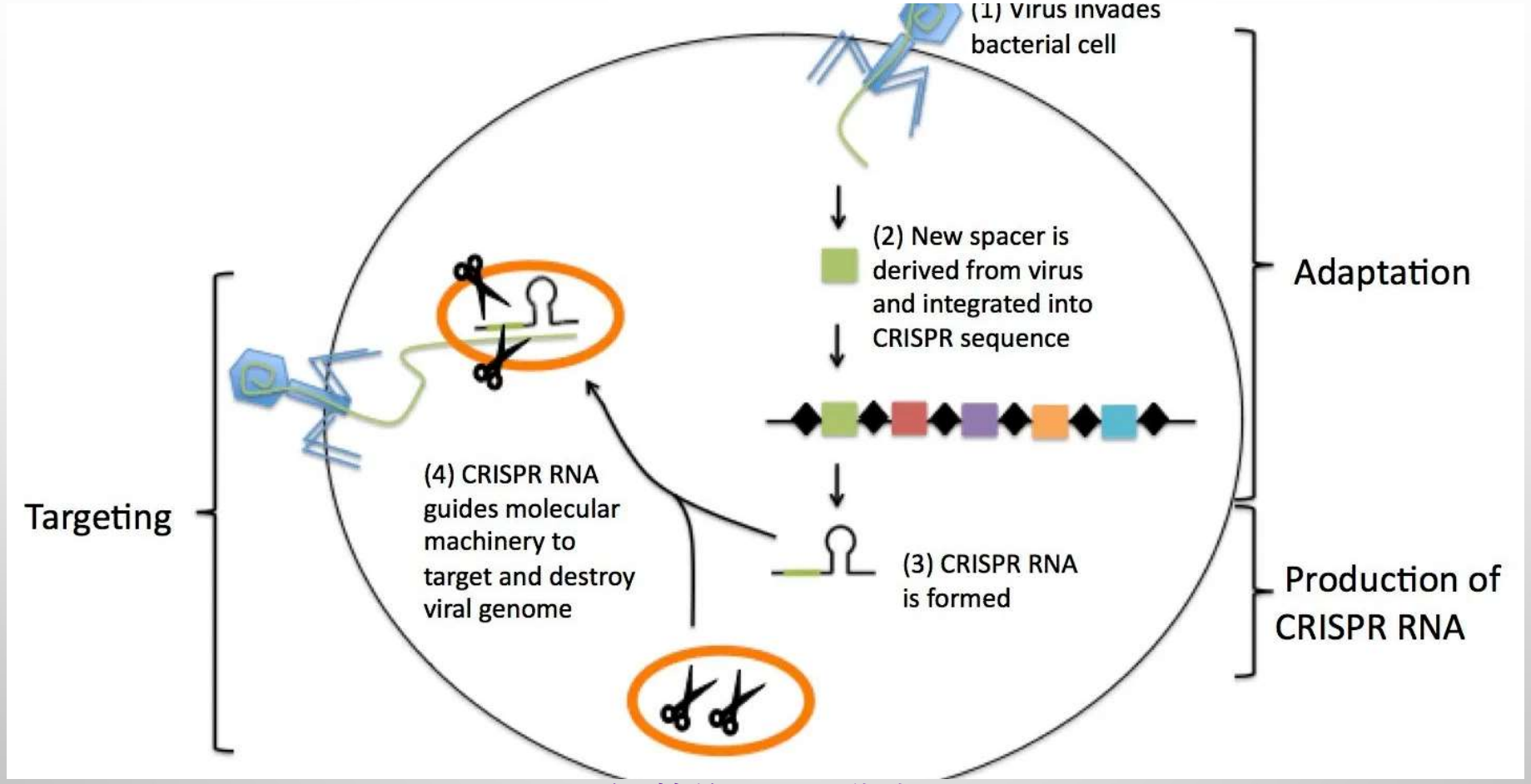
CRISPR/cas9 genome editing
(基因編輯技術)

CRISPR-CAS9(Clustered regularly interspaced short palindromic repeat / CRISPR associated protein 9)

- The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR- associated (Cas) system is the latest addition to the genome editing toolbox, offering a simple, rapid, and efficient solution.

- Derived from components of a simple bacterial immune system, the CRISPR-Cas9 system permits targeted gene cleavage and gene editing in a variety of eukaryotic cells, and because the endonuclease cleavage specificity in CRISPR-Cas9 system is guided by RNA sequences, editing can be directed to virtually any genomic locus by engineering the guide RNA sequence and delivering it along with the Cas endonuclease to your target cell.

Bacterial immune system



細菌的Cas9 酵素

Nobel Prize in Chemistry (2020)

https://mcbblog.nsfbio.com

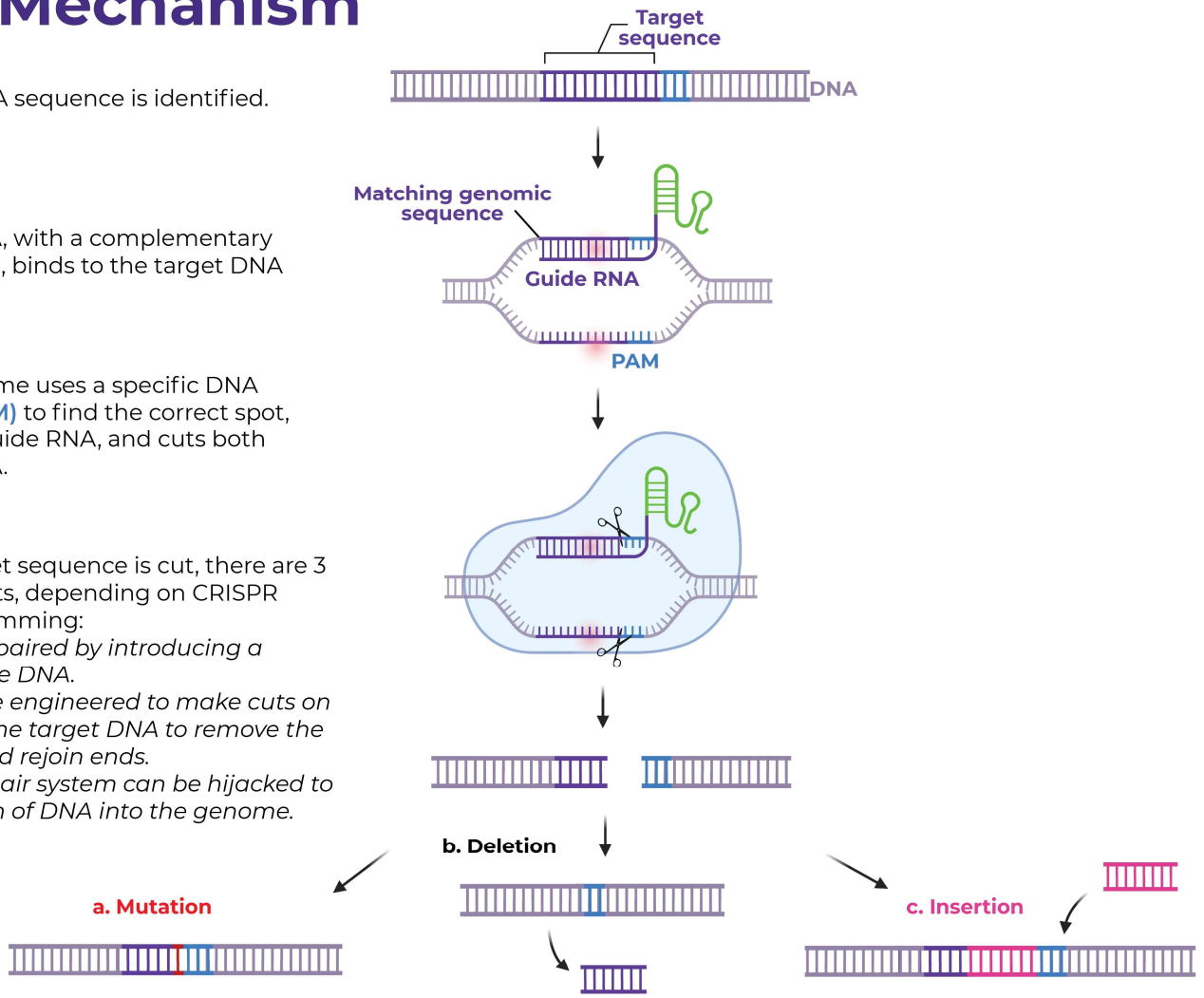
CRISPR Mechanism

① The target DNA sequence is identified.

② The guide RNA, with a complementary DNA sequence, binds to the target DNA sequence.

③ The Cas9 enzyme uses a specific DNA sequence **(PAM)** to find the correct spot, binds to the guide RNA, and cuts both strands of DNA.

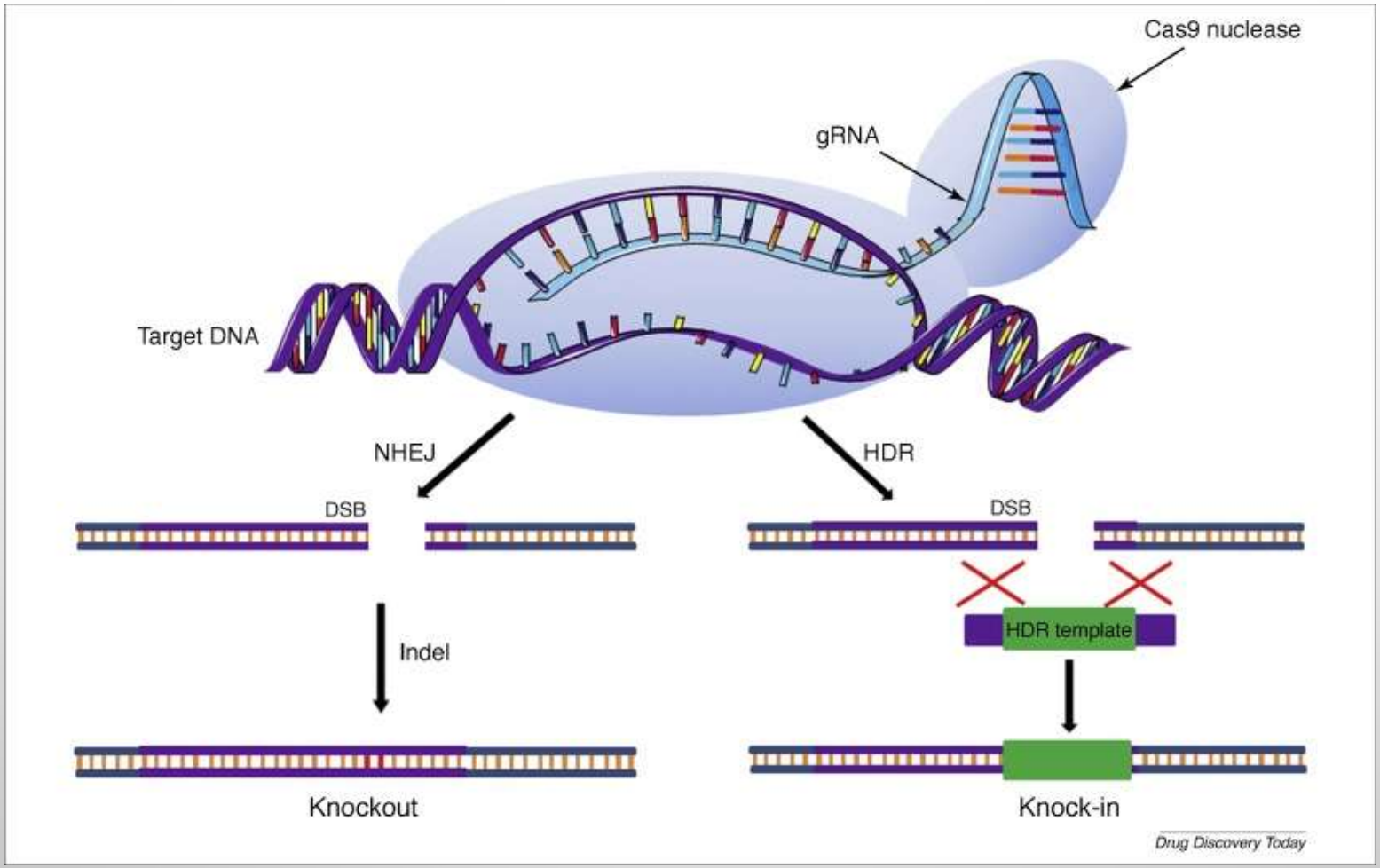
④ Once the target sequence is cut, there are 3 potential results, depending on CRISPR system programming:
a. The cut is repaired by introducing a mutation in the DNA.
b. Enzymes are engineered to make cuts on either side of the target DNA to remove the target DNA and rejoin ends.
c. The DNA repair system can be hijacked to insert a section of DNA into the genome.



https://www.youtube.com/watch?v=UKbrwPL3wXE

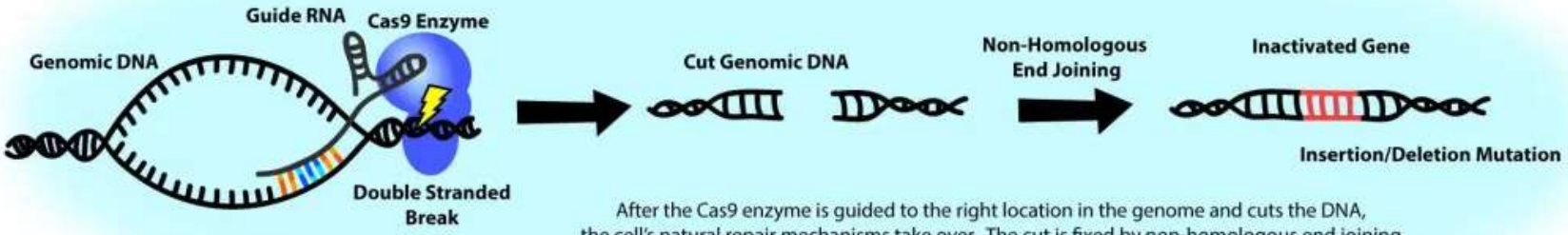
https://www.umassmed.edu/rti/biology/crispr-cas9/

Functions of CRISPR-cas9



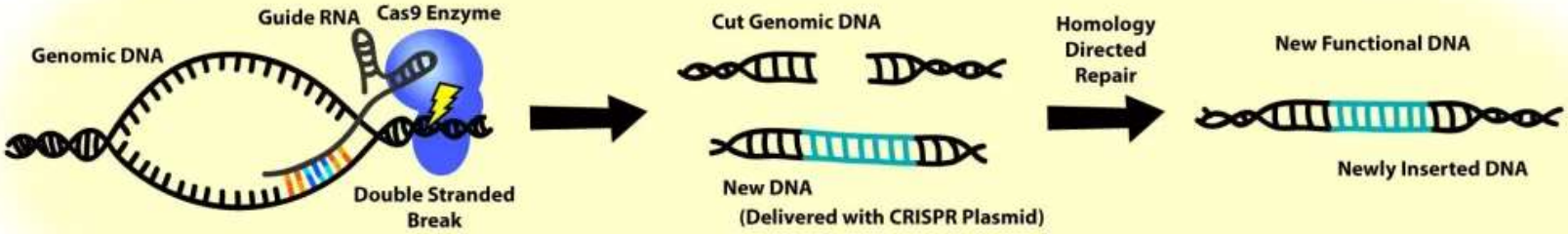
https://www.youtube.com/watch?v=g7bkE1krgFM

Gene Silencing with CRISPR



After the Cas9 enzyme is guided to the right location in the genome and cuts the DNA, the cell's natural repair mechanisms take over. The cut is fixed by non-homologous end joining. This process is error-prone and does not perfectly replace the cut DNA, often resulting in an insertion or deletion mutation which silences the gene.

Gene Insertion with CRISPR



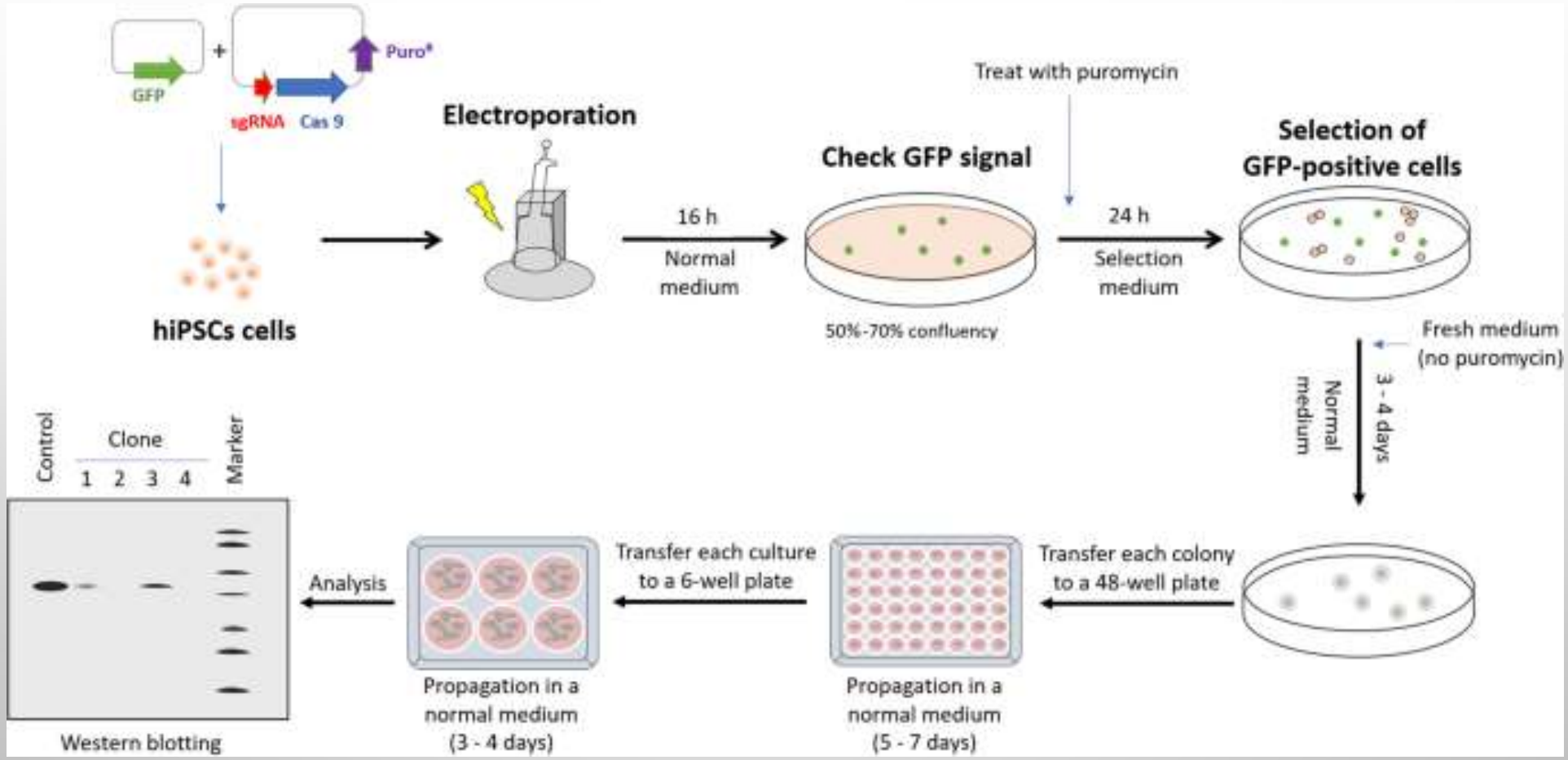
To insert a gene, the new gene is added into the original CRISPR plasmid. It is designed to line up perfectly with the cut DNA strands, so the cell uses a different technique, homology directed repair, to incorporate a new stretch of DNA into the genome.

https://www.youtube.com/watch?v=pNseb_U6gu0

by Christopher Gerry

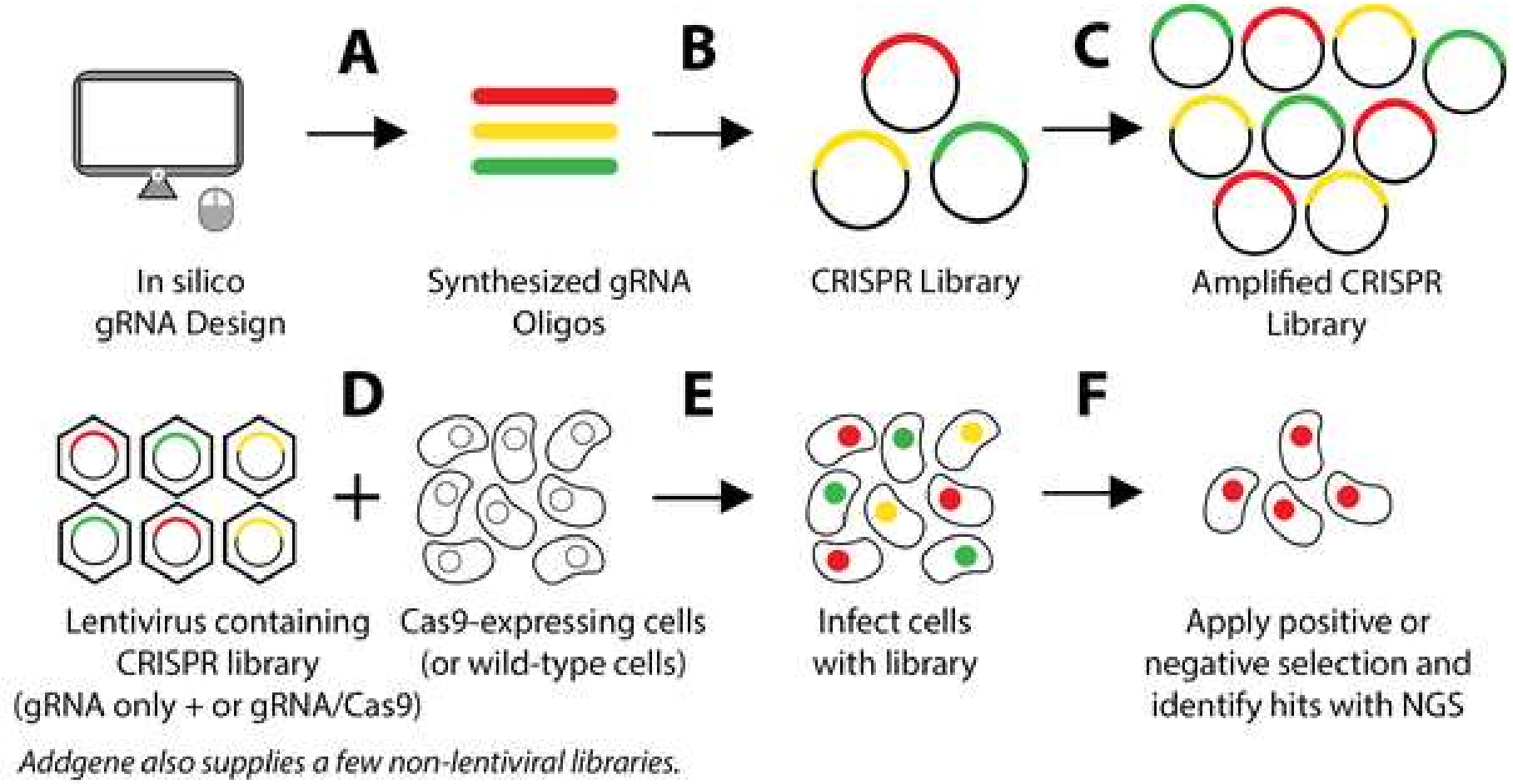
CRISPR protocol

(a) single sgRNA

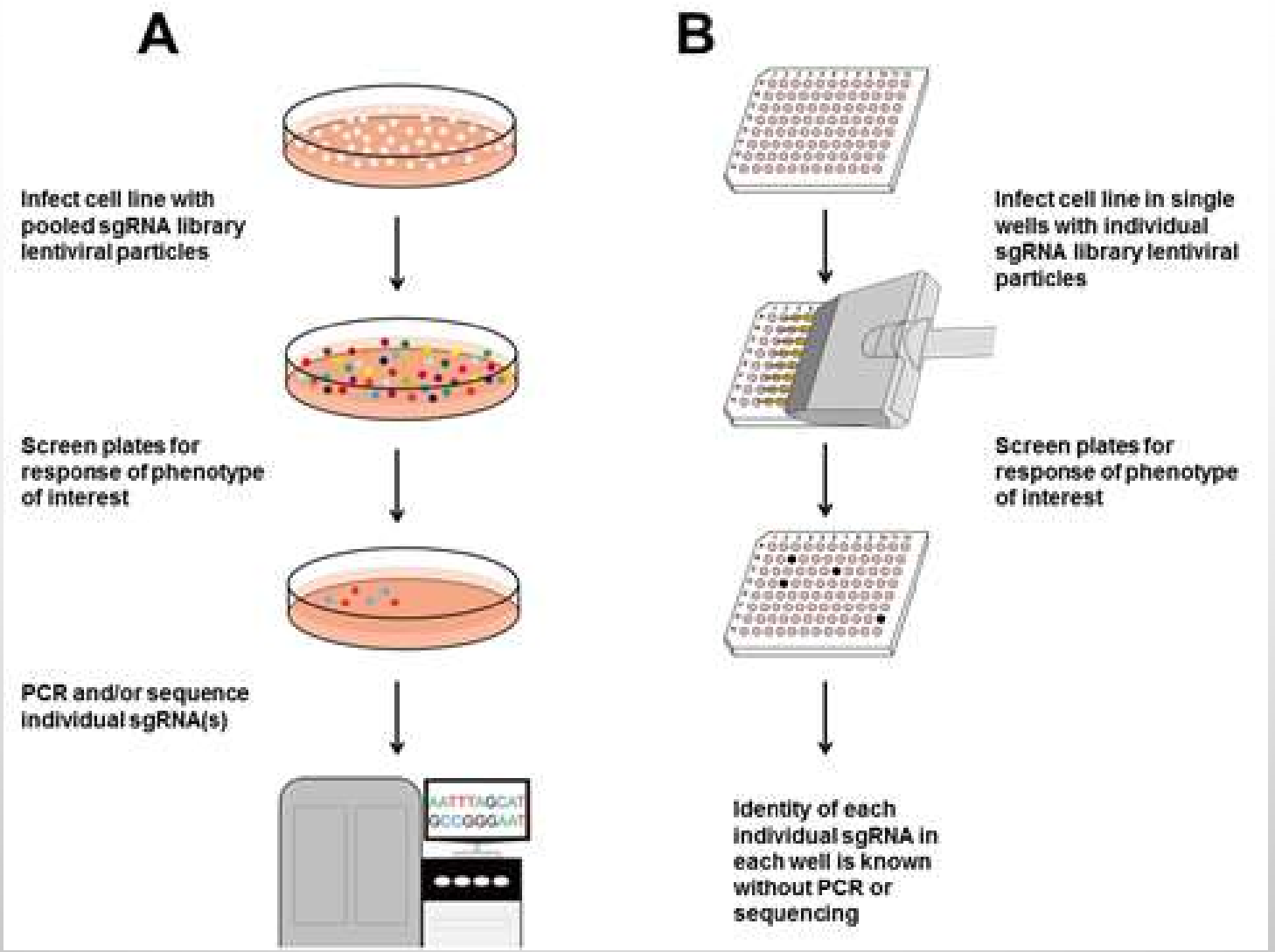


Acta Pharmacologica Sinica volume 41, pages1427–1432 (2020)

(b) CRISPR sgRNA library screening

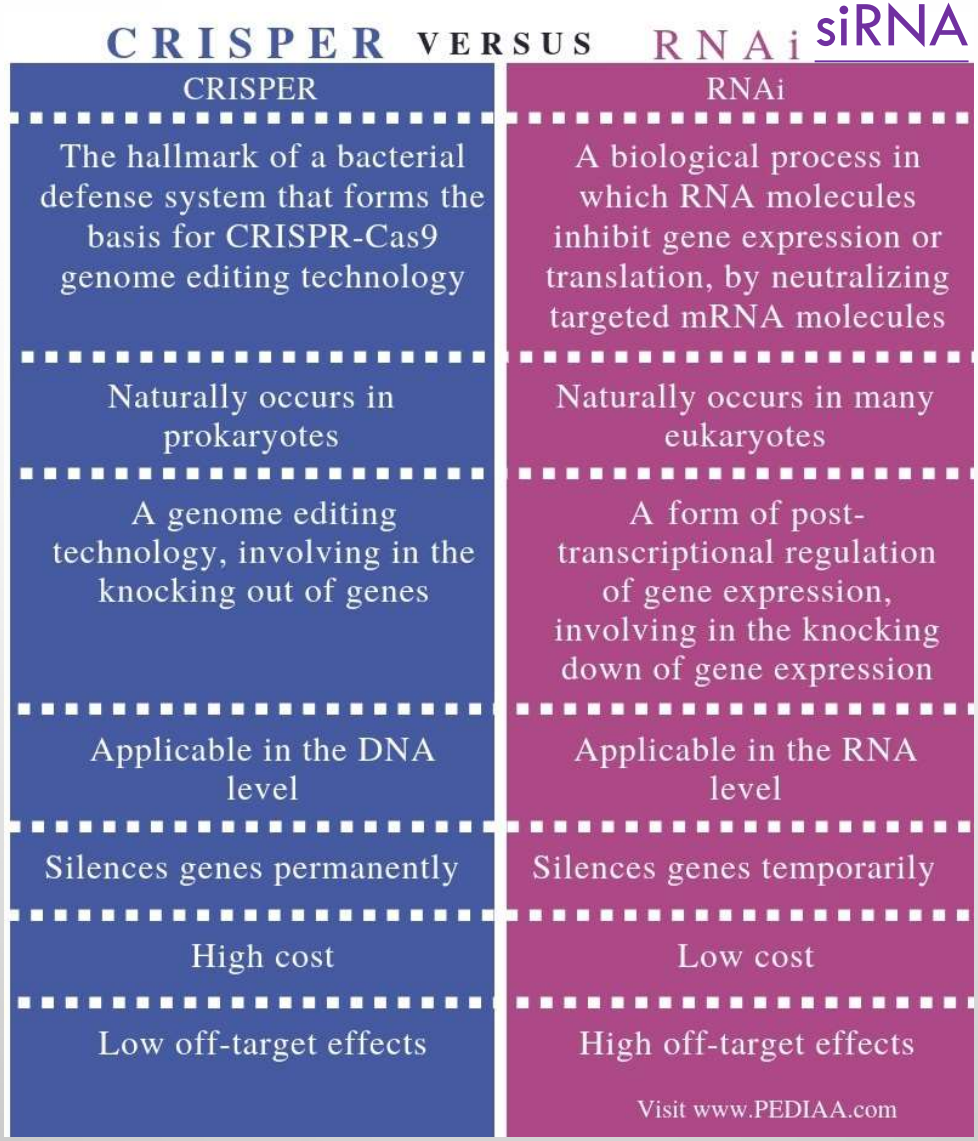


addgene



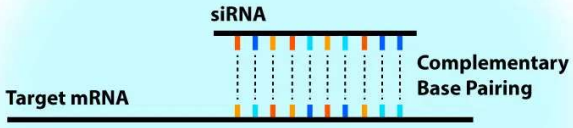
GeneCopoeia™
Expressway to Discovery

Compared to RNAi

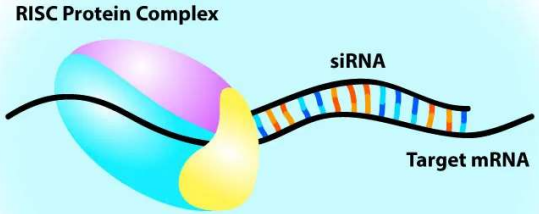


https://cn.weblogographic.com/what-is-difference-between-crispr

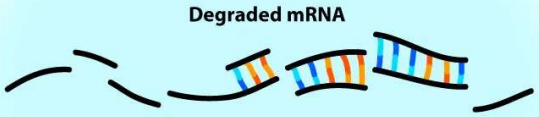
RNAi



1) Small interfering RNA (siRNA) is designed to specifically target mRNA that's derived from a particular gene.

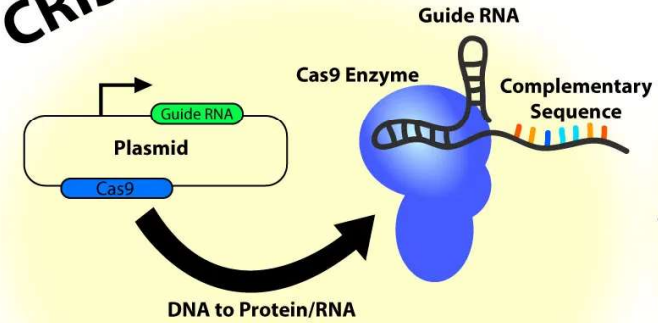


2) Once it's been introduced into a cell, siRNA binds to target mRNA with the help of a piece of cellular machinery called the RISC protein complex.

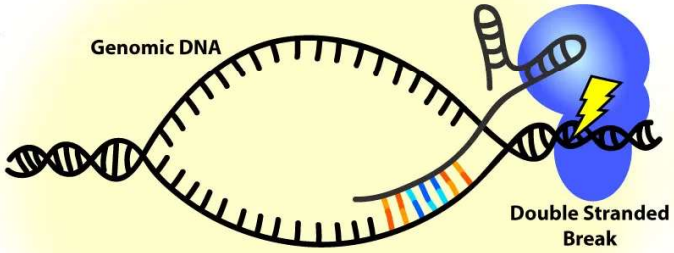


3) After binding, the siRNA promotes degradation of the target mRNA, resulting in translational repression and lower levels of the corresponding protein.

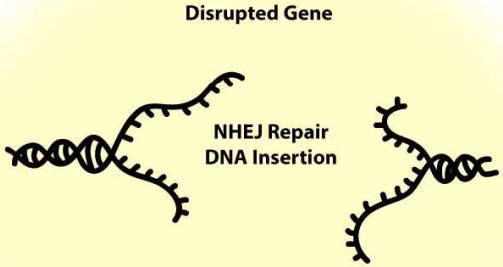
CRISPR



1) A plasmid, a circular piece of DNA, encodes for both the Cas9 enzyme and a "guide RNA" that's complementary to the target gene's DNA sequence.



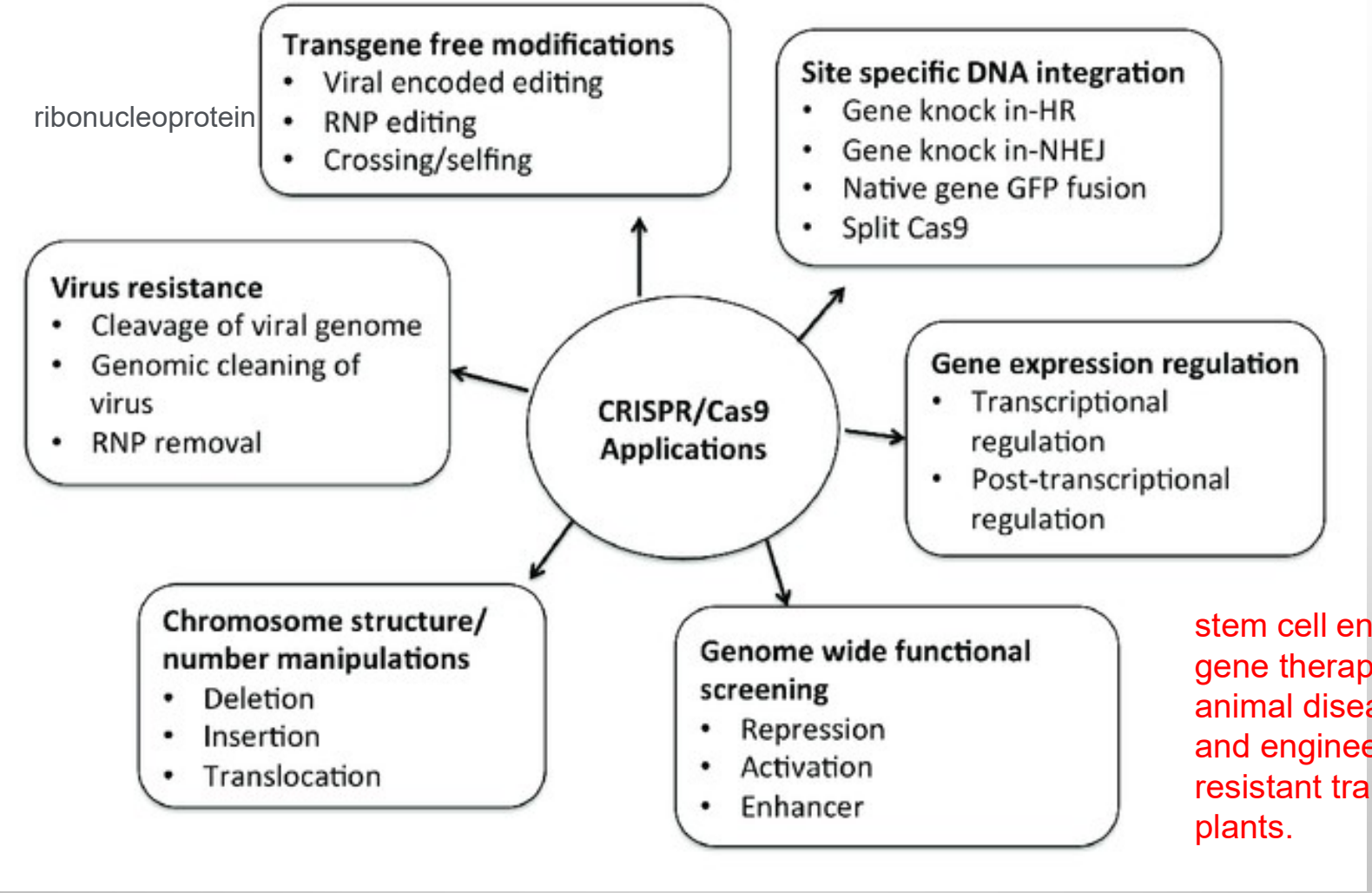
2) After the plasmid enters the cell, the resulting Cas9/guide RNA complex binds the target DNA sequence and Cas9 cuts both strands of DNA.



3) The cell repairs the DNA break via non-homologous end joining (NHEJ), which is prone to errors that can silence the gene, or homology directed repair, which enables high-precision editing of the gene.

by Christopher Gerry

Applications of CRISPR system



stem cell engineering, gene therapy, tissue and animal disease models, and engineering disease-resistant transgenic plants.

Frontiers in Plant Science 8:1932 · November 2017

Genome analysis
(基因體分析)

(1) Open reading frame(ORF): 開放讀序框架

— 基因組中能被轉譯成氨基酸序列的DNA片段
— 一條DNA序列共含有六種可能性的讀取框架(reading frame)
— 框架讀取： 每3個鹼基為一組
— 若讀取框架中，含有起始(ATG, Methionine)及終止碼(TAA, TAG, TGA),且該讀取框架夠長，則其可能為一個基因
—若讀取框架片段沒有包含起始或終止碼，則該片段則有可能是某基因片段

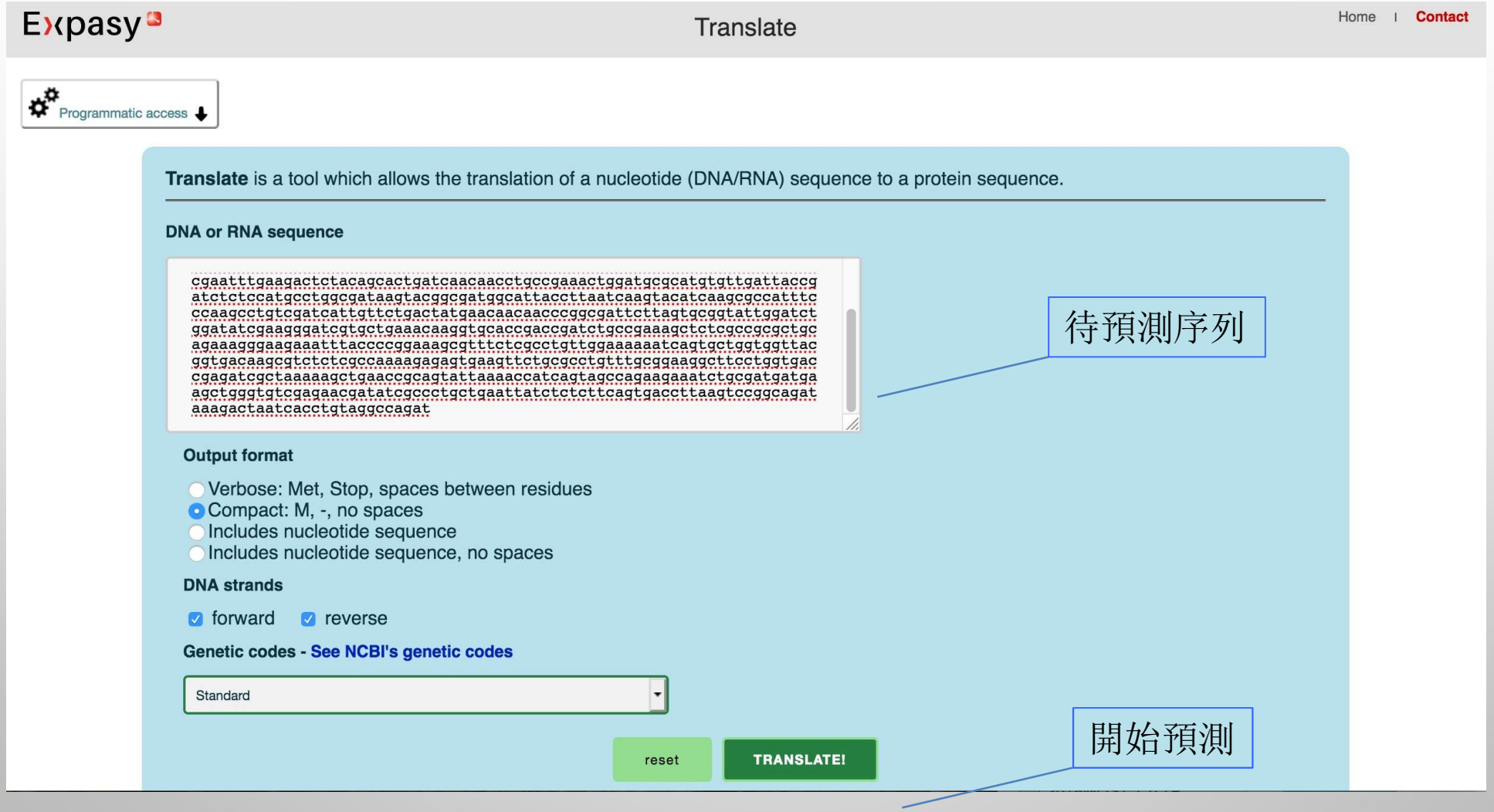
Frame 1 -> K P C - I W - A S S C E
Frame 2 -> N H A K S G R Q A V V K
Frame 3 -> T M L N L V G K Q L - K
G H - I Q Y A L L Q S F <- Frame 4
V M S F R T P L C N H F <- Frame 5
F W A L D P L C A T T F <- Frame 6

Q：假設你有一條來自大腸桿菌序列如下，請問你要如何以尋找ORF的方式預測該片段是否含有基因片段？

aaacggctcatcgtcttaaaggcgtatttgccatgctaaatctggtacccggcaagcagttat
gtgaaacgctggaacatctgattcgtgagaaggatgttccaggaatagaaaaatacatcag
cgacattgacagttatgtcaagagcttgctgtagcaaggtagcctattacatgaacaatatg
aacgtaattattgccgatgaccatccgatagtcttgttcggtattcgcaaatcacttgagcaa
attgagtgggtgaatgttgtcggcgaatttgaagactctacagcactgatcaacaacctgcc
gaaactggatgcgcatgtgttgattaccgatctctccatgcctggcgataagtacggcgatg
gcattaccttaatcaagtacatcaagcgccatttcccaagcctgtcgatcattgttctgactat
gaacaacaacccggcgattcttagtgcggtattggatctggatatcgaagggatcgtgctga
aacaaggtgcaccgaccgatctgccgaaagctctcgccgcgctgcagaaagggaagaaat
ttaccccggaaagcgtttctcgcctgttggaaaaaatcagtgctggtggttacggtgacaag
cgtctctcgccaaaagagagtgaagttctgcgcctgtttgcggaaggcttcctggtgaccga
gatcgctaaaaagctgaaccgcagtattaaaaccatcagtagccagaagaaatctgcgatg
atgaagctgggtgtcgagaacgatatcgccctgctgaattatctctcttcagtgaccttaagt
ccggcagataaagactaatcacctgtaggccagat

ORF預測軟體－ TRANSLATE TOLLS IN EXPASY

https://web.expasy.org/translate/

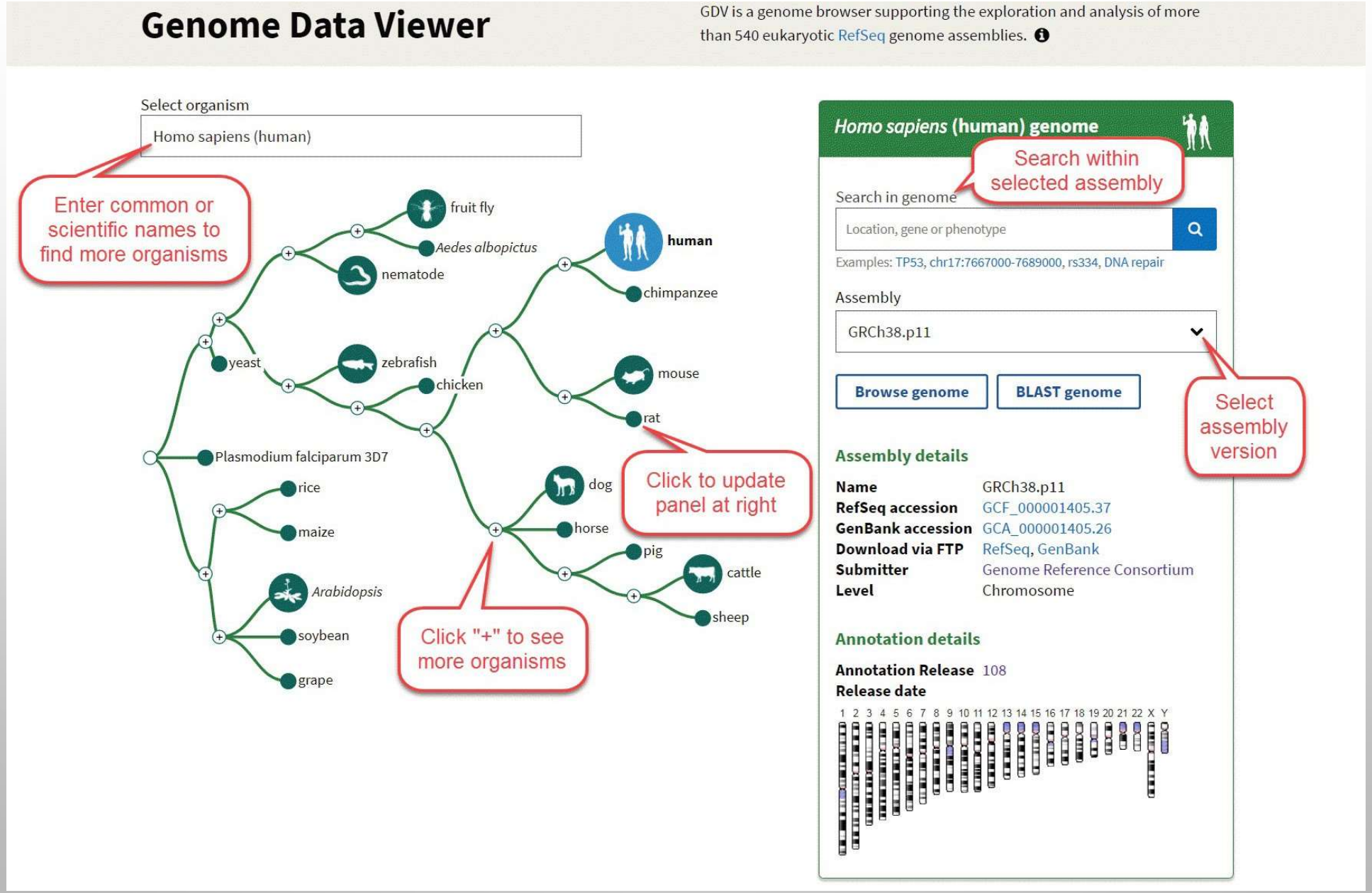


ORF預測結果

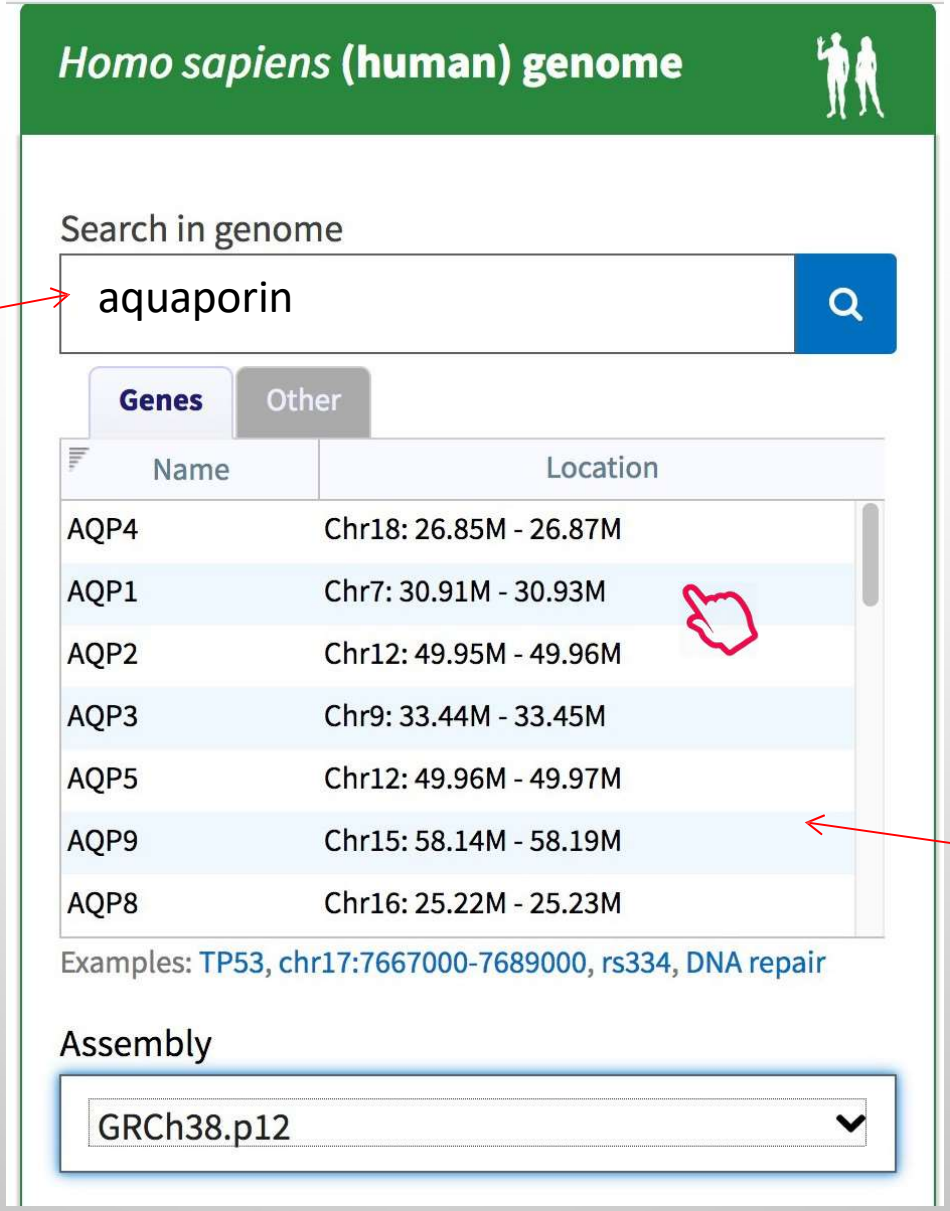

5'3' Frame 1
KRLIVLKAYLPC-IWYPASSYVKRWNI-FVRRMFQE-KNTSATLTVMSRACCSKVAYYMNNMNVIIADDHPIVLFGIRKSLEQIEWVNVVGEFEDSTAL
INNLPKLDAHVLITDLSMPGDKYGDGITLIKYIKRHFPSLSIIVLTMNNNPAILSAVLDLDIEGIVLKQGAPTDLPKALAALQKGKKFTPESVSRLLEK
ISAGGYGDKRLSPKESEVLRLFAEGFLVTEIAKKLNRSIKTISSQKKSAMMKLGVENDIALLNYLSSVTLSPADKD-SPVGQ
5'3' Frame 2
NGSSS-RRICHAKSGTRQAVM-NAGTSDS-EGCSRNRKIHQRH-QLCQELAVAR-PIT-TI-T-LLPMTIR-SCSVFANHLSKLSG-MLSANLKTLQH-
STTCRNWMRMC-LPISPCLAISTAMALP-SSTSSAISQACRSLF-L-TTTRRFLVRYWIWISKGSC-NKVHRPICRKLSPRCRKGRNLPRKAFLACWKK
SVLVVTVTSVSRQKRVKFCACLRKASW-PRSLKS-TAVLKPSVARRNLR--SWVSRTISPC-IISLQ-P-VRQIKTNHL-AR
5'3' Frame 3
TAHRLKGVFAMLNLVPGKQLCETLEHLIREKDVPGIEKYISDIDSYVKSLL-QGSLLHEQYERNYCR-PSDSLVRYSQIT-AN-VGECCRRI-RLYSTD
QQPAETGCACVDYRSLHAWR-VRRWHYLNQVHQAPFPKPVDHCSDYEQQPGDS-CGIGSGYRRDRAETRCTDRSAESSRRAAEREEIYPGKRFSPVGKN
QCWWLR-QASLAKRE-SSAPVCGRLPGDRDR-KAEPQY-NHQ-PEEICDDEAGCRERYRPAELSLFSDLKSGR-RLITCRPD
3'5' Frame 1
IWPTGD-SLSAGLKVTEER-FSRAISFSTPSFIIADFFWLLMVLILRFSFLAISVTRKPSANRRRTSLSFGERRLSP-PPALIFSNRRETLSGVNFFPF
CSAARAFGRSVGAPCFSTIPSISRSNTALRIAGLLFIVRTMIDRLGKWRLMYLIKVMPSPYLSPGMERSVINTCASSFGRLLISAVESSNSPTTFTHSI
CSSDLRIPNKTIGWSSAIITFILFM--ATLLQQALDITVNVADVFFYSWNILLTNQMFQRFT-LLAGYQI-HGKYAFKTMSR
3'5' Frame 2
SGLQVISLYLPDLRSLKRDNSAGRYRSRHPASSSQISSGY-WF-YCGSAF-RSRSPGSLPQTGAELHSLLARDACHRNHQH-FFPTGEKRFPG-ISSLS
AARRELSADRSVHLVSARSLRYPDPIPH-ESPGCCS-SEQ-STGLGNGA-CT-LR-CHRRTYRQAWRDR-STHAHPVSAGC-SVL-SLQIRRQHSPTQF
AQVICEYRTRLSDGHRQ-LRSYCSCNRLPCYSKLLT-LSMSLMYFSIPGTSFSRIRCSSVSHNCLPGTRFSMANTPLRR-AV
3'5' Frame 3
LAYR-LVFICRT-GH-REIIQQGDIVLDTQLHHRRFLLATDGFNTAVQLFSDLGHQEAFRKQAQNFTLFWRETLVTVTTSTDFFQQARNAFRGKFLPFL
QRGESFRQIGRCTLFQHDPFDIQIQYRTKNRRVVVHSQNNDRQAWEMALDVLD-GNAIAVLIARHGEIGNQHMRIQFRQVVDQCCRVFKFADNIHPLNL
LK-FANTEQDYRMVIGNNYVHIVHVIGYLATASS-HNCQCR-CIFLFLEHPSHESDVPAFHITACRVPDLAWQIRL-DDEPF

(2) NCBI genome data viewer

https://www.ncbi.nlm.nih.gov/gdv/



Search “Aquaporin”



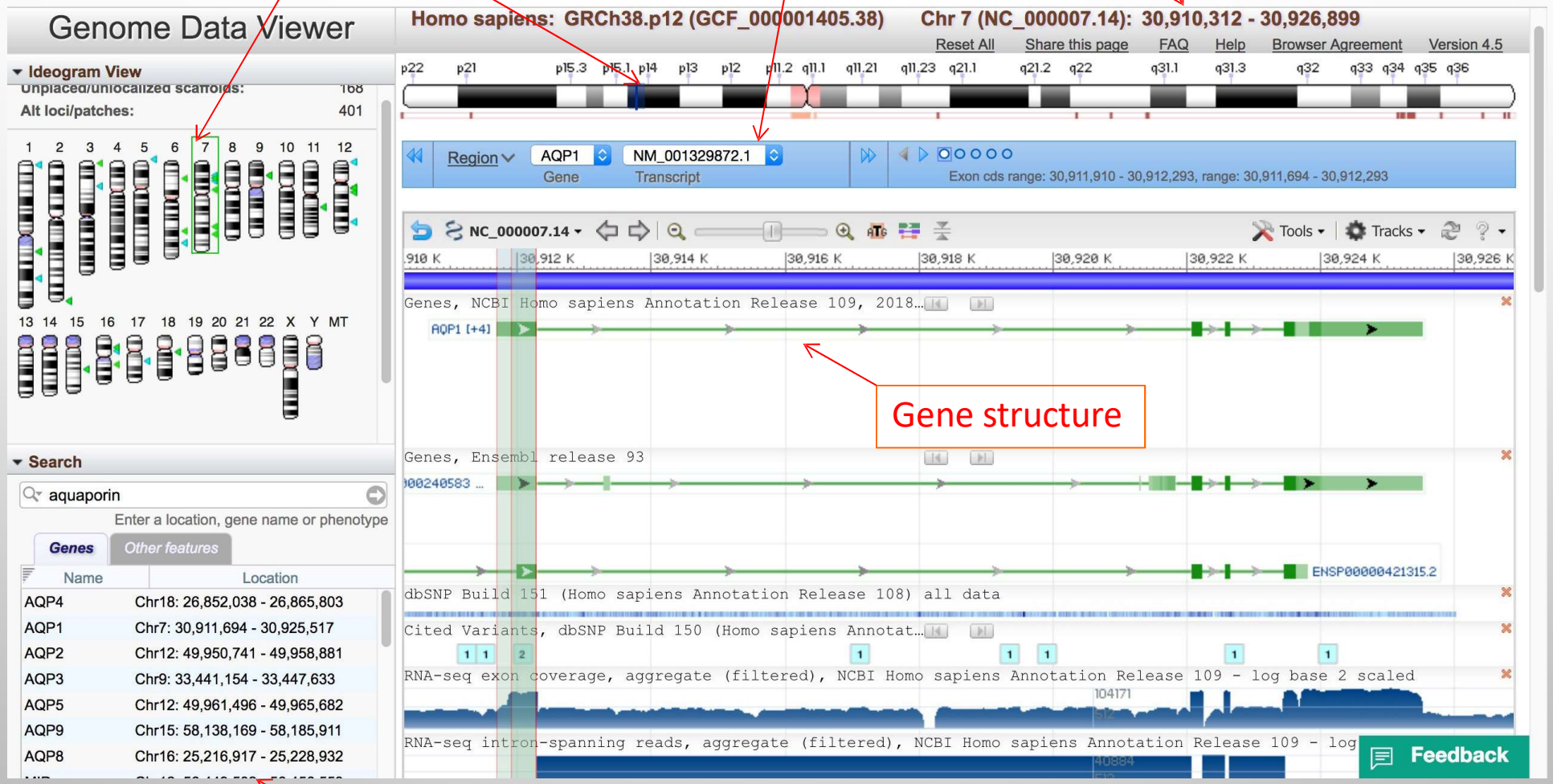
result

Chromosome 7

location: 7p14.3
(cytological map)

Transcript ID

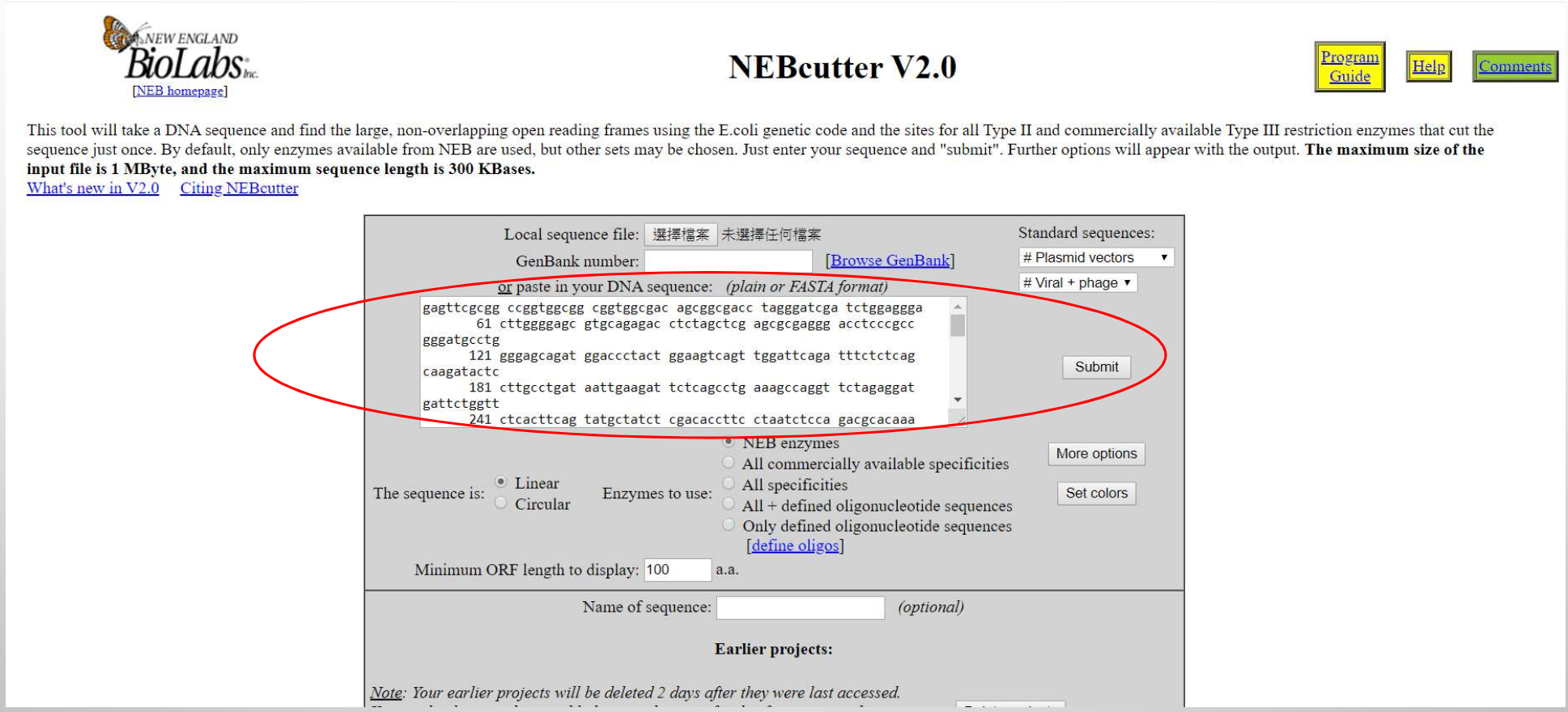
Current position
(physical map)

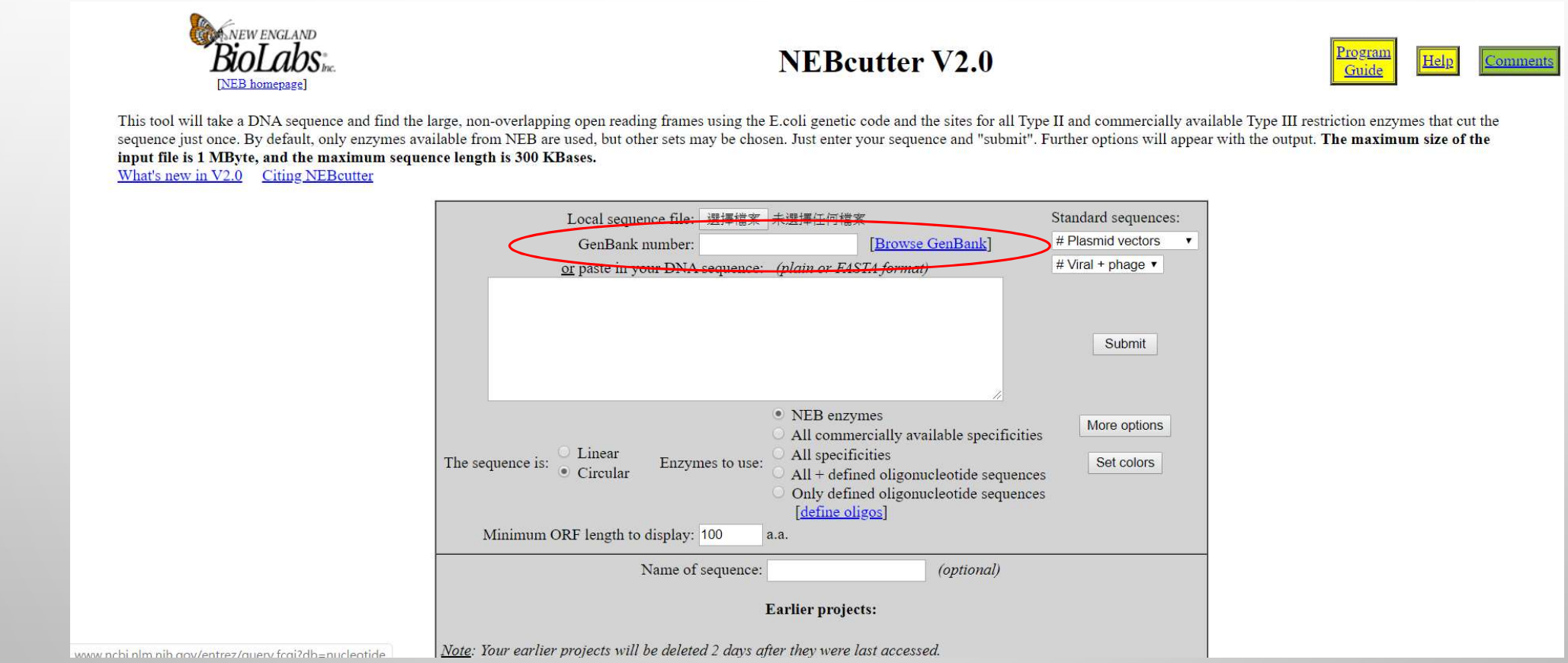


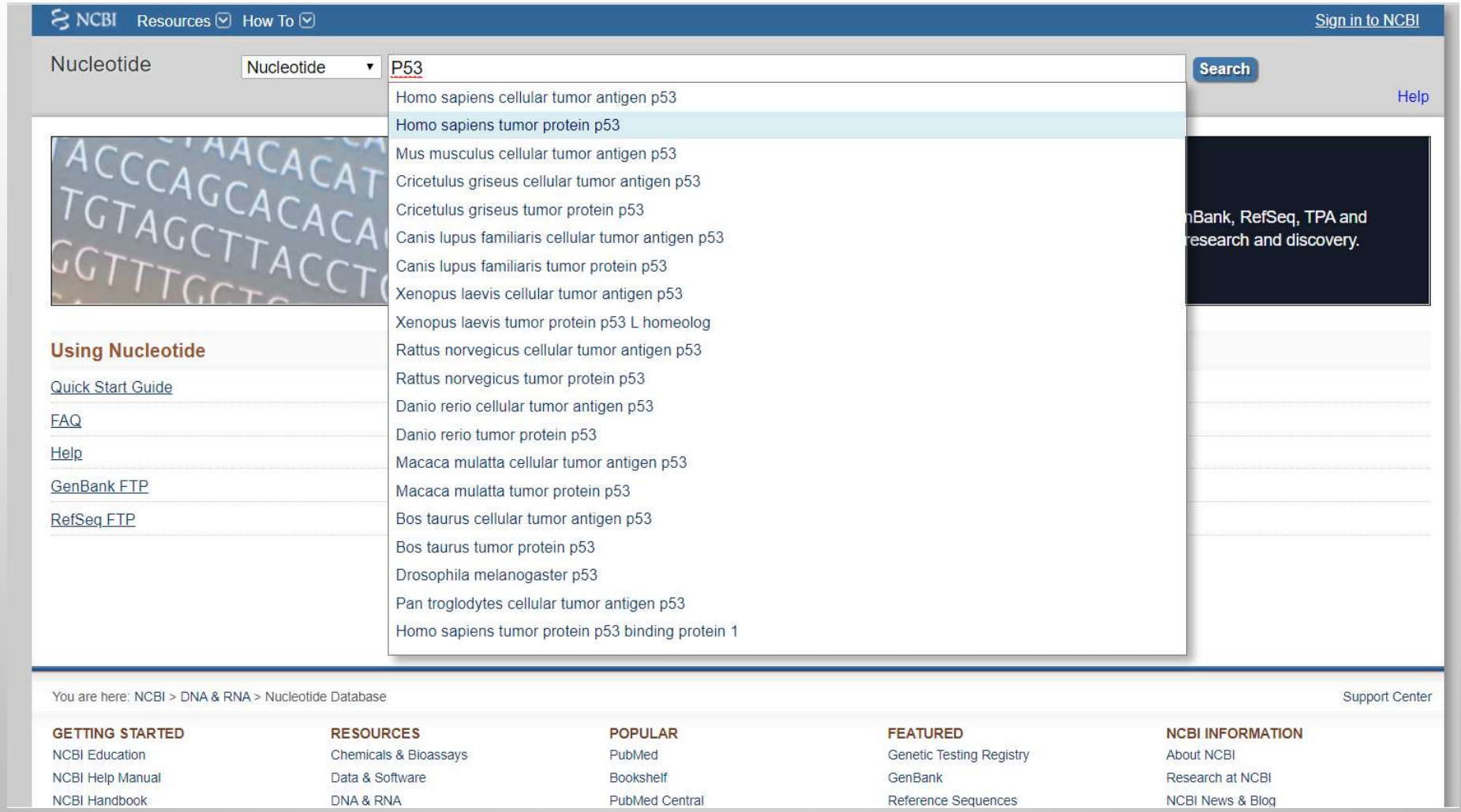
Search result

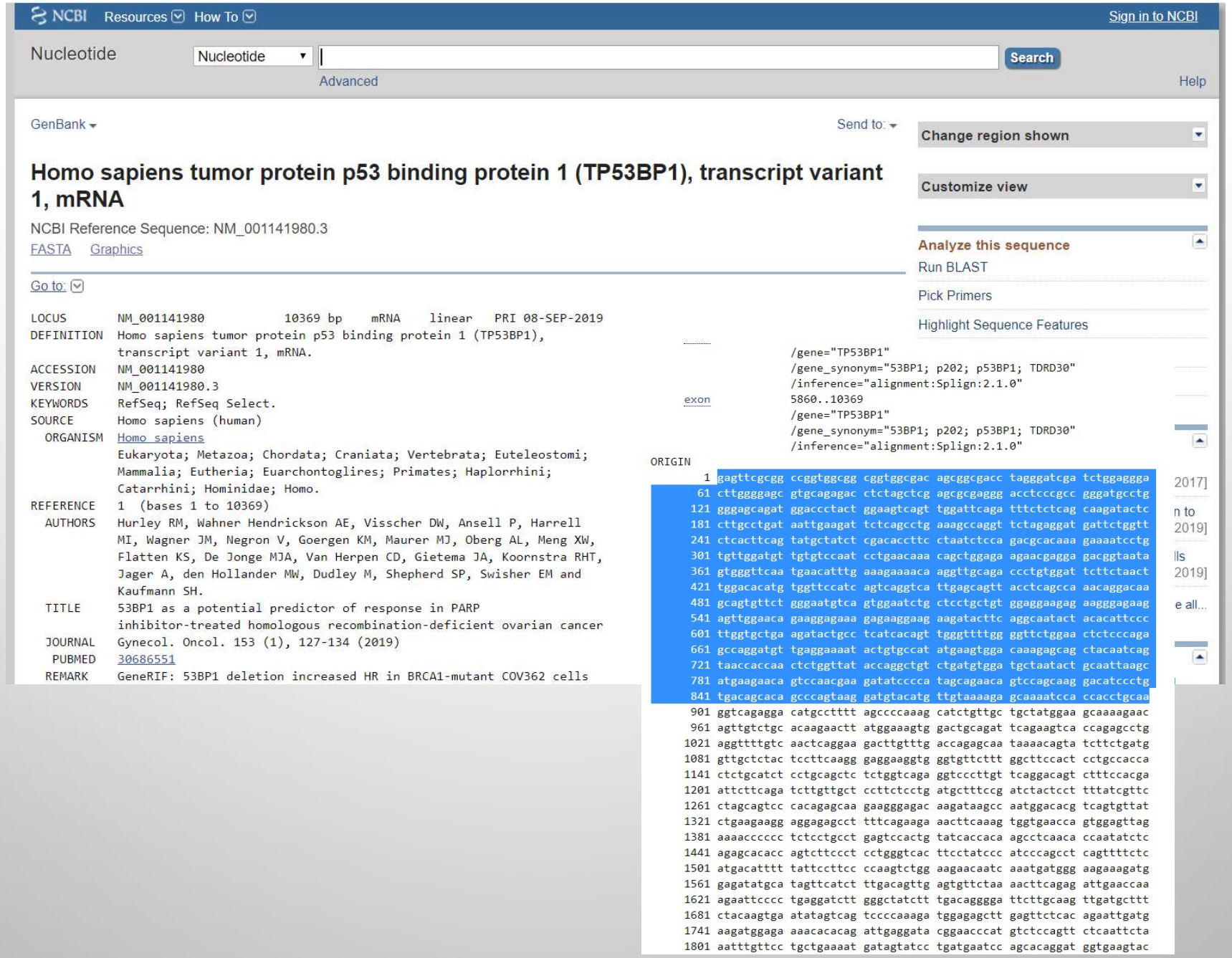
(3) Restriction Enzyme Tool-NEBcutter

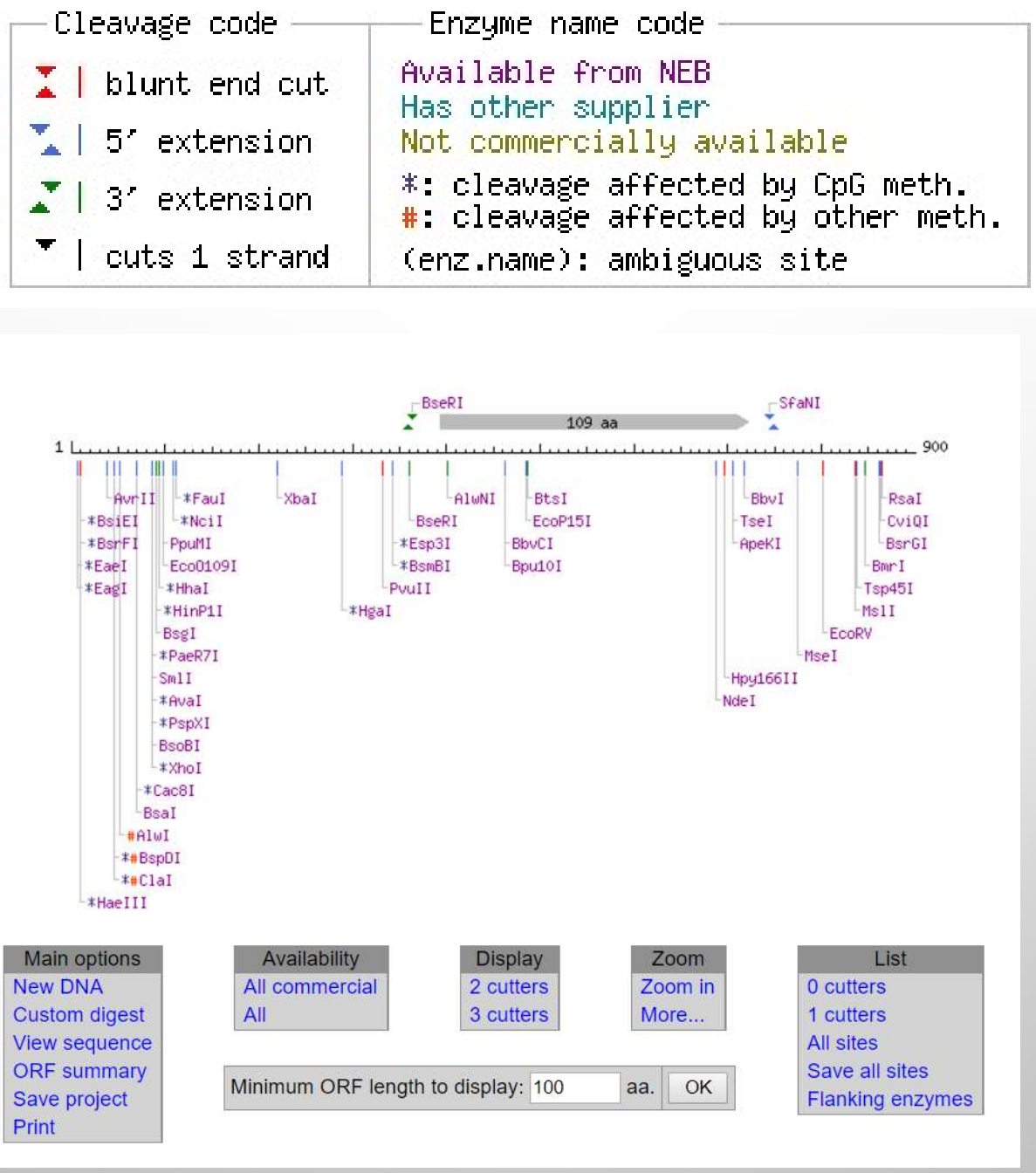
http://nc2.neb.com/NEBcutter2/



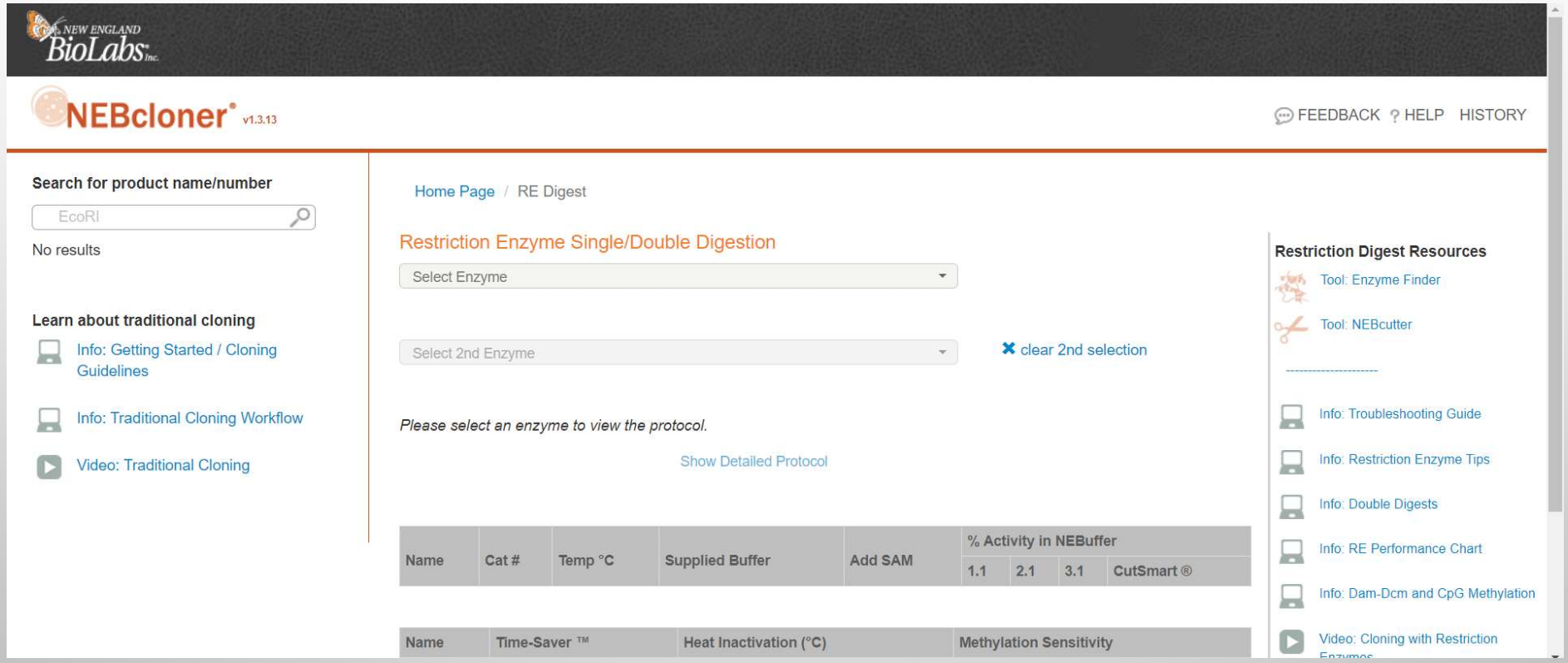

NEW ENGLAND BioLabs Inc.
[NEB homepage]
NEBcutter V2.0
Program Guide
Help
Comments
This tool will take a DNA sequence and find the large, non-overlapping open reading frames using the E.coli genetic code and the sites for all Type II and commercially available Type III restriction enzymes that cut the sequence just once. By default, only enzymes available from NEB are used, but other sets may be chosen. Just enter your sequence and "submit". Further options will appear with the output. The maximum size of the input file is 1 MByte, and the maximum sequence length is 300 KBases.
What's new in V2.0
Citing NEBcutter
Local sequence file: 選擇檔案 未選擇任何檔案
GenBank number: [Browse GenBank]
or paste in your DNA sequence: (plain or FASTA format)
Standard sequences:
Plasmid vectors
Viral + phage
Submit
The sequence is: Linear Circular
Enzymes to use: NEB enzymes; All commercially available specificities; All specificities; All + defined oligonucleotide sequences; Only defined oligonucleotide sequences [define oligos]
More options
Set colors
Minimum ORF length to display: 100 a.a.
Name of sequence: (optional)
Earlier projects:
Note: Your earlier projects will be deleted 2 days after they were last accessed.
www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=nucleotide


NCBI Resources How To
Sign in to NCBI
Nucleotide
Nucleotide
P53
Search
Help
Homo sapiens cellular tumor antigen p53
Homo sapiens tumor protein p53
Mus musculus cellular tumor antigen p53
Cricetulus griseus cellular tumor antigen p53
Cricetulus griseus tumor protein p53
Canis lupus familiaris cellular tumor antigen p53
Canis lupus familiaris tumor protein p53
Xenopus laevis cellular tumor antigen p53
Xenopus laevis tumor protein p53 L homeolog
Rattus norvegicus cellular tumor antigen p53
Rattus norvegicus tumor protein p53
Danio rerio cellular tumor antigen p53
Danio rerio tumor protein p53
Macaca mulatta cellular tumor antigen p53
Macaca mulatta tumor protein p53
Bos taurus cellular tumor antigen p53
Bos taurus tumor protein p53
Drosophila melanogaster p53
Pan troglodytes cellular tumor antigen p53
Homo sapiens tumor protein p53 binding protein 1
Using Nucleotide
Quick Start Guide
FAQ
Help
GenBank FTP
RefSeq FTP
You are here: NCBI > DNA & RNA > Nucleotide Database
Support Center
GETTING STARTED
NCBI Education
NCBI Help Manual
NCBI Handbook
RESOURCES
Chemicals & Bioassays
Data & Software
DNA & RNA
POPULAR
PubMed
Bookshelf
PubMed Central
FEATURED
Genetic Testing Registry
GenBank
Reference Sequences
NCBI INFORMATION
About NCBI
Research at NCBI
NCBI News & Blog

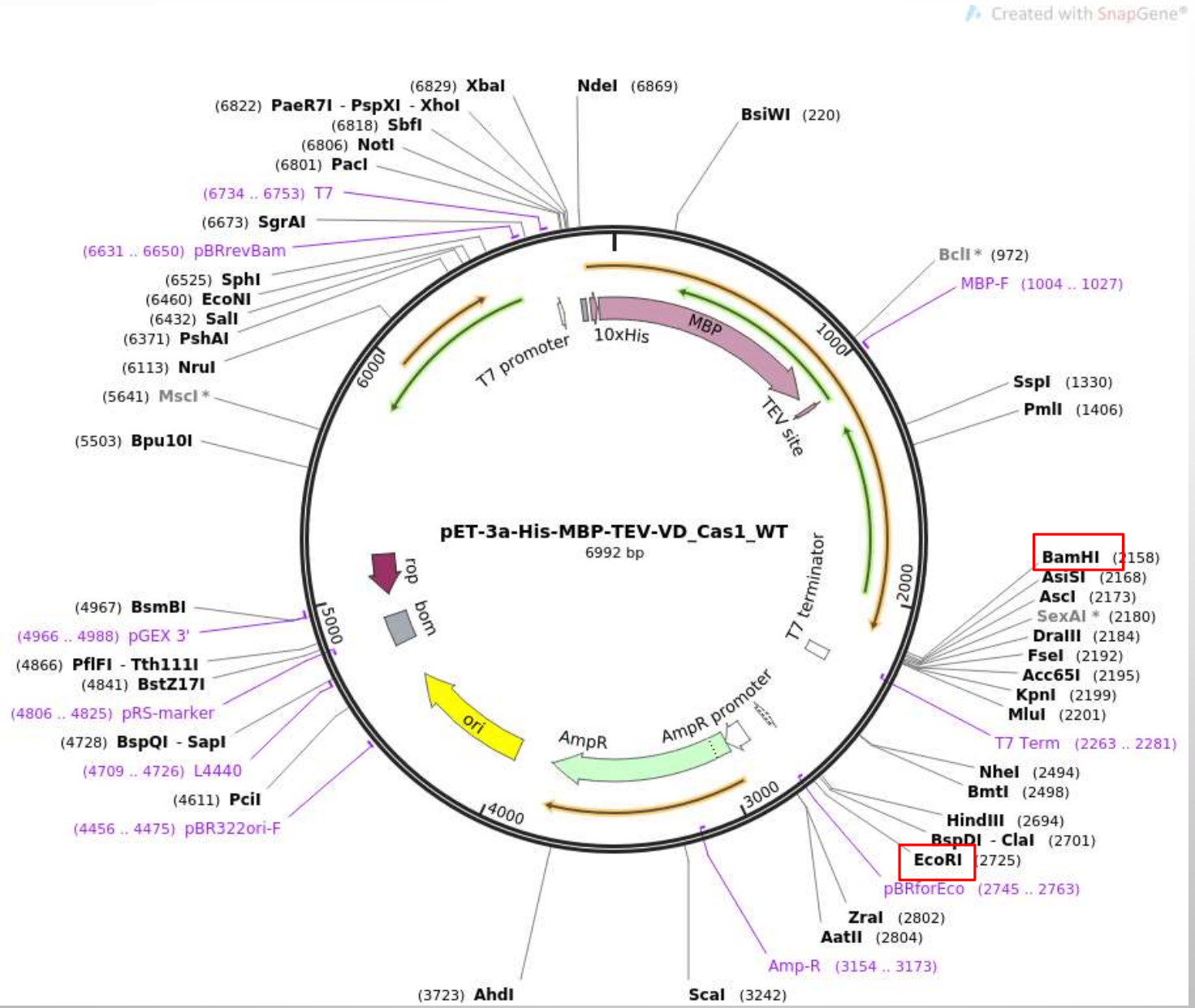

NCBI Resources How To
Sign in to NCBI
Nucleotide
Nucleotide
Search
Advanced
Help
GenBank
Send to:
Change region shown
Customize view
Homo sapiens tumor protein p53 binding protein 1 (TP53BP1), transcript variant 1, mRNA
NCBI Reference Sequence: NM_001141980.3
FASTA Graphics
Go to:
LOCUS NM_001141980 10369 bp mRNA linear PRI 08-SEP-2019
DEFINITION Homo sapiens tumor protein p53 binding protein 1 (TP53BP1), transcript variant 1, mRNA.
ACCESSION NM_001141980
VERSION NM_001141980.3
KEYWORDS RefSeq; RefSeq Select.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 10369)
AUTHORS Hurley RM, Wahner Hendrickson AE, Visscher DW, Ansell P, Harrell MI, Wagner JM, Negron V, Goergen KM, Maurer MJ, Oberg AL, Meng XW, Flatten KS, De Jonge MJA, Van Herpen CD, Gietema JA, Koornstra RHT, Jager A, den Hollander MW, Dudley M, Shepherd SP, Swisher EM and Kaufmann SH.
TITLE 53BP1 as a potential predictor of response in PARP inhibitor-treated homologous recombination-deficient ovarian cancer
JOURNAL Gynecol. Oncol. 153 (1), 127-134 (2019)
PUBMED 30686551
REMARK GeneRIF: 53BP1 deletion increased HR in BRCA1-mutant COV362 cells
Analyze this sequence
Run BLAST
Pick Primers
Highlight Sequence Features
/gene="TP53BP1"
/gene_synonym="53BP1; p202; p53BP1; TDRD30"
/inference="alignment:Splign:2.1.0"
exon 5860..10369
/gene="TP53BP1"
/gene_synonym="53BP1; p202; p53BP1; TDRD30"
/inference="alignment:Splign:2.1.0"
ORIGIN
1 gagttcgcgg ccggtggcgg cggtggcgac agcggcgacc tagggatcga tctggaggga
61 cttggggagc gtgcagagac ctctagctcg agcgcgaggg acctcccgcc gggatgcctg
121 gggagcagat ggaccctact ggaagtcagt tggattcaga tttctctcag caagatactc
181 cttgcctgat aattgaagat tctcagcctg aaagccaggt tctagaggat gattctggtt
241 ctcacttcag tatgctatct cgacaccttc ctaatctcca gacgcacaaa gaaaatcctg
301 tgttggatgt tgtgtccaat cctgaacaaa cagctggaga agaacgagga gacggtaata
361 gtgggttcaa tgaacatttg aaagaaaaca aggttgcaga ccctgtggat tcttctaact
421 tggacacatg tggttccatc agtcaggtca ttgagcagtt acctcagcca aacaggacaa
481 gcagtgttct gggaatgtca gtggaatctg ctcctgctgt ggaggaagag aagggagaag
541 agttggaaca gaaggagaaa gagaaggaag aagatacttc aggcaatact acacattccc
601 ttggtgctga agatactgcc tcatcacagt tgggttttgg ggttctggaa ctctcccaga
661 gccaggatgt tgaggaaaat actgtgccat atgaagtgga caaagagcag ctacaatcag
721 taaccaccaa ctctggttat accaggctgt ctgatgtgga tgctaatact gcaattaagc
781 atgaagaaca gtccaacgaa gatatcccca tagcagaaca gtccagcaag gacatccctg
841 tgacagcaca gcccagtaag gatgtacatg ttgtaaaaga gcaaaatcca ccacctgcaa
901 ggtcagagga catgcctttt agccccaaag catctgttgc tgctatggaa gcaaaagaac
961 agttgtctgc acaagaactt atggaaagtg gactgcagat tcagaagtca ccagagcctg
1021 aggttttgtc aactcaggaa gacttgtttg accagagcaa taaaacagta tcttctgatg
1081 gttgctctac tccttcaagg gaggaaggtg ggtgttcttt ggcttccact cctgccacca
1141 ctctgcatct cctgcagctc tctggtcaga ggtcccttgt tcaggacagt ctttccacga
1201 attcttcaga tcttgttgct ccttctcctg atgctttccg atctactcct tttatcgttc
1261 ctagcagtcc cacagagcaa gaagggagac aagataagcc aatggacacg tcagtgttat
1321 ctgaagaagg aggagagcct tttcagaaga aacttcaaag tggtgaacca gtggagttag
1381 aaaacccccc tctcctgcct gagtccactg tatcaccaca agcctcaaca ccaatatctc
1441 agagcacacc agtcttccct cctgggtcac ttcctatccc atcccagcct cagttttctc
1501 atgacatttt tattccttcc ccaagtctgg aagaacaatc aaatgatggg aagaaagatg
1561 gagatatgca tagttcatct ttgacagttg agtgttctaa aacttcagag attgaaccaa
1621 agaattcccc tgaggatctt gggctatctt tgacagggga ttcttgcaag ttgatgcttt
1681 ctacaagtga atatagtcag tccccaaaga tggagagctt gagttctcac agaattgatg
1741 aagatggaga aaacacacag attgaggata cggaacccat gtctccagtt ctcaattcta
1801 aatttgttcc tgctgaaaat gatagtatcc tgatgaatcc agcacaggat ggtgaagtac


Cleavage code
blunt end cut
5′ extension
3′ extension
cuts 1 strand
Enzyme name code
Available from NEB
Has other supplier
Not commercially available
*: cleavage affected by CpG meth.
#: cleavage affected by other meth.
(enz.name): ambiguous site
BseRI
109 aa
SfaNI
1
900
AvrII
*BsiEI
*BsrFI
*EaeI
*EagI
*FauI
*NciI
PpuMI
Eco0109I
*HhaI
*HinP1I
BsgI
*PaeR7I
SmlI
*AvaI
*PspXI
BsoBI
*XhoI
*Cac8I
BsaI
#AlwI
*#BspDI
*#ClaI
*HaeIII
XbaI
*HgaI
AlwNI
BseRI
*Esp3I
*BsmBI
PvuII
BtsI
EcoP15I
BbvCI
Bpu10I
BbvI
TseI
ApeKI
Hpy166II
NdeI
EcoRV
MseI
RsaI
CviQI
BsrGI
BmrI
Tsp45I
MslI
Main options
New DNA
Custom digest
View sequence
ORF summary
Save project
Print
Availability
All commercial
All
Display
2 cutters
3 cutters
Zoom
Zoom in
More...
List
0 cutters
1 cutters
All sites
Save all sites
Flanking enzymes
Minimum ORF length to display: 100 aa. OK

Double digest finder



PET3A PLASMID



Two Fragments:

4128 bp

512 bp

NEW ENGLAND BioLabs Inc.

NEBcloner® v1.3.13

Search for product name/number

EcoRI

No results

Learn about traditional cloning

Info: Getting Started / Cloning Guidelines

Info: Traditional Cloning Workflow

Video: Traditional Cloning

Home Page / RE Digest

Restriction Enzyme Single/Double Digestion

EcoRI

BamHI ✖ clear 2nd selection

Digest in NEBuffer 3.1 Show Detailed Protocol

Name	Cat #	Temp °C	Supplied Buffer	Add SAM	% Activity in NEBuffer			
					1.1	2.1	3.1	CutSmart ®
EcoRI	R0101	37	NEBuffer EcoRI	No	25	100*	50	50*
BamHI	R0136	37	NEBuffer 3.1	No	75*	100*	100	100*

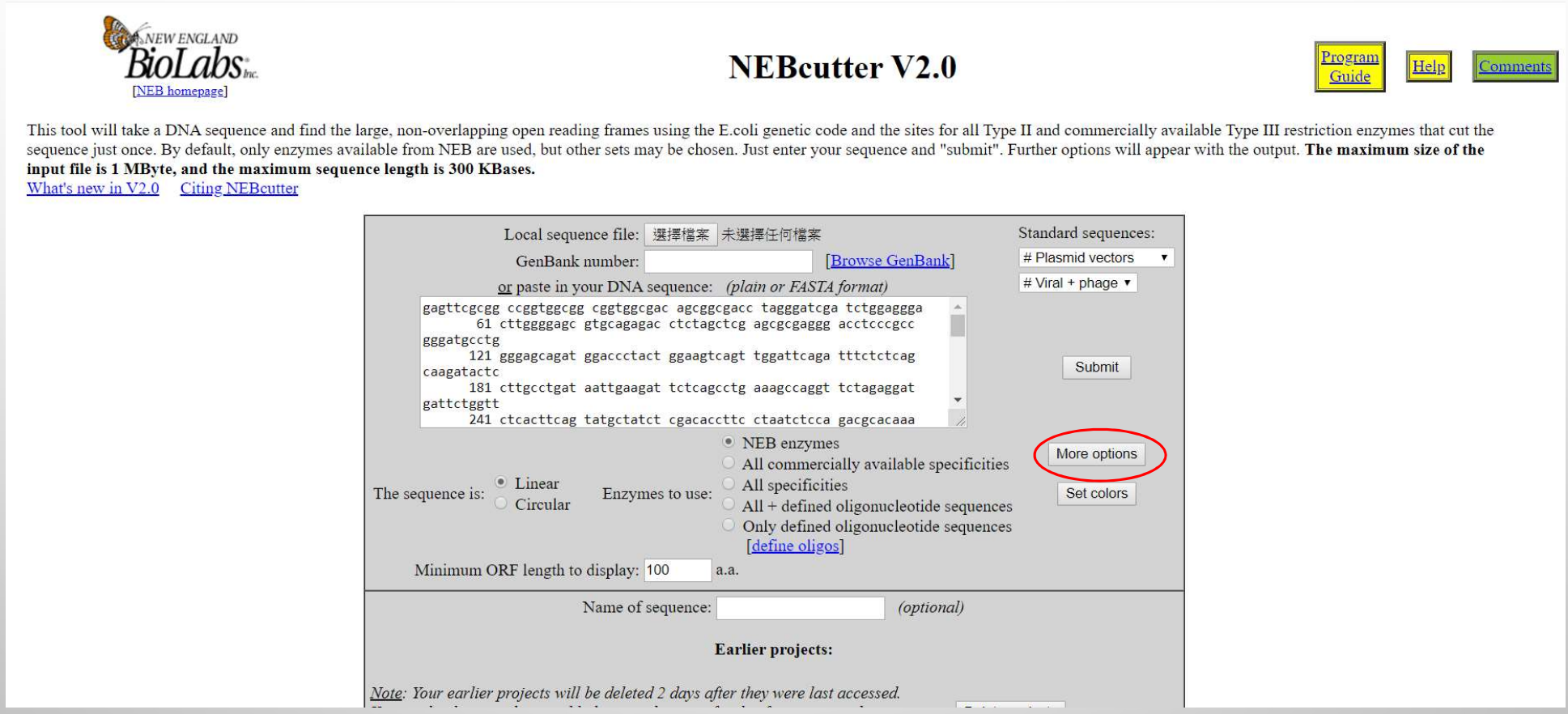
* May exhibit star activity in this buffer.

Name	Time-Saver ™	Heat Inactivation (°C)	Methylation Sensitivity
EcoRI	Yes	65	cpg (Blocked by Some Combinations of Overlapping)
BamHI	Yes	No	None

Notes:

1. Digest in NEBuffer 3.1 (or NEBuffer 3 + BSA) at 37 °C.
2. At least one enzyme has < 100% activity in this buffer, so additional units of enzyme and/or longer incubation time may be necessary.
3. EcoRI has a High Fidelity version EcoRI-HF. BamHI has a High Fidelity version BamHI-HF. High Fidelity (HF) Restriction Enzymes have been engineered for reduced star activity and have 100% activity in CutSmart Buffer which may simplify your double digest.

MORE OPTIONS



A nicking enzyme (or **nicking endonuclease**) is an enzyme that cuts one strand of a double-stranded DNA at a specific recognition nucleotide sequences

In addition to Type II and commercially available Type III enzymes, also look for:
- Type I & III enzymes
- Homing endonucleases
- Nicking enzymes

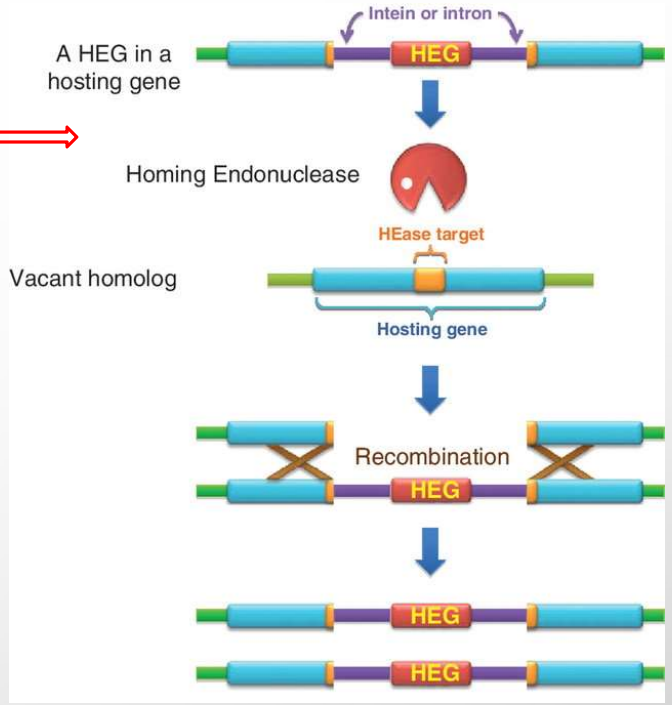
- Ignore CpG methylation
- Ignore Dam methylation
- Ignore Dcm methylation
- Ignore EcoBI methylation
- Ignore EcoKI methylation

Genetic code to use when searching for ORFs:
11. Bacterial and Plant Plastid

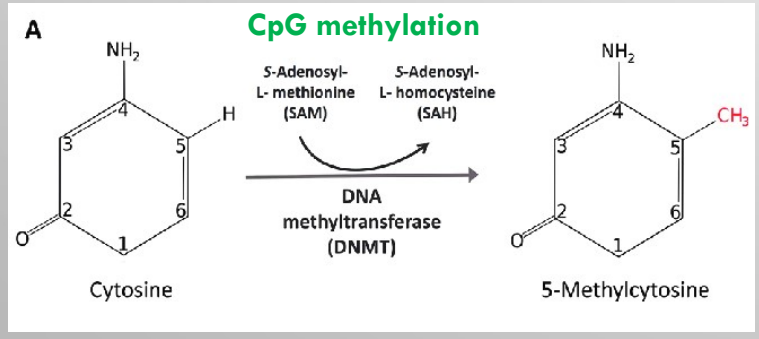
Sequence is a fragment

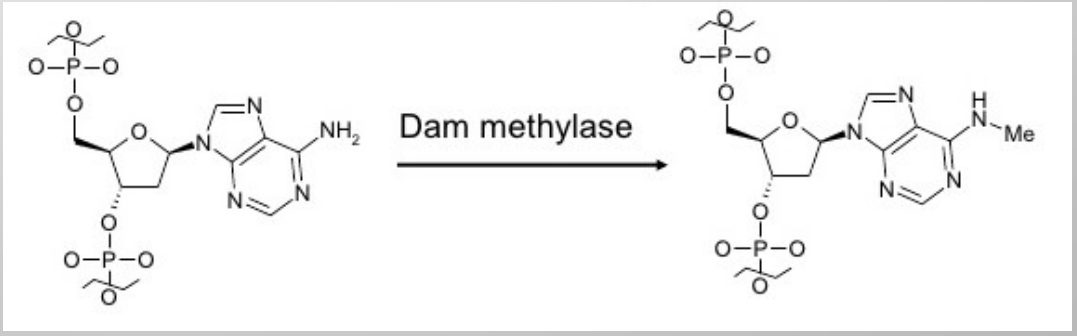
Process this region only: - bp

OK Cancel

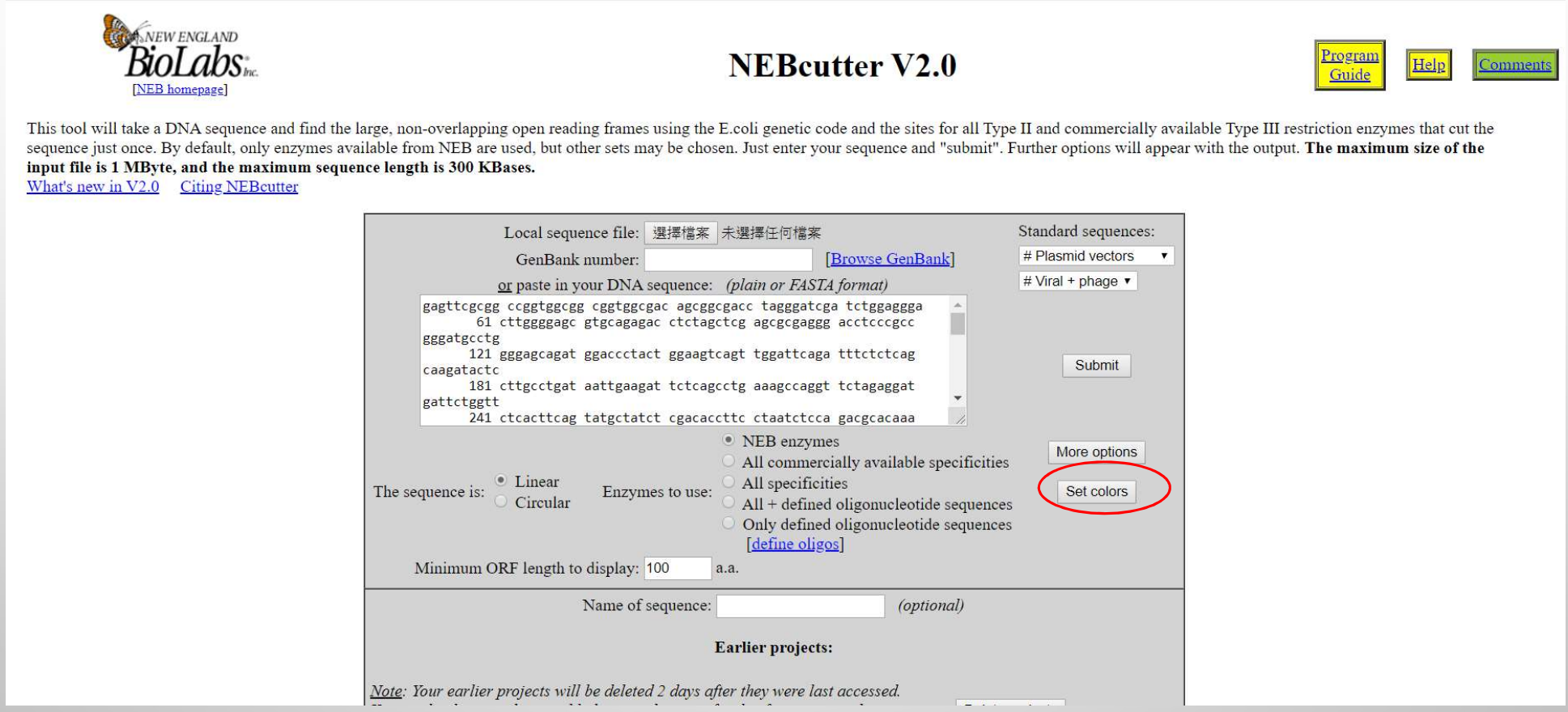


EcoKI or EcoBI methyltransferases modify adenine residues within their respective recognition sequences:
AAC(N)6GTGC for EcoKI and TGA(N)8TGCT for EcoBI





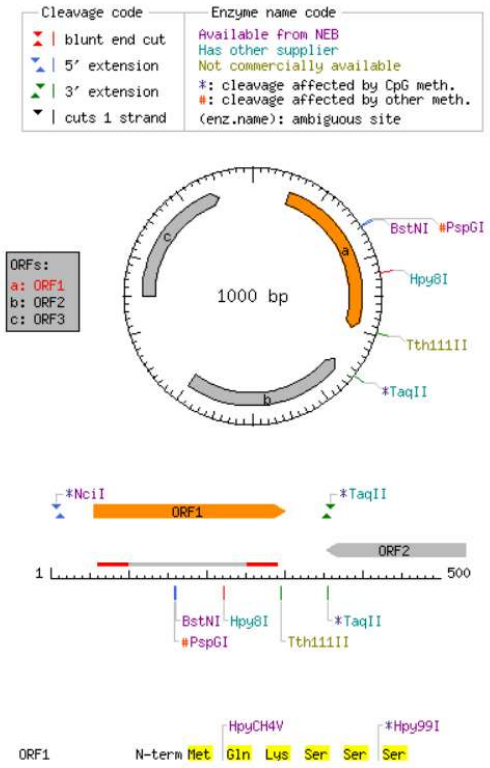
Set Colors


NEW ENGLAND BioLabs Inc.
[NEB homepage]
NEBcutter V2.0
Program Guide
Help
Comments
This tool will take a DNA sequence and find the large, non-overlapping open reading frames using the E.coli genetic code and the sites for all Type II and commercially available Type III restriction enzymes that cut the sequence just once. By default, only enzymes available from NEB are used, but other sets may be chosen. Just enter your sequence and "submit". Further options will appear with the output. The maximum size of the input file is 1 MByte, and the maximum sequence length is 300 KBases.
What's new in V2.0
Citing NEBcutter
Local sequence file: 選擇檔案 未選擇任何檔案
GenBank number: [Browse GenBank]
or paste in your DNA sequence: (plain or FASTA format)
gagttcgcgg ccggtggcgg cggtggcgac agcggcgacc tagggatcga tctggaggga
61 cttggggagc gtgcagagac ctctagctcg agcgcgaggg acctcccgcc
gggatgcctg
121 gggagcagat ggaccctact ggaagtcagt tggattcaga tttctctcag
caagatactc
181 cttgcctgat aattgaagat tctcagcctg aaagccaggt tctagaggat
gattctggtt
241 ctcacttcag tatgctatct cgacaccttc ctaatctcca gacgcacaaa
Standard sequences:
Plasmid vectors
Viral + phage
Submit
More options
Set colors
The sequence is: Linear Circular
Enzymes to use:
NEB enzymes
All commercially available specificities
All specificities
All + defined oligonucleotide sequences
Only defined oligonucleotide sequences
[define oligos]
Minimum ORF length to display: 100 a.a.
Name of sequence: (optional)
Earlier projects:
Note: Your earlier projects will be deleted 2 days after they were last accessed.

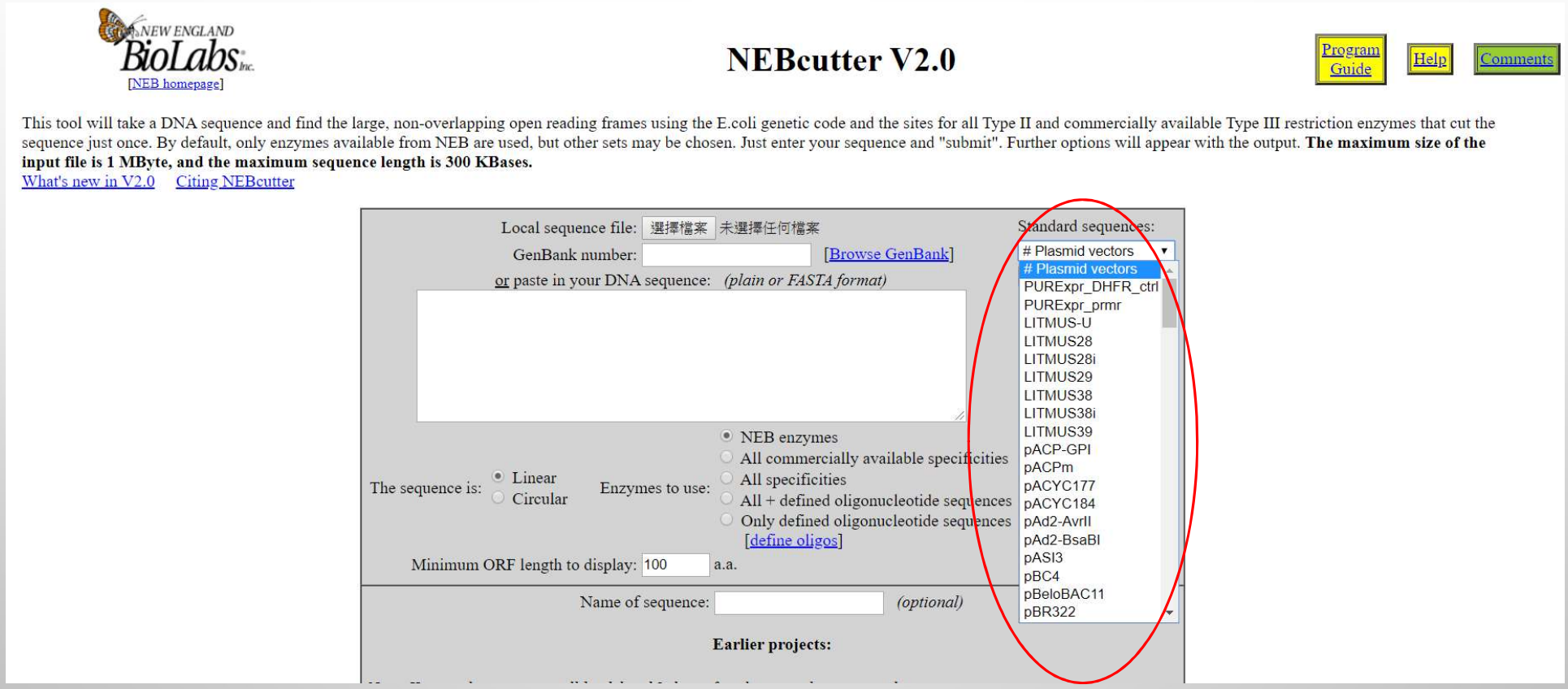
NEW ENGLAND BioLabs Inc.
NEBcutter

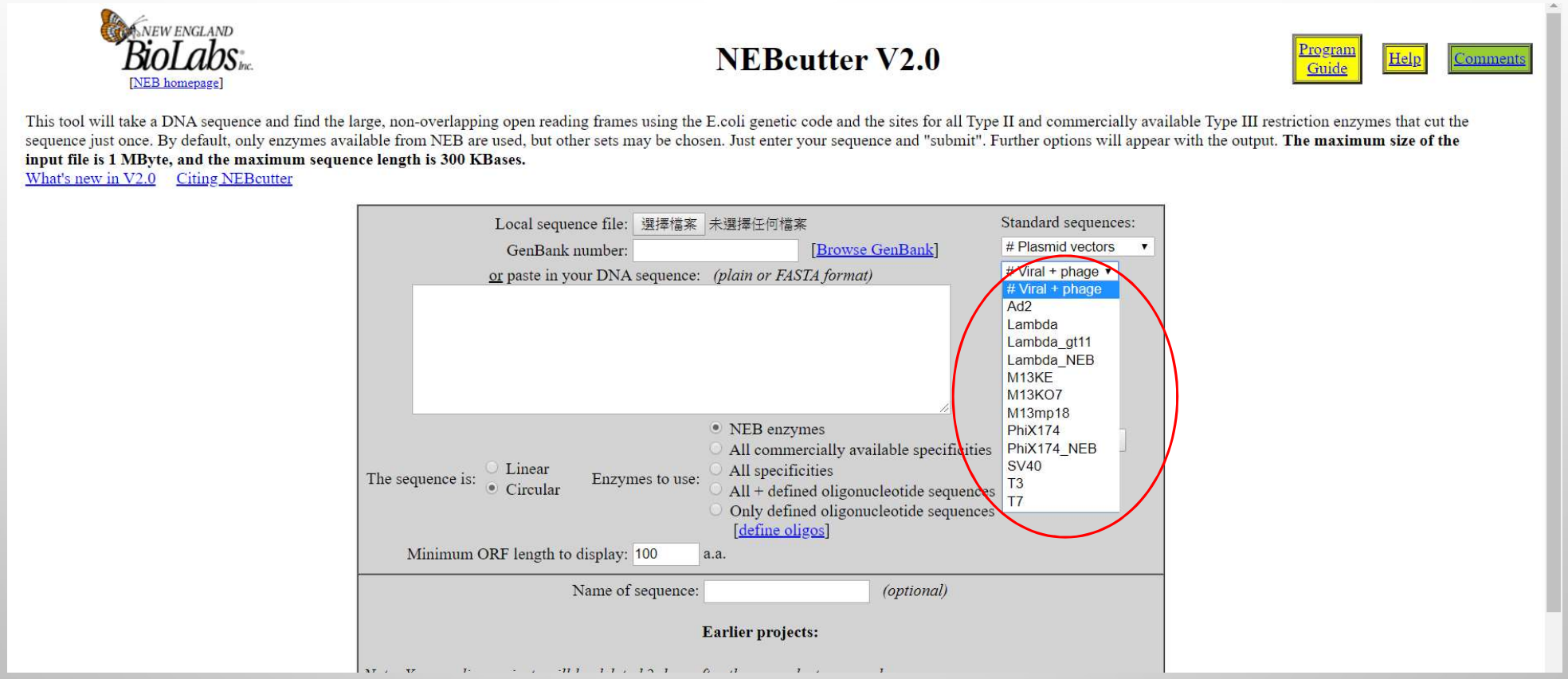
Set Colors

Help | Comments



Element	Color	Change
Scale	black	Change
Cut site - blunt	red	Change
Cut site - 5' ext.	royalblue	Change
Cut site - 3' ext.	green	Change
Cut site - one strand	black	Change
Cut site - unknown	black	Change
Cut site - multiple overlapping	darkmagenta	Change
Guide line	silver	Change
Supplier code - NEB	darkmagenta	Change
Supplier code - other commercial	darkcyan	Change
Supplier code - none	olive	Change
MS - CpG	darkslateblue	Change
MS - other	orangered	Change
Basepair hilite - unique	darkred	Change
Basepair hilite - degenerate	darkblue	Change
a.a. name	black	Change
a.a. background	yellow	Change
ORF body	mediumgray	Change
ORF outline (linear)	mediumgray	Change
ORF outline (circular)	black	Change
ORF name	black	Change
Highlighted ORF body	darkorange	Change
Highlighted ORF outline (linear)	darkorange	Change


NEW ENGLAND BioLabs Inc.
[NEB homepage]
NEBcutter V2.0
Program Guide
Help
Comments
This tool will take a DNA sequence and find the large, non-overlapping open reading frames using the E.coli genetic code and the sites for all Type II and commercially available Type III restriction enzymes that cut the sequence just once. By default, only enzymes available from NEB are used, but other sets may be chosen. Just enter your sequence and "submit". Further options will appear with the output. The maximum size of the input file is 1 MByte, and the maximum sequence length is 300 KBases.
What's new in V2.0
Citing NEBcutter
Local sequence file: 選擇檔案 未選擇任何檔案
GenBank number: [Browse GenBank]
or paste in your DNA sequence: (plain or FASTA format)
Standard sequences:
Plasmid vectors
Plasmid vectors
PURExpr_DHFR_ctrl
PURExpr_prmr
LITMUS-U
LITMUS28
LITMUS28i
LITMUS29
LITMUS38
LITMUS38i
LITMUS39
pACP-GPI
pACPm
pACYC177
pACYC184
pAd2-AvrII
pAd2-BsaBI
pASI3
pBC4
pBeloBAC11
pBR322
The sequence is: Linear Circular
Enzymes to use: NEB enzymes
All commercially available specificities
All specificities
All + defined oligonucleotide sequences
Only defined oligonucleotide sequences
[define oligos]
Minimum ORF length to display: 100 a.a.
Name of sequence: (optional)
Earlier projects:


NEW ENGLAND BioLabs Inc.
[NEB homepage]
NEBcutter V2.0
Program Guide
Help
Comments
This tool will take a DNA sequence and find the large, non-overlapping open reading frames using the E.coli genetic code and the sites for all Type II and commercially available Type III restriction enzymes that cut the sequence just once. By default, only enzymes available from NEB are used, but other sets may be chosen. Just enter your sequence and "submit". Further options will appear with the output. The maximum size of the input file is 1 MByte, and the maximum sequence length is 300 KBases.
What's new in V2.0
Citing NEBcutter
Local sequence file: 選擇檔案 未選擇任何檔案
GenBank number: [Browse GenBank]
or paste in your DNA sequence: (plain or FASTA format)
Standard sequences:
Plasmid vectors
Viral + phage
Viral + phage
Ad2
Lambda
Lambda_gt11
Lambda_NEB
M13KE
M13KO7
M13mp18
PhiX174
PhiX174_NEB
SV40
T3
T7
The sequence is: Linear Circular
Enzymes to use: NEB enzymes
All commercially available specificities
All specificities
All + defined oligonucleotide sequences
Only defined oligonucleotide sequences
[define oligos]
Minimum ORF length to display: 100 a.a.
Name of sequence: (optional)
Earlier projects: