

生物資訊與因果生物學 Bioinformatics & Bio-Causality (BBC)

尋找causal genes以自閉症為例

Trees-Juen Chuang (莊樹諄)

中央研究院 基因體研究中心 研究員/教授
Research fellow/Professor, Genomics Research Center, Academia Sinica

不必痛的領悟
學因果



Email: trees@gate.sinica.edu.tw

<https://www.genomics.sinica.edu.tw/tw/chuang-trees-juen->



December 28, 2024

生物資訊與因果生物學 Bioinformatics & Bio-Causality (BBC)

尋找致病基因(causal genes)以自閉症為例

1. 分子演化簡介
2. 基因體(DNA)、轉錄體(RNA)、與蛋白質體(protein)
3. DNA和疾病間的關係
4. RNA和疾病間的關係
5. DNA和RNA間的關係
6. DNA、RNA和疾病間的關係

Bioinformatics & Bio-Causality (BBC) Lab.

我們致力於探討人類複雜疾病，基因型(genotype – DNA level)和表現型(phenotype – molecular level)變化間的因果關係。

- **方向:** 比較與演化基因體/轉錄體、因果生物學、系統生物學、分子生物學
- **主題:** 分子內/分子間選擇性剪接、環狀RNA、RNA編輯、後轉錄調控、因果推論
- **策略:** 資訊演算法、統計模型、演化推演、分生實驗
- **流程:** raw data → information → validation → knowledge → new data
- **疾病:** 神經發育/退化疾病、腦膠質瘤、乳癌

基因、演化及大數據分析



Trees-Juen Chuang (莊樹諄)

<http://idv.sinica.edu.tw/trees/>

Email: trees@gate.sinica.edu.tw



Academia Sinica

Genomics Research Center



猩球崛起 劇情簡介

美國科學家威爾·羅德曼利用黑猩猩進行實驗，研究藥物治療阿茲海默症。猩猩主角凱撒因此智力大增，牠眼見人類對待猩猩的種種不人道，憤而領導猩猩族群，反抗人類世界，希望解救所有受到不人道待遇的猩猩們，並帶領大家重返真正的家—森林。

<http://zh.wikipedia.org/wiki/%E7%8C%BF%E4%BA%BA%E7%88%AD%E9%9C%B8%E6%88%B0%E7%BC%9A%E7%8C%A9%E5%87%B6%E9%9D%A9%E5%91%BD#.E6.95.85.E4.BA.8B.E7.B0.A1.E4.BB.8B>



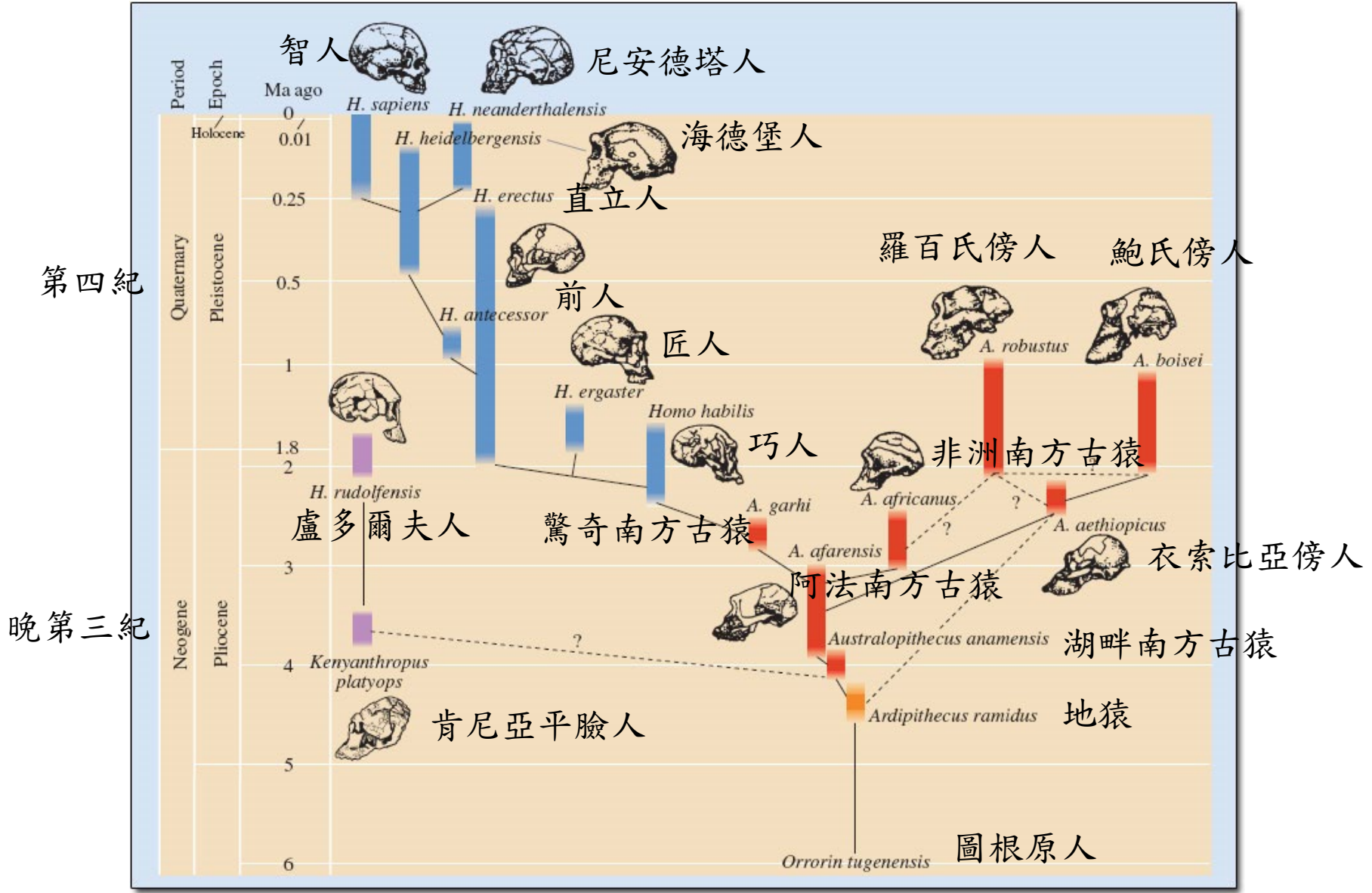
<學院報告> (短篇小說)

法蘭茲 · 卡夫卡

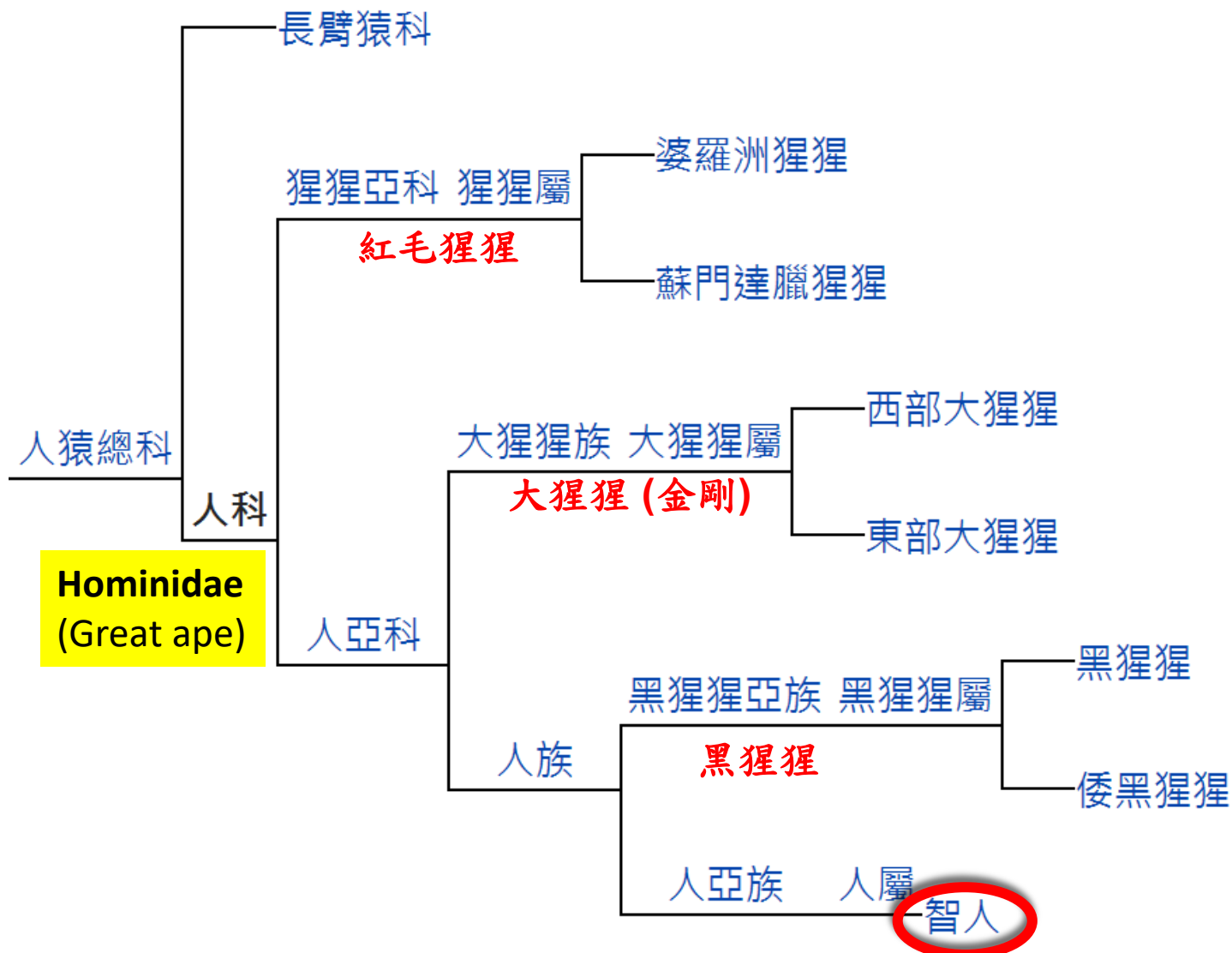
卡夫卡以黑猩猩的觀點來看自由與身為人類的意義。

內容描述一隻猩猩在非洲被人類抓走並關在籠內運往歐洲去，這隻猩猩在籠內便開始思考，他如果要活下，他就必須找尋“出路”。

而他的“出路”就是必須向籠子外面的那些“人”一樣，所以他開始學習“人類”的行為、語言，甚至達到一般歐洲人的教育水準，並受邀到學院裡作一場演講。

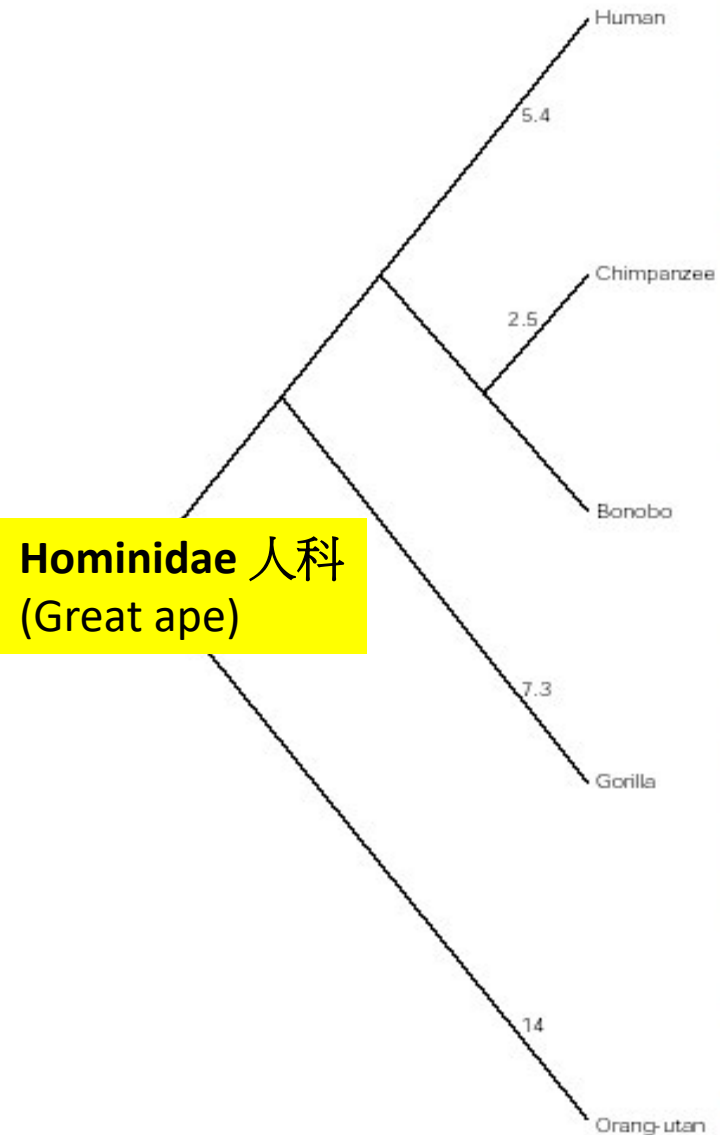


界、門、綱、目、科、屬、種



國際自然保護聯盟分類中，智人是人科物種中唯一無危物種。

單位：百萬年



黑猩猩
(Chimpanzee)

侏儒黑猩猩
(Bonobo)

大猩猩
(Gorilla)

紅毛猩猩 婆羅洲跟蘇門達臘
(Orangutan)

非洲

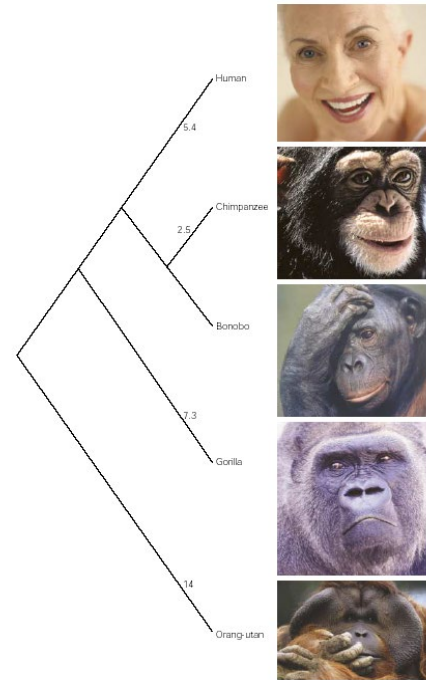
Common ancestor

共同祖先

最近의共同祖先

人跟黑猩猩的基因非常相像
➔ 有人說人是「第三種黑猩猩」。
(The Third Chimpanzee--Jared Diamond)

醫藥上的應用：研究為何有些疾病，在人身上會致命，在黑猩猩身上卻不發病或僅產生輕微病症，如愛滋病、阿茲海默症、B/C型肝炎等。



- 人有23對染色體: 1, 2, ..., 22, X, Y
- 黑猩猩有24對染色體: 1, 2, ..., 23, X, Y
- 人的第2號染色體相當於黑猩猩的第12和13號染色體
- 人和黑猩猩的基因體大小差不多
- 人和黑猩猩在DNA上的相似度超過98% (1.23%的差異)
- 幾乎所有的人的基因和黑猩猩的基因皆具很高的相似度
 - 人 vs. 老鼠 → 88%
 - 人 vs. 雞 → 60%

新世界猴

舊世界猴

小猿

大猿

長臂猿

紅毛猩猩

大猩猩

黑猩猩

侏儒黑猩猩

Lesser apes

Great apes

New World
monkeys

Old World
monkeys

Gibbons

Orang-utan

Gorilla

Common
chimpanzee

Bonobo

Human

Time

界、門、綱、目、科、屬、種

靈長目

Hominidae 人科
(Great ape)

原猴科

(樹鼯、狐猴、眼鏡猴)

猴科

新大陸猴 (蜘蛛猴)

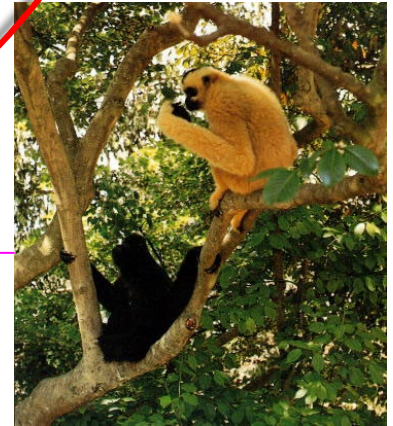
舊大陸猴 (獼猴、狒狒)

猿科

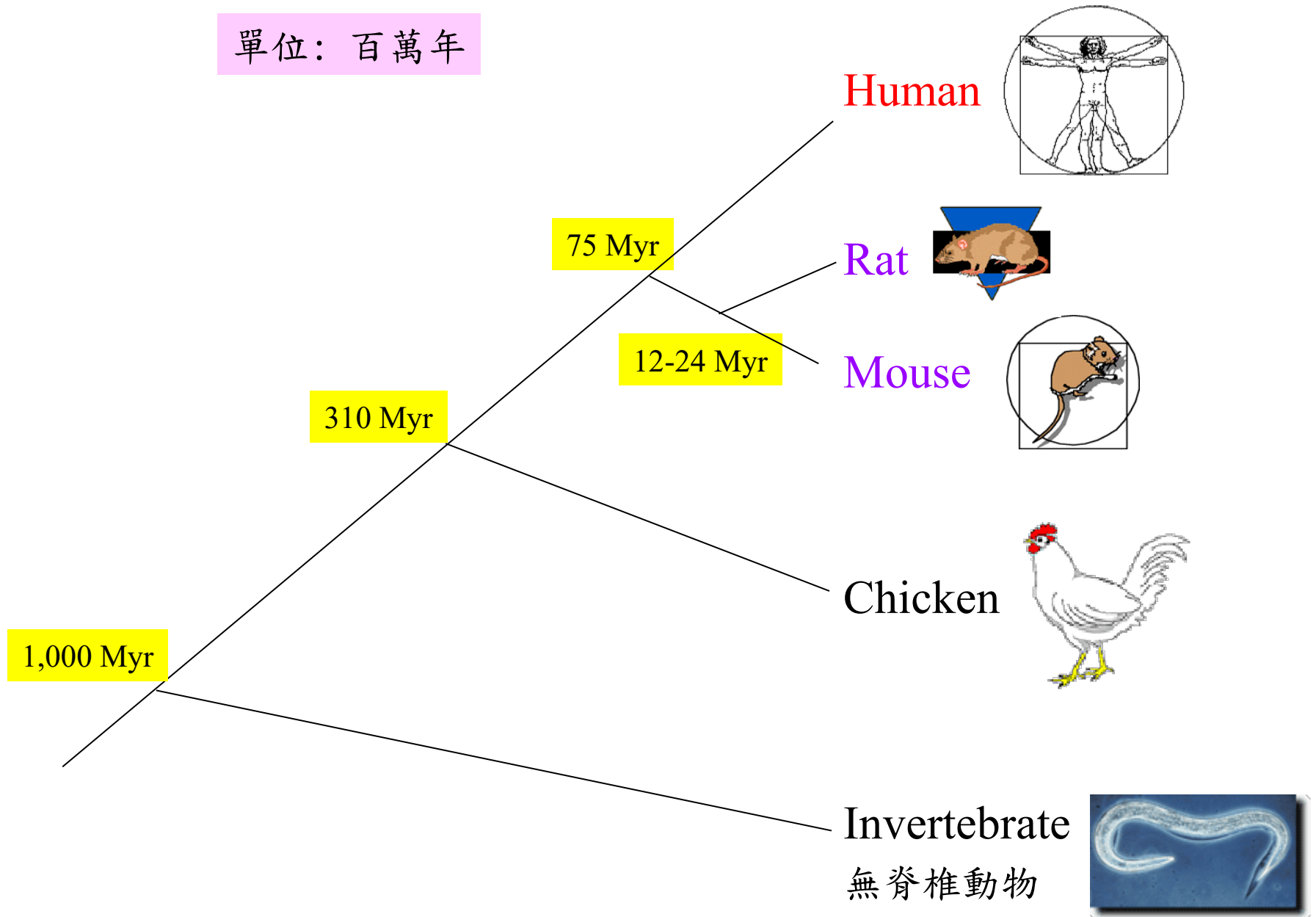
小猿 (長臂猿)

大猿

人科



單位：百萬年



基因體序列 (DNA 序列):

aagtacgatatgccgagtcccatatgcgtagtcatgagc

由四種核苷酸 (**nucleotide acid**) 排列組合而成

A, G, C, T



A (adenine, 腺嘌呤)

G (guanine, 鳥嘌呤)

T (thymine, 胸腺嘧啶)

C (cytosine, 胞嘧啶)

人類的DNA序列由30億個A、G、T、C核苷酸排列組合而成

A (adenine, 腺嘌呤) T (thymine, 胸腺嘧啶)
G (guanine, 鳥嘌呤) C (cytosine, 胞嘧啶)

DNA → nucleotide acid (核苷酸) {
① Phosphoric acid (磷酸)
② Deoxyribose (去氧核糖)
③ Nitrogenous base (含氮鹽基)

Nitrogenous base (含氮鹽基) {
① Purines : {
● Adenine (A, 腺嘌呤)
● Guanine (G, 鳥糞嘌呤)
② Pyrimidine : {
● Cytosine (C, 胞嘧啶)
● Thymine (T, 胸腺嘧啶)

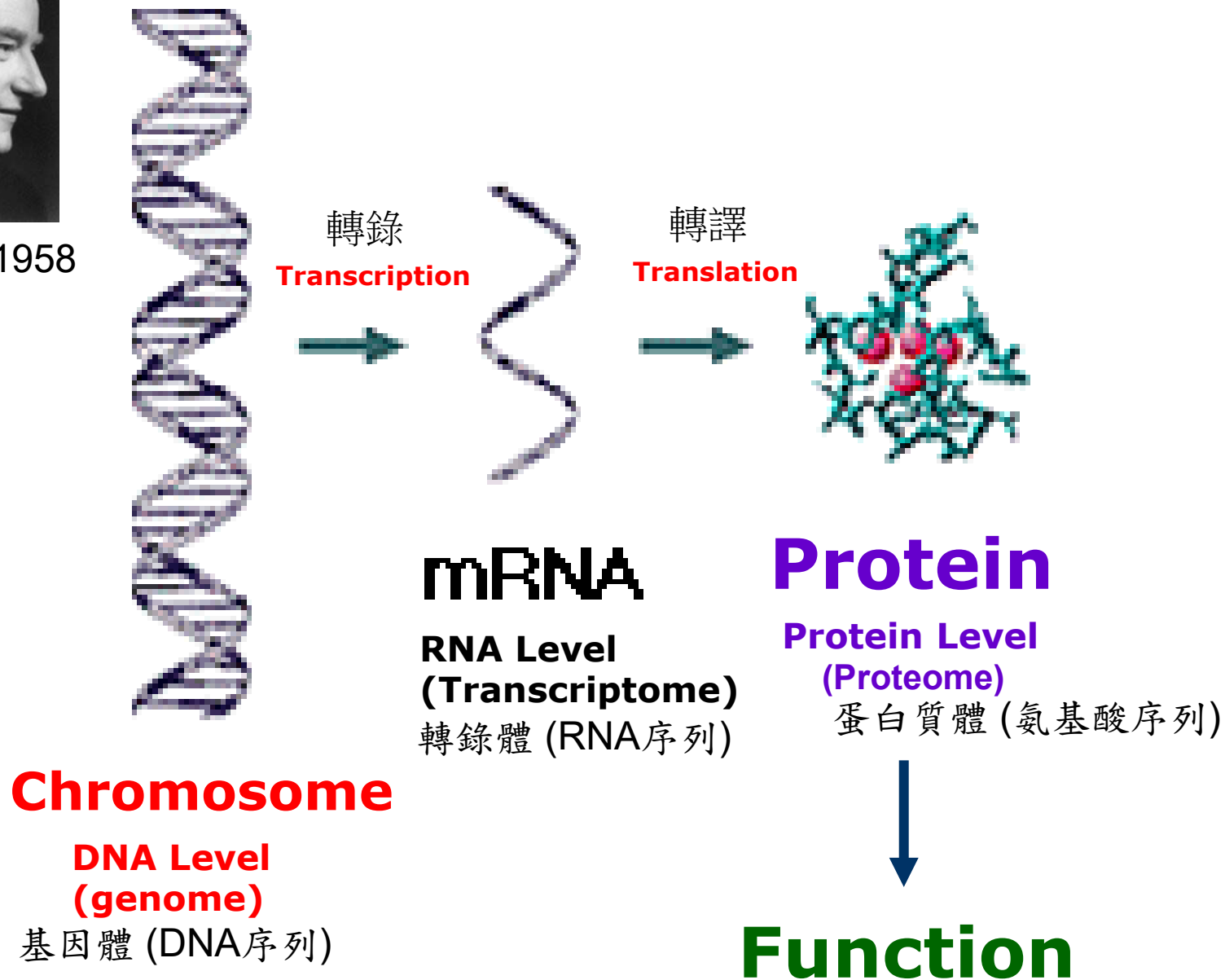
● DNA sequence: **A, C, G, T** --- 4 letters

● RNA sequence: **A, C, G, U** (Uracil, (U), 尿嘧啶) --- 4 letters

分子生物學的中心法則 (central dogma)



克拉克; 1958

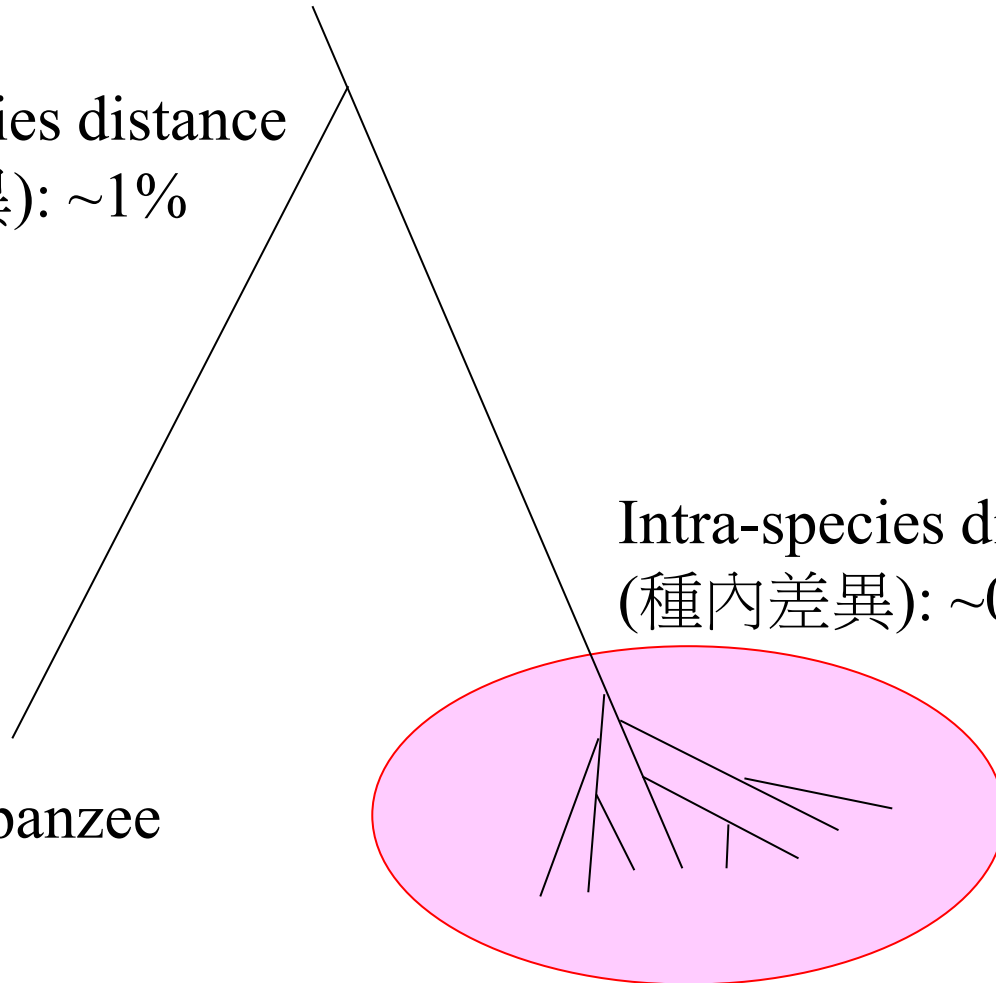


Inter-species distance
(種間差異): ~1%

Chimpanzee

Intra-species distance
(種内差異): ~0.1%

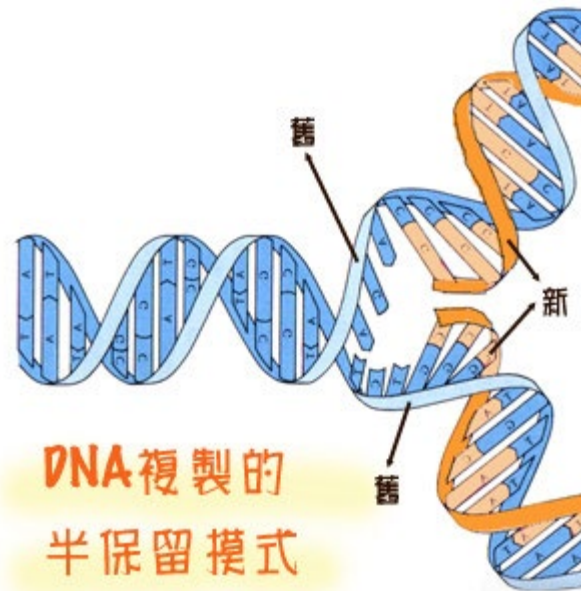
Human



DNA的雙螺旋模型

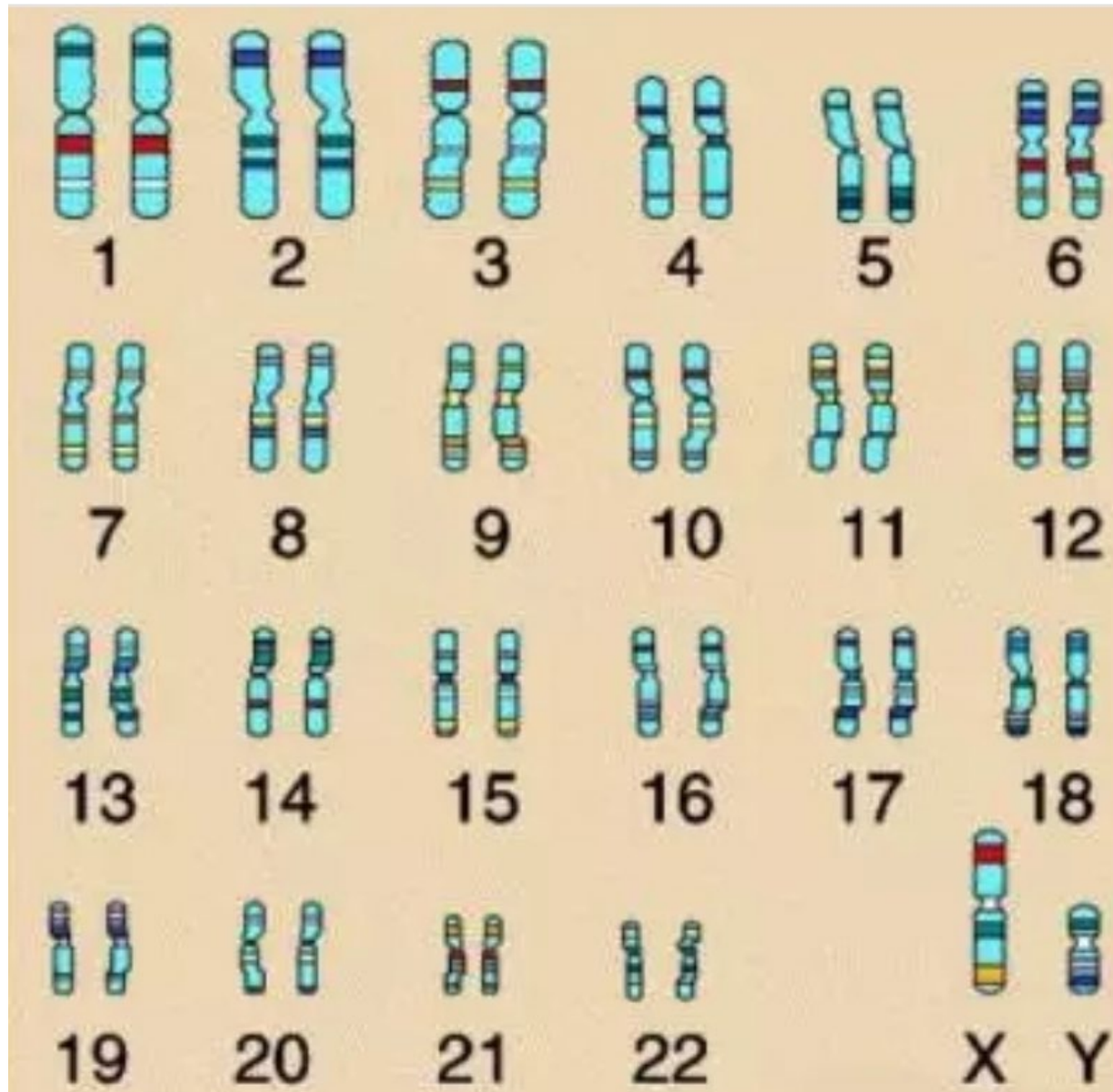


1953年，華生與克拉克發表了DNA的雙螺旋模型，它的結構以及所揭發的生命密碼系統，堪稱上一個世紀生物學最重大的發現。



1955年12月:印尼裔美國籍遺傳學家 **Joe Hin Tjio**, 人有23對染色體 (chromosomes)

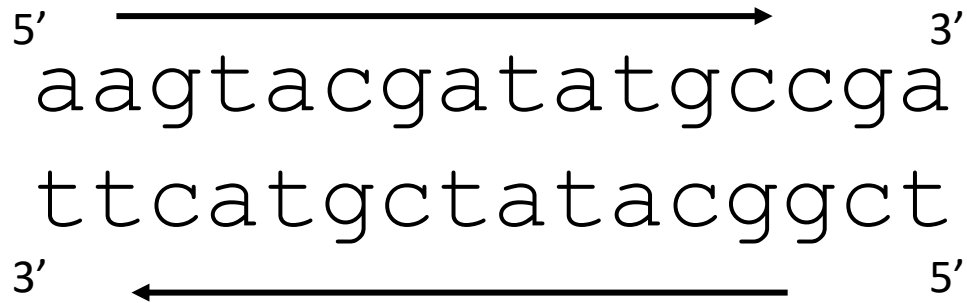
➔ 所有 DNA 序列的總稱叫做: 基因體 (genome)



科學家經過**50**年的努力，在本世紀的初期，也完成了人類基因組的定序計畫，寫出了組成**23**對染色體的**30**億個字母**A、G、C、T**的序列。

人類的DNA序列由**30億 (3×10^9)** 個**A、G、C、T**核苷酸排列組合而成

一對染色體片段序列：



演化的幾個力量

- Mutation (突變)
- Selection (選擇)
- Genetic drift (基因漂變)
- Gene flow (基因流動)

突變是隨機發生的。

選擇則讓有利於生存與繁殖的遺傳性狀在族群間擴大，有害的性狀在族群間漸漸減少甚至消失。

天擇 (natural selection)

- 負向選擇 (negative selection)
- 正向選擇 (positive selection)
- 中性選擇 (neutral selection)
- 平衡選擇 (balancing selection):血紅蛋白基因

人擇 (artificial selection)

天擇

平衡選擇

鐮刀型細胞貧血 vs. 瘧原蟲侵襲

血紅素: 紅血球中負責攜帶氧氣的蛋白質。

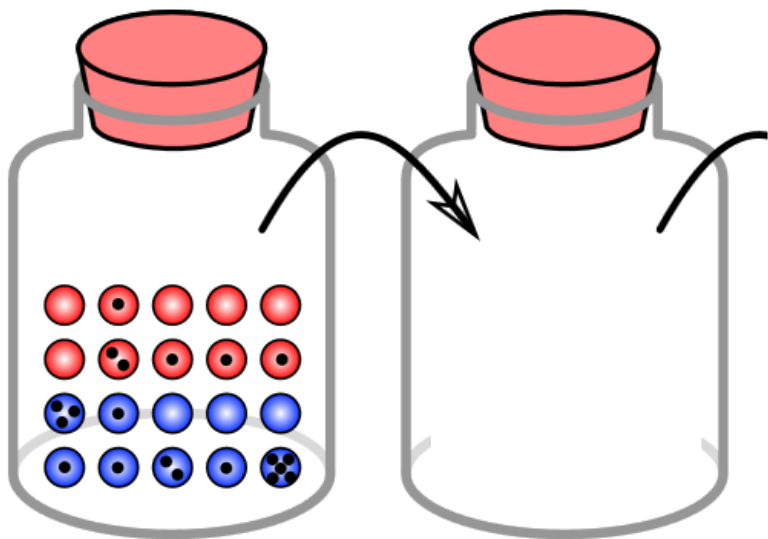
鐮型血球貧血症患者: 轉譯成血紅素的基因(HbB)發生點突變。

→ 紅血球的攜氧量降低

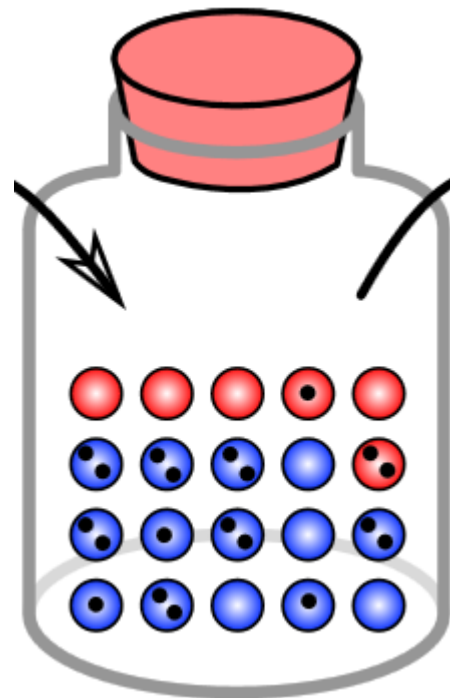
→ 嚴重貧血

非洲瘧疾流行的區域，瘧原蟲無法在鐮刀形紅血球內成長。

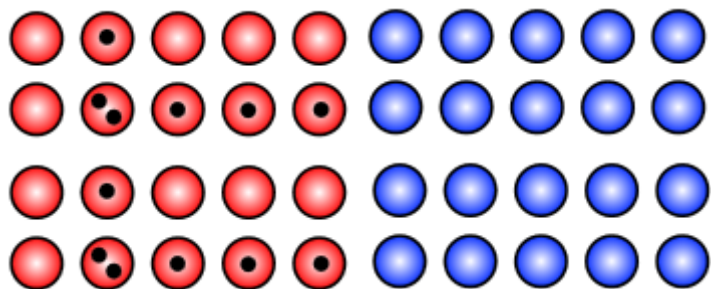
•Genetic drift (基因漂變)



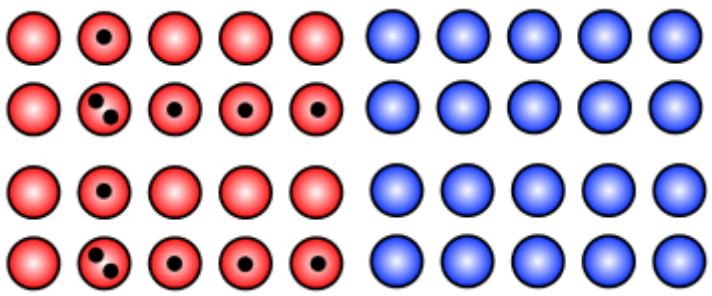
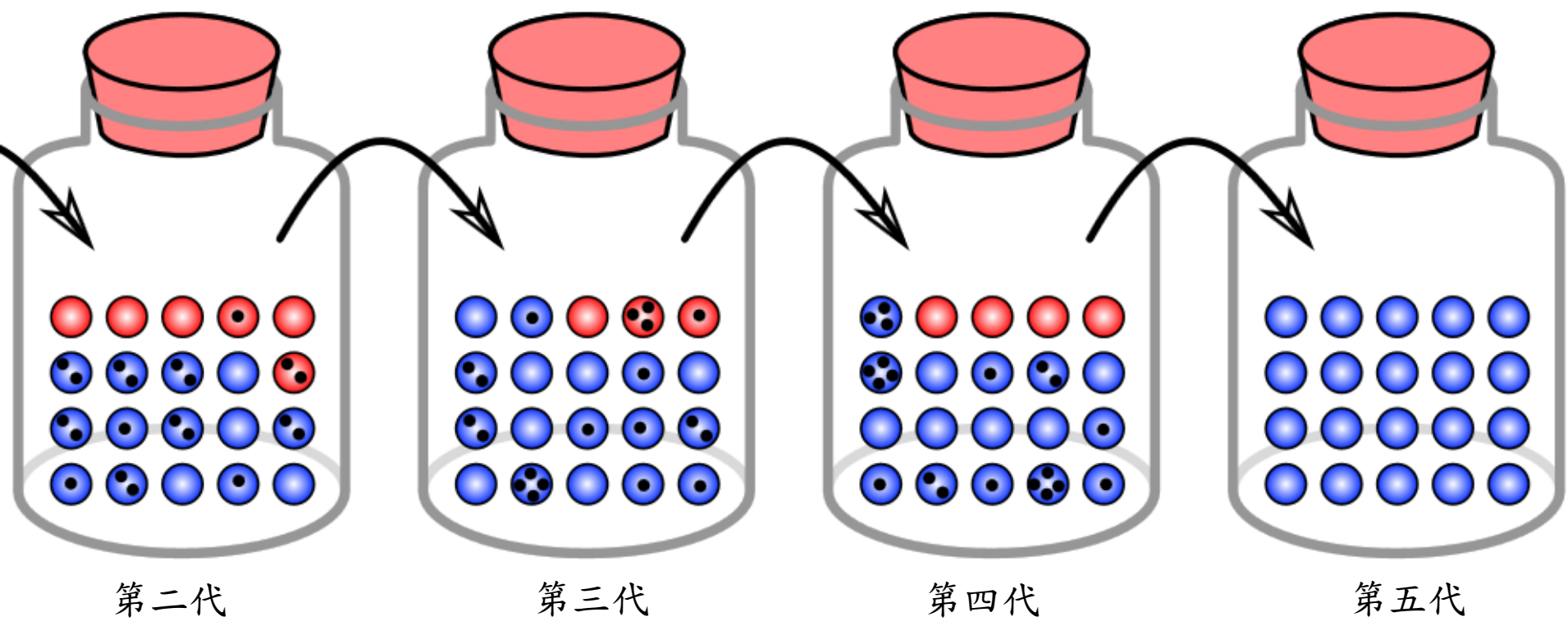
第一代



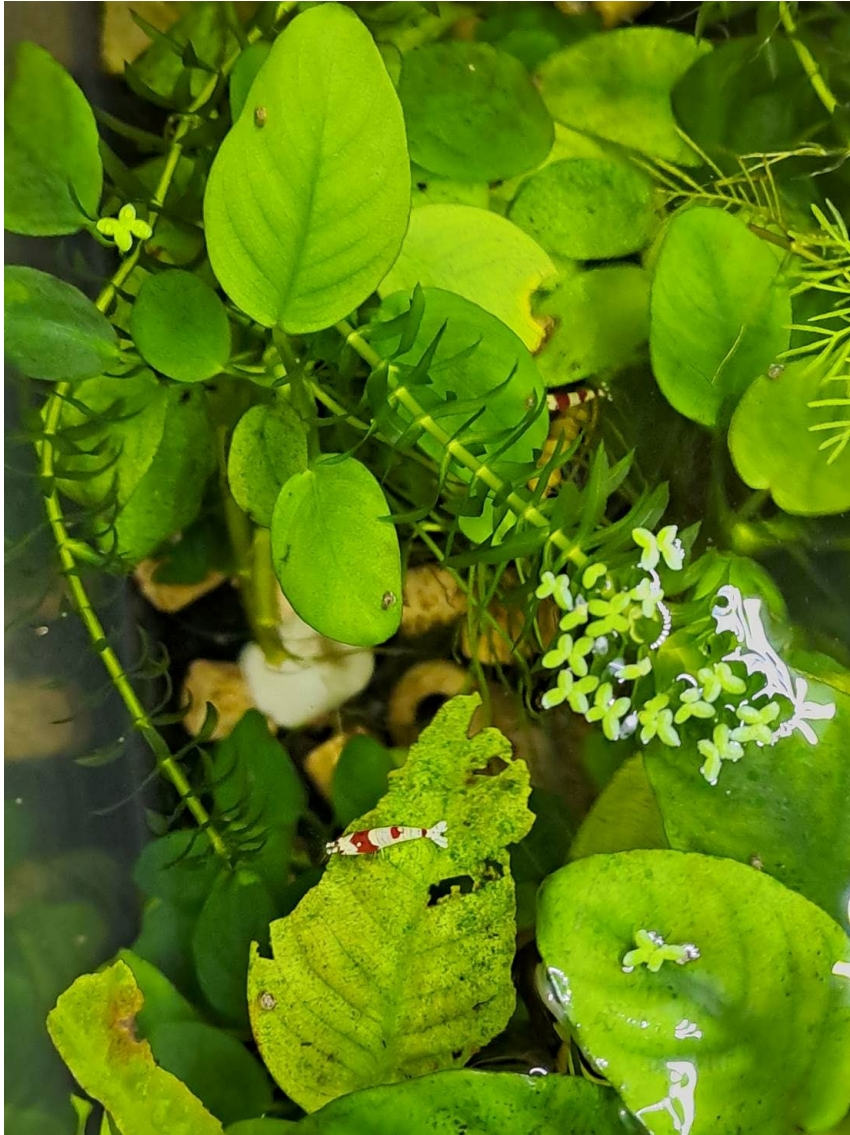
第二代



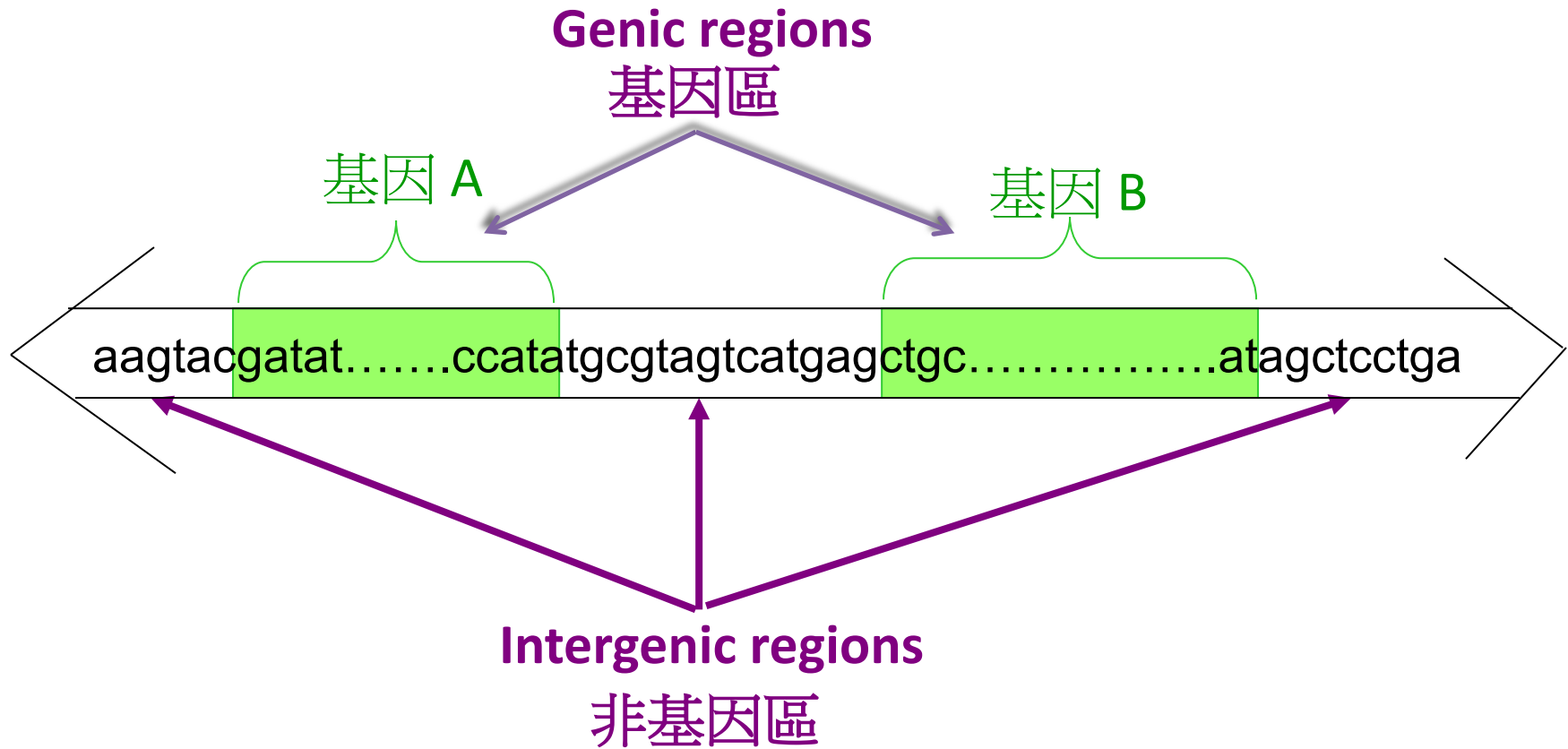
•Genetic drift (基因漂變)



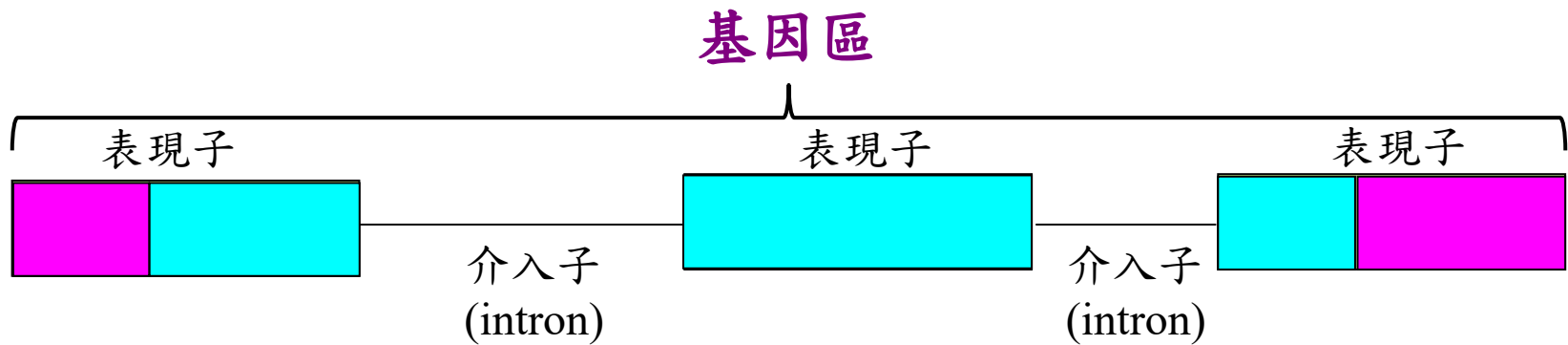
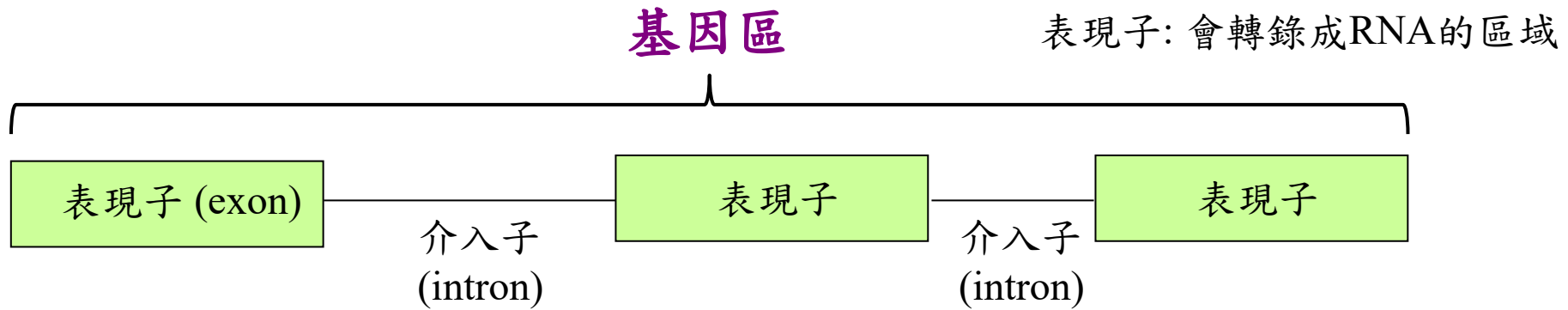
Gene flow (基因流動)



突變是隨機發生的。



突變是隨機發生的。

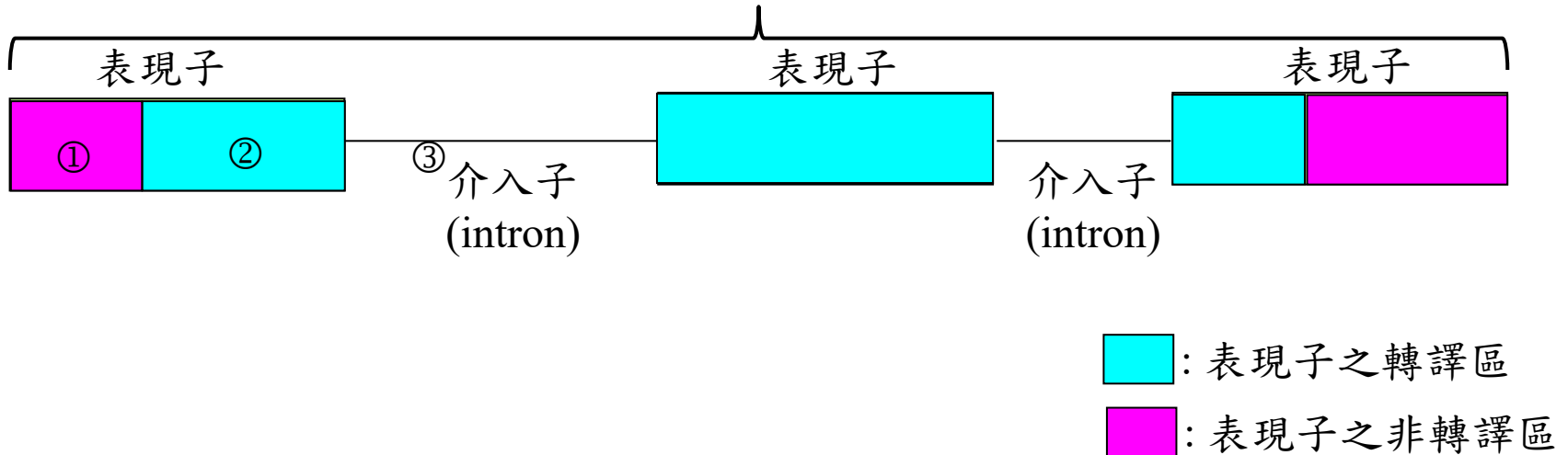


: 表現子之轉譯區

: 表現子之非轉譯區

突變是隨機發生的。

基因區



基因區 (genic region)

- Exon (表現子; 外顯子)
 - ✓ ORF (open reading frame; 轉譯區)
 - ✓ UTR (untranslated region; 非轉譯區)
- Intron (介入子; 內含子)

Genetic code

基因密碼

DNA sequence: A, C, G, T --- 4 letters

RNA sequence: A, C, G, U --- 4 letters

Amino acid sequence: --- 20 letters

64種密碼子 → 20種胺基酸
(鴿籠原理)

First Position (5')	Second position				Third Position (3')
	U	C	A	G	
U	Phe (F)	Ser (S)	Tyr (Y)	Cys (C)	U
	Phe (F)	Ser (S)	Tyr (Y)	Cys (C)	C
	Leu (L)	Ser (S)	Stop	Stop	A
	Leu (L)	Ser (S)	Stop	Trp (W)	G
C	Leu (L)	Pro (P)	His (H)	Arg (R)	U
	Leu (L)	Pro (P)	His (H)	Arg (R)	C
	Leu (L)	Pro (P)	Gln (Q)	Arg (R)	A
	Leu (L)	Pro (P)	Gln (Q)	Arg (R)	G
A	Ile (I)	Thr (T)	Asn (N)	Ser (S)	U
	Ile (I)	Thr (T)	Asn (N)	Ser (S)	C
	Ile (I)	Thr (T)	Lys (K)	Arg (R)	A
	Met (M)	Thr (T)	Lys (K)	Arg (R)	G
G	Val (V)	Ala (A)	Asp (D)	Gly (G)	U
	Val (V)	Ala (A)	Asp (D)	Gly (G)	C
	Val (V)	Ala (A)	Glu (E)	Gly (G)	A
	Val (V)	Ala (A)	Glu (E)	Gly (G)	G

開始編碼的密碼子(codon): AUG ← 必要但非充分條件

結束編碼密碼子: UAA, UAG, UGA ← 必要且充分條件

充分條件

A: 中樂透頭獎。 B: 覺得很快樂。

A: 下雨。 B: 無遮雨篷的操場濕掉。

A 是 B 的充分條件 (A發生一定導致B)

必要條件

A: 空氣的存在。 B: 人可以生存。

A: 努力。 B: 成功。

A: 父母。 B: 小孩。

A 是 B 的必要條件 (有B一定要有A)

充分且必要條件

A: 可被2整除的整數。 B: 偶數。

A: 可燃物、助燃物、溫度達到燃點。 B: 燃燒。

A: 吃某種食品或藥。 B: 病好了

Genetic code

基因密碼

64種密碼子 → 20種胺基酸
(鴿籠原理)

First Position (5')	Second position				Third Position (3')
	U(T)	C	A	G	
U(T)	Phe (F)	Ser (S)	Tyr (Y)	Cys (C)	U
	Phe (F)	Ser (S)	Tyr (Y)	Cys (C)	C
	Leu (L)	Ser (S)	Stop	Stop	A
	Leu (L)	Ser (S)	Stop	Trp (W)	G
C	Leu (L)	Pro (P)	His (H)	Arg (R)	U
	Leu (L)	Pro (P)	His (H)	Arg (R)	C
	Leu (L)	Pro (P)	Gln (Q)	Arg (R)	A
	Leu (L)	Pro (P)	Gln (Q)	Arg (R)	G
A	Ile (I)	Thr (T)	Asn (N)	Ser (S)	U
	Ile (I)	Thr (T)	Asn (N)	Ser (S)	C
	Ile (I)	Thr (T)	Lys (K)	Arg (R)	A
	Met (M)	Thr (T)	Lys (K)	Arg (R)	G
G	Val (V)	Ala (A)	Asp (D)	Gly (G)	U
	Val (V)	Ala (A)	Asp (D)	Gly (G)	C
	Val (V)	Ala (A)	Glu (E)	Gly (G)	A
	Val (V)	Ala (A)	Glu (E)	Gly (G)	G

aagtacgatatg **a**atagtaacata**aa**agtagtc**a**gctgg.....
M N S N I K V V M S W

First Position (5')	Second position				Third Position (3')
	U(T)	C	A	G	
U(T)	Phe (F)	Ser (S)	Tyr (Y)	Cys (C)	U
	Phe (F)	Ser (S)	Tyr (Y)	Cys (C)	C
	Leu (L)	Ser (S)	Stop	Stop	A
	Leu (L)	Ser (S)	Stop	Trp (W)	G
C	Leu (L)	Pro (P)	His (H)	Arg (R)	U
	Leu (L)	Pro (P)	His (H)	Arg (R)	C
	Leu (L)	Pro (P)	Gln (Q)	Arg (R)	A
	Leu (L)	Pro (P)	Gln (Q)	Arg (R)	G
A	Ile (I)	Thr (T)	Asn (N)	Ser (S)	U
	Ile (I)	Thr (T)	Asn (N)	Ser (S)	C
	Ile (I)	Thr (T)	Lys (K)	Arg (R)	A
	Met (M)	Thr (T)	Lys (K)	Arg (R)	G
G	Val (V)	Ala (A)	Asp (D)	Gly (G)	U
	Val (V)	Ala (A)	Asp (D)	Gly (G)	C
	Val (V)	Ala (A)	Glu (E)	Gly (G)	A
	Val (V)	Ala (A)	Glu (E)	Gly (G)	G

沉默突變 (silent mutation)：密碼改變，但對應的胺基酸不變。

aagtacgatatg aatagtaacataaagtagtcatgagctgg.....
M N S N I K V V M S W

aag
K

First Position (5')	Second position				Third Position (3')
	U(T)	C	A	G	
U(T)	Phe (F)	Ser (S)	Tyr (Y)	Cys (C)	U
	Phe (F)	Ser (S)	Tyr (Y)	Cys (C)	C
	Leu (L)	Ser (S)	Stop	Stop	A
	Leu (L)	Ser (S)	Stop	Trp (W)	G
C	Leu (L)	Pro (P)	His (H)	Arg (R)	U
	Leu (L)	Pro (P)	His (H)	Arg (R)	C
	Leu (L)	Pro (P)	Gln (Q)	Arg (R)	A
	Leu (L)	Pro (P)	Gln (Q)	Arg (R)	G
A	Ile (I)	Thr (T)	Asn (N)	Ser (S)	U
	Ile (I)	Thr (T)	Asn (N)	Ser (S)	C
	Ile (I)	Thr (T)	Lys (K)	Arg (R)	A
	Met (M)	Thr (T)	Lys (K)	Arg (R)	G
G	Val (V)	Ala (A)	Asp (D)	Gly (G)	U
	Val (V)	Ala (A)	Asp (D)	Gly (G)	C
	Val (V)	Ala (A)	Glu (E)	Gly (G)	A
	Val (V)	Ala (A)	Glu (E)	Gly (G)	G

錯義突變 (missense mutation)：使密碼所對應的胺基酸改變。

aagtacgatatg aatagtaacataaaagtagtcatgagctgg.....

M N S N I K V V M S W

aat

N

First Position (5')	Second position				Third Position (3')
	U(T)	C	A	G	
U(T)	Phe (F)	Ser (S)	Tyr (Y)	Cys (C)	U
	Phe (F)	Ser (S)	Tyr (Y)	Cys (C)	C
	Leu (L)	Ser (S)	Stop	Stop	A
	Leu (L)	Ser (S)	Stop	Trp (W)	G
C	Leu (L)	Pro (P)	His (H)	Arg (R)	U
	Leu (L)	Pro (P)	His (H)	Arg (R)	C
	Leu (L)	Pro (P)	Gln (Q)	Arg (R)	A
	Leu (L)	Pro (P)	Gln (Q)	Arg (R)	G
A	Ile (I)	Thr (T)	Asn (N)	Ser (S)	U
	Ile (I)	Thr (T)	Asn (N)	Ser (S)	C
	Ile (I)	Thr (T)	Lys (K)	Arg (R)	A
	Met (M)	Thr (T)	Lys (K)	Arg (R)	G
G	Val (V)	Ala (A)	Asp (D)	Gly (G)	U
	Val (V)	Ala (A)	Asp (D)	Gly (G)	C
	Val (V)	Ala (A)	Glu (E)	Gly (G)	A
	Val (V)	Ala (A)	Glu (E)	Gly (G)	G

無義突變 (nonsense mutation)：使原本可製造蛋白質的密碼變成停止密碼。

aagtacgatatg **a**tagt**a**acata**a**aagtagt**c**atg**a**g**t**gg.....
M N S N I K V V M S W

tag

Stop codon

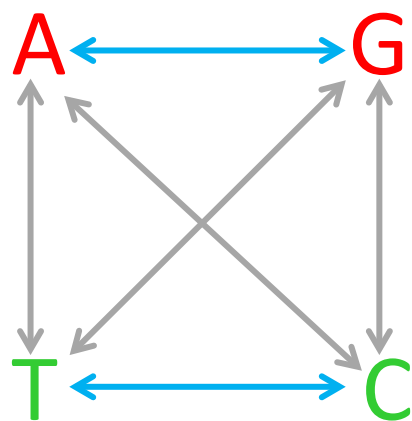
突變是隨機發生的。

A (adenine, 腺嘌呤)

G (guanine, 鳥嘌呤)

T (thymine, 胸腺嘧啶)

C (cytosine, 胞嘧啶)



嘔呤類

嘧啶類

突變是隨機發生的。

A (adenine, 腺嘌呤)

G (guanine, 鳥嘌呤)

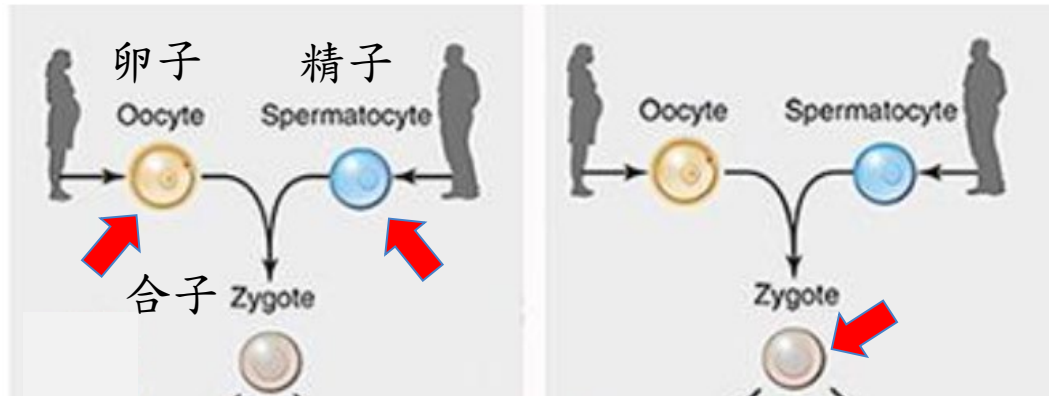
T (thymine, 胸腺嘧啶)

C (cytosine, 胞嘧啶)

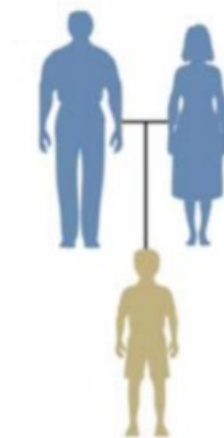
地址：

115201臺北市南港區研究院路一段128號

de novo mutation (新發生的突變)

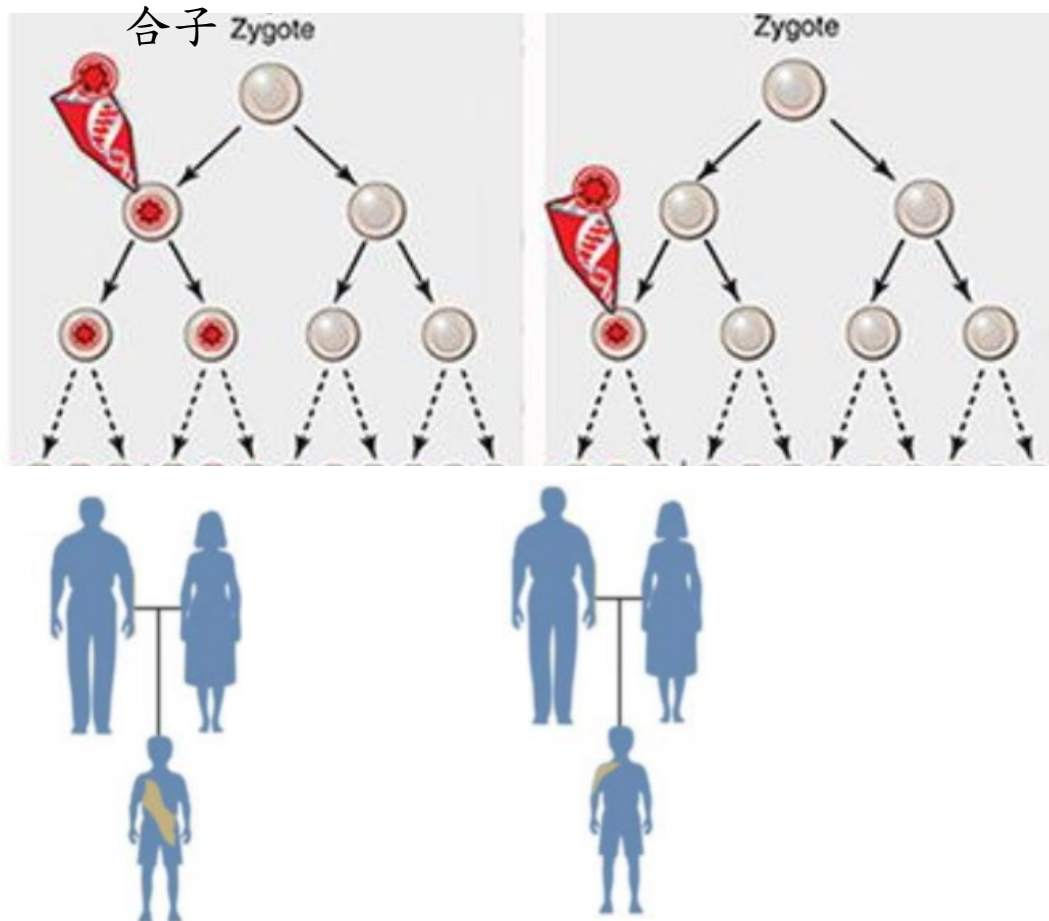


[Courtesy of Science/AAAS, Poduri et al., 2013.]



Modified from *Genes* **2014**, 5(4), 1064-1094.

Mosaicism (鑲嵌現象)



[Courtesy of
Science/AAAS, Poduri et
al., 2013.]

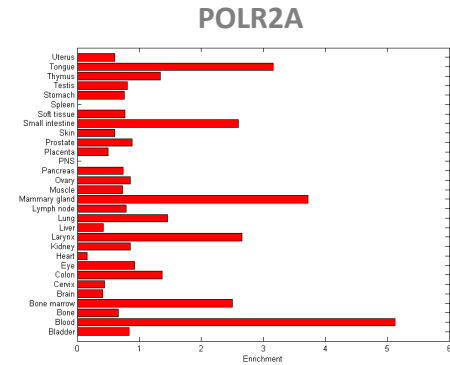
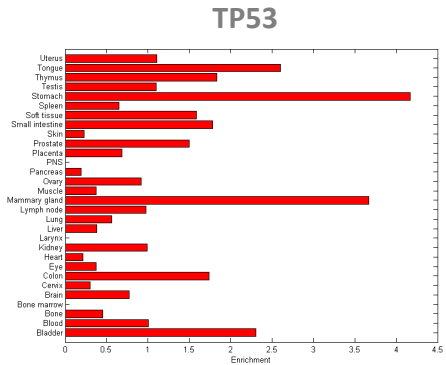
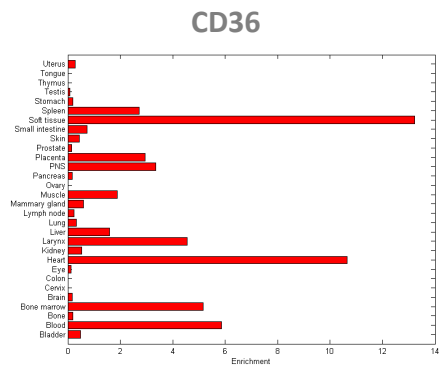
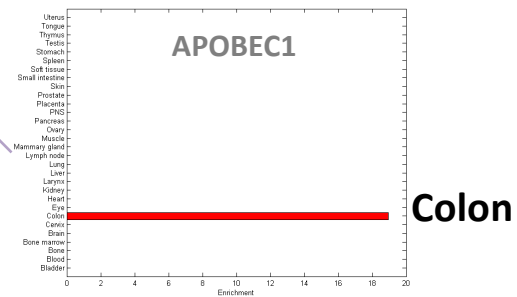
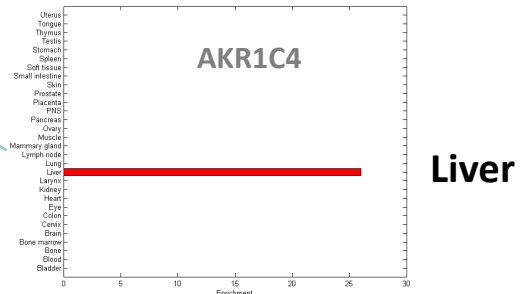
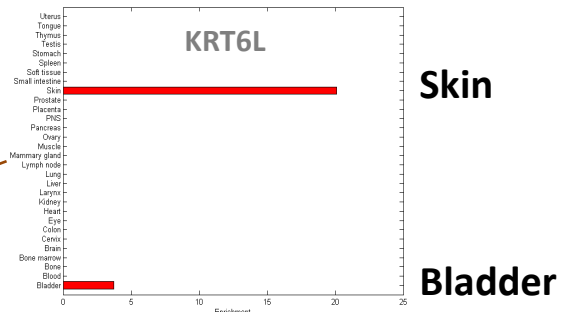
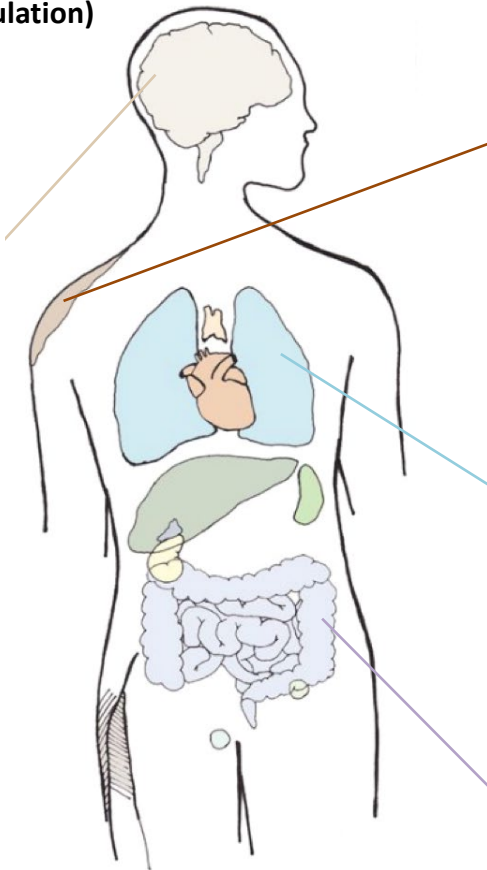
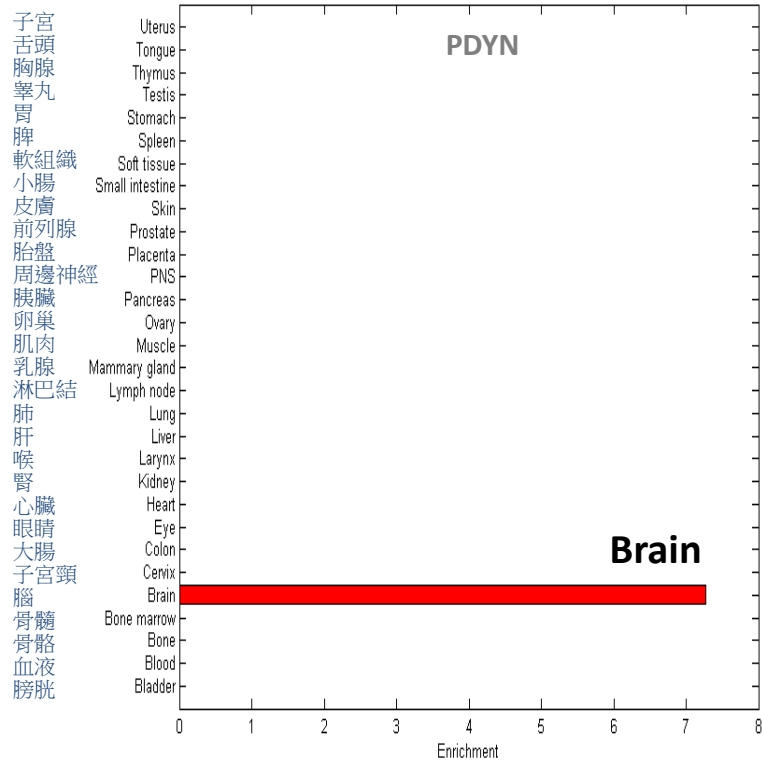
人有多少個基因？

	Organism	# of protein-coding genes	Genome size
viruses	人類免疫缺陷病毒 HIV 1	9	10
	甲型流感病毒 Influenza A virus	10-11	14
	噬菌體 Bacteriophage λ	66	49
	人類皰疹病毒第四型 Epstein Barr virus	80	170
prokaryotes	鬼羽箭屬菌 <i>Buchnera sp.</i>	610	640
	海棲熱袍菌 <i>T. maritima</i>	1,900	1,900
	金黃色葡萄球菌 <i>S. aureus</i>	2,700	2,900
	霍亂弧菌 <i>V. cholerae</i>	3,900	4,000
	枯草桿菌 <i>B. subtilis</i>	4,400	4,200
	大腸桿菌 <i>E. coli</i>	4,300	4,600
	釀酒酵母 <i>S. cerevisiae</i>	6,600	12,000
eukaryotes	秀麗隱桿線蟲 <i>C. elegans</i>	20,000	100,000
	阿拉伯芥 <i>A. thaliana</i>	27,000	140,000
	黑腹果蠅 <i>D. melanogaster</i>	14,000	140,000
	紅鰭東方魴 <i>F. rubripes</i>	19,000	400,000
	玉米 <i>Z. mays</i>	33,000	2,300,000
	小鼠 <i>M. musculus</i>	20,000	2,800,000
	智人 <i>H. sapiens</i>	21,000	3,200,000
	普通小麥 <i>T. aestivum</i> (hexaploid)	95,000	16,800,000

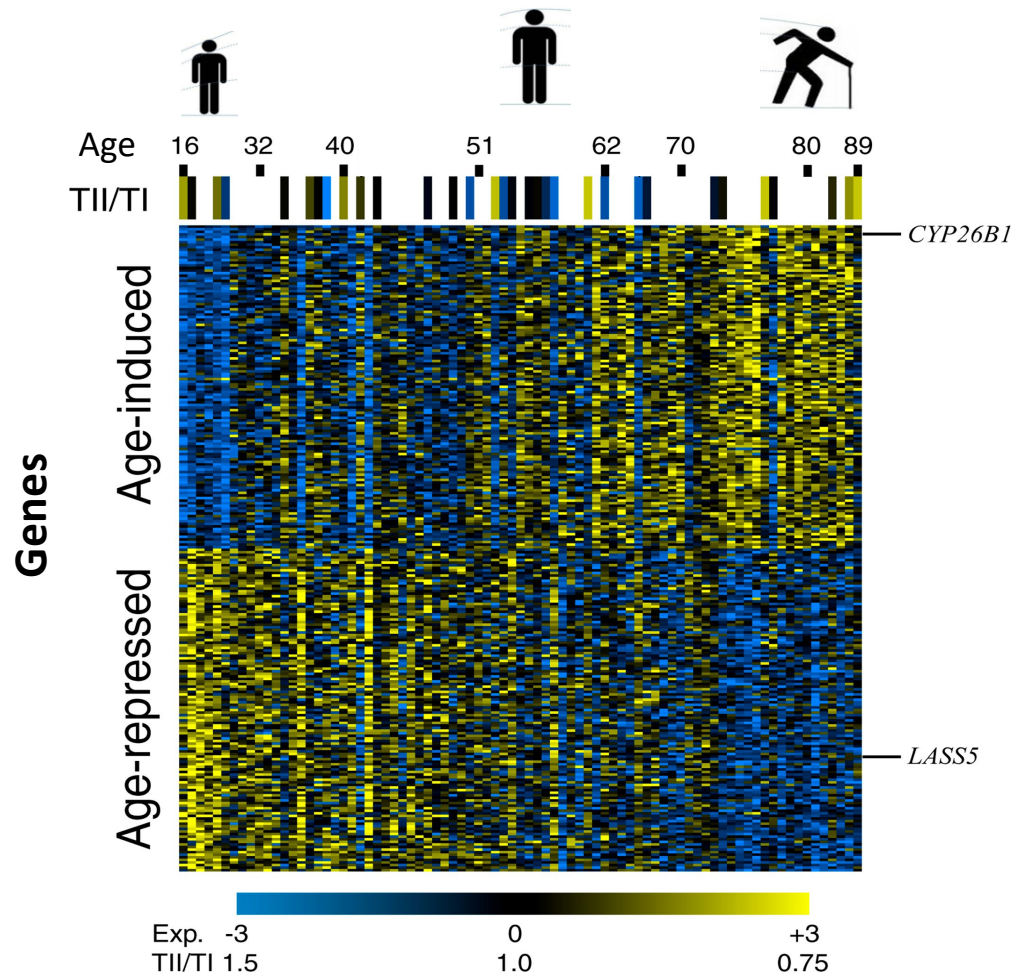
($\times 10^3$)

(<http://book.bionumbers.org/how-many-genes-are-in-a-genome/>)

TiGER database (Tissue-specific Gene Expression and Regulation)



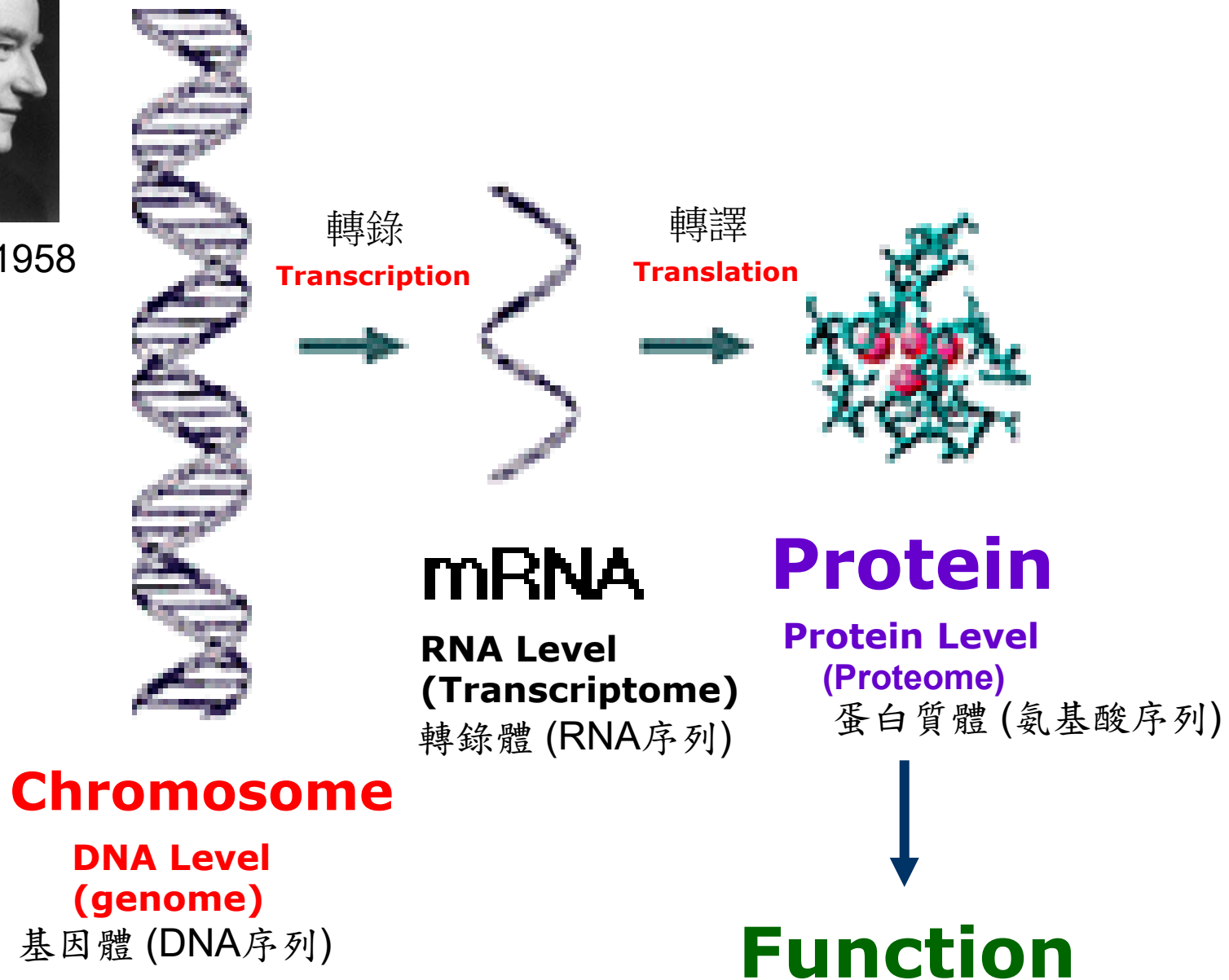
年輕人的基因與老年人的基因



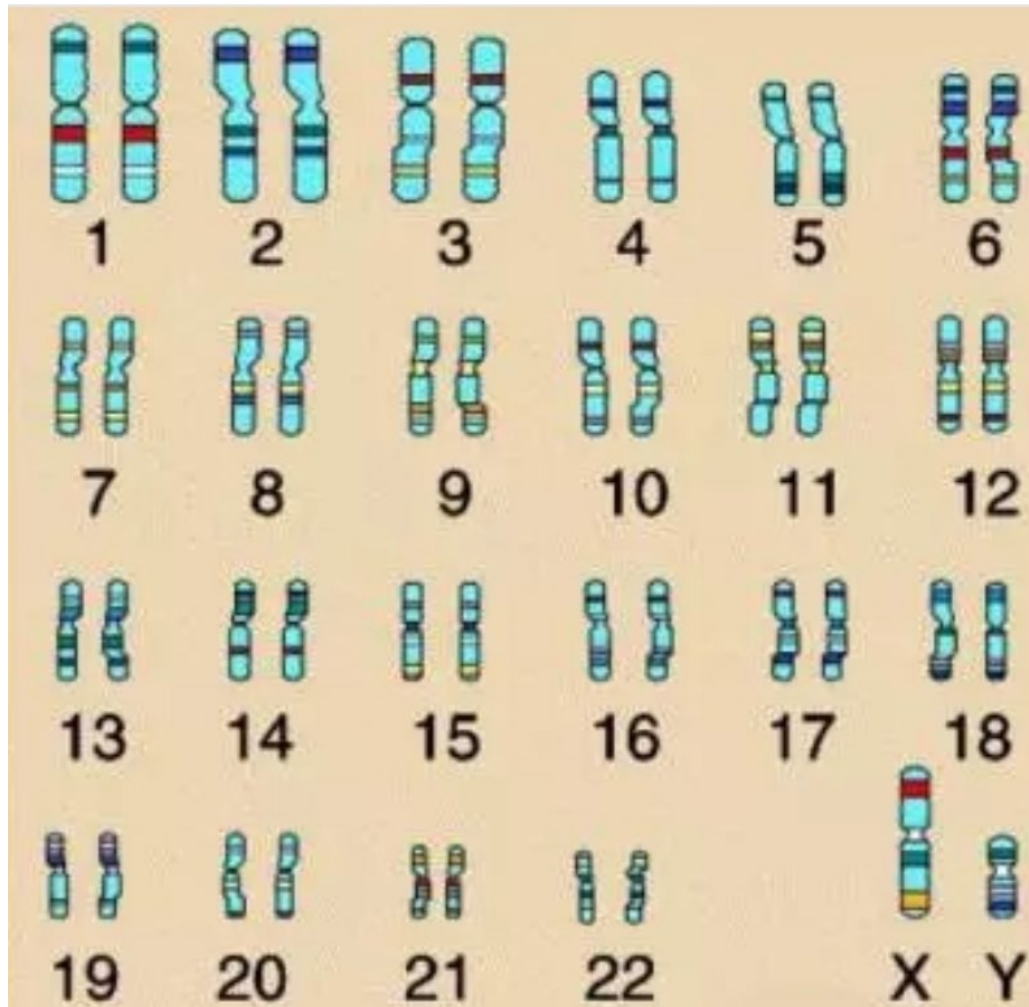
分子生物學的中心法則 (central dogma)



克拉克; 1958



所有 DNA 序列的總稱叫做: 基因體 (genome)



常染色體 (autosomes)

性染色體 (sex chromosome)

人類的DNA序列由30億 (3×10^9) 個A、G、T、C核苷酸排列組合而成

Causative Biology (因果生物學)

- DNA和疾病間的關係
- RNA和疾病間的關係
- DNA和RNA間的關係
- DNA、RNA和疾病間的關係

DNA 間的差異和疾病

GWAS: Genome-Wide Association Study

全基因體關聯性分析研究

單核苷酸多態性 (Single Nucleotide Polymorphism; SNP)



aagtacgatcggccga

aagtgcgatcggccga

aagtacgatcggccga

aagtacgatcggccga

aagtgcgatcggccga

aagtgcgatcggccga

基因型 (genotype): aa, gg, ag

同型合子 (homozygous) → aa, gg

異型合子 (heterozygous) → ag

單核苷酸多態性 (Single Nucleotide Polymorphism; SNP)

aagt**a**cgat**c**ggccga
aagt**g**cgat**t**ggccga

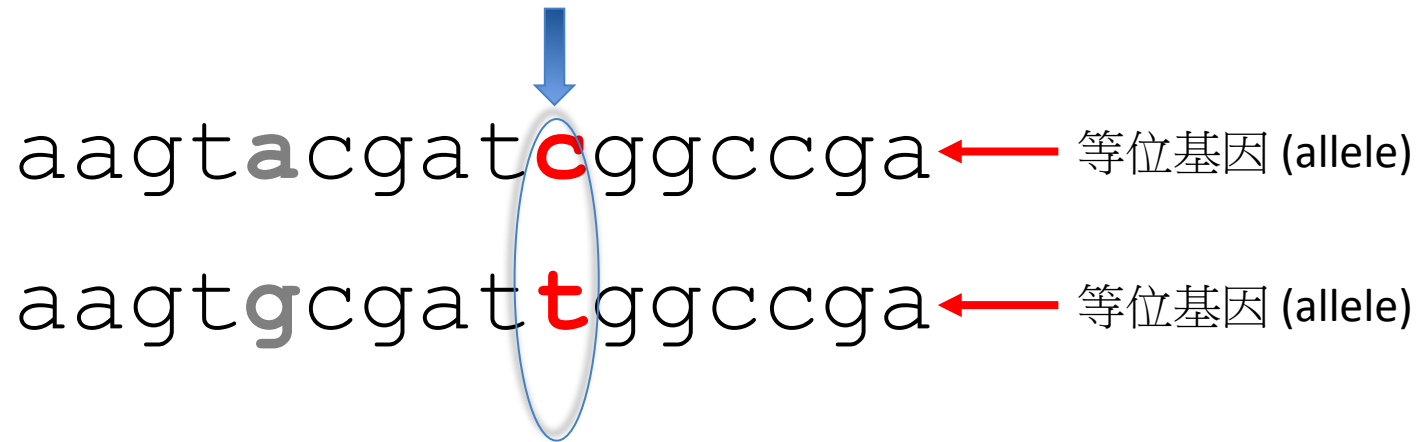


基因型 (genotype): cc, ct, tt

同型合子 (homozygous)

異型合子 (heterozygous)

單核苷酸多態性 (Single Nucleotide Polymorphism; SNP)



基因型 (genotype): cc, ct, tt

同型合子 (homozygous)

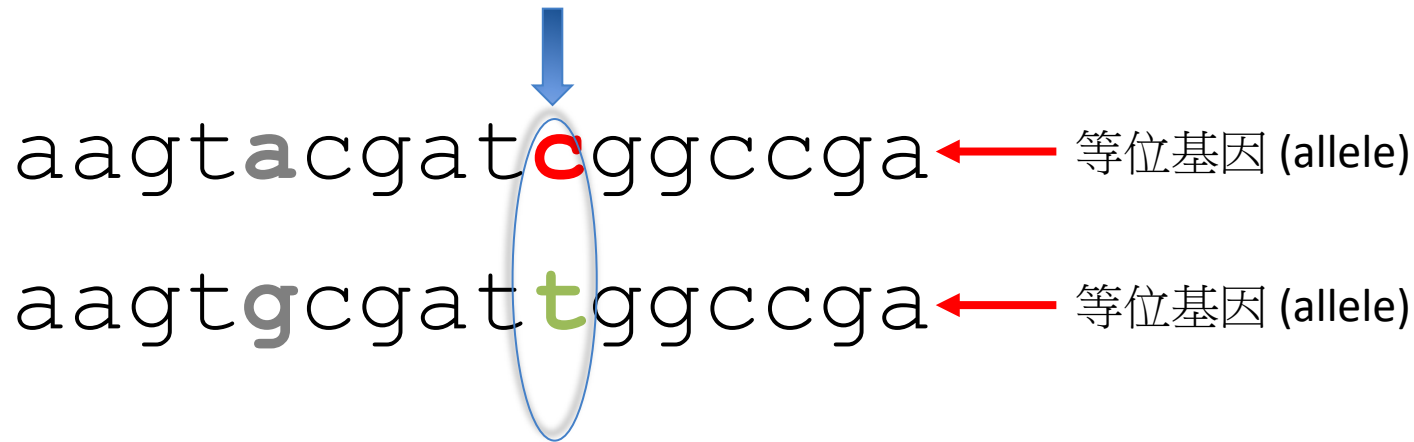
異型合子 (heterozygous)

人是二倍體

只有精子和卵子是單倍體

其他體細胞都是雙倍體

單核苷酸多態性 (Single Nucleotide Polymorphism; SNP)



基因型 (genotype): cc, ct, tt

100個人共有幾條alleles?

同型合子 (homozygous)

90個人基因型為 **cc**

異型合子 (heterozygous)

3個人基因型為 **ct**

7個人基因型為 **tt**

c的allele frequency=? $(90 \times 2 + 3) / (100 \times 2) = 91.5\%$

t的allele frequency=? $(7 \times 2 + 3) / (100 \times 2) = 8.5\%$

遺傳性疾病

是指以基因為主要致病原因的疾病。

單一基因缺陷造成的遺傳疾病

- 鎌刀型貧血症
- 白化症

多重基因共同影響所造成的遺傳疾病

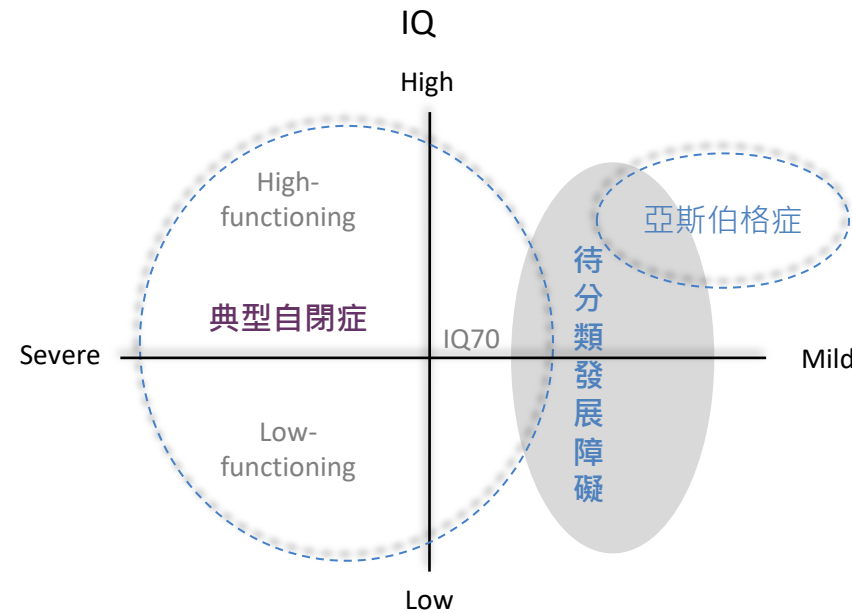
- 冠狀心臟疾病
- 高血壓
- 中風
- 許多種類的癌症
- 自閉症

Autism spectrum disorders (ASD)

自閉症譜系障礙

自閉症譜系障礙是一種腦部發育障礙所導致的複雜疾病。

患者的主要特徵是在社交溝通、互動以及表達上有障礙，對限制性行為且重複的動作有強烈興趣。譜系障礙意指此病症患者因障礙程度不同而跨距譜系很大，依不同程度的症狀、智力及語言發展，自閉症一般可分為幾種不同亞型如：典型自閉症、亞斯伯格症、待分類廣泛性發展障礙等。



Coplan, James. Making sense of autistic spectrum disorders: Create the brightest future for your child with the best treatment options. Bantam, 2010.

自閉症的盛行率逐年攀升，在已開發國家中盛行率約為1-2%。

美國：

2016普查結果顯示，每54位8歲以前的兒童就有1位 (1.85%)被診斷為自閉症。

臺灣：

根據我國衛生福利部統計資料，截至2023年，我國的自閉症患者已超過20,656人。

DNA和疾病間的關係

GWAS: Genome-Wide Association Study
全基因體關聯性分析研究

Grove et al. *Nature Genetics* (2019)

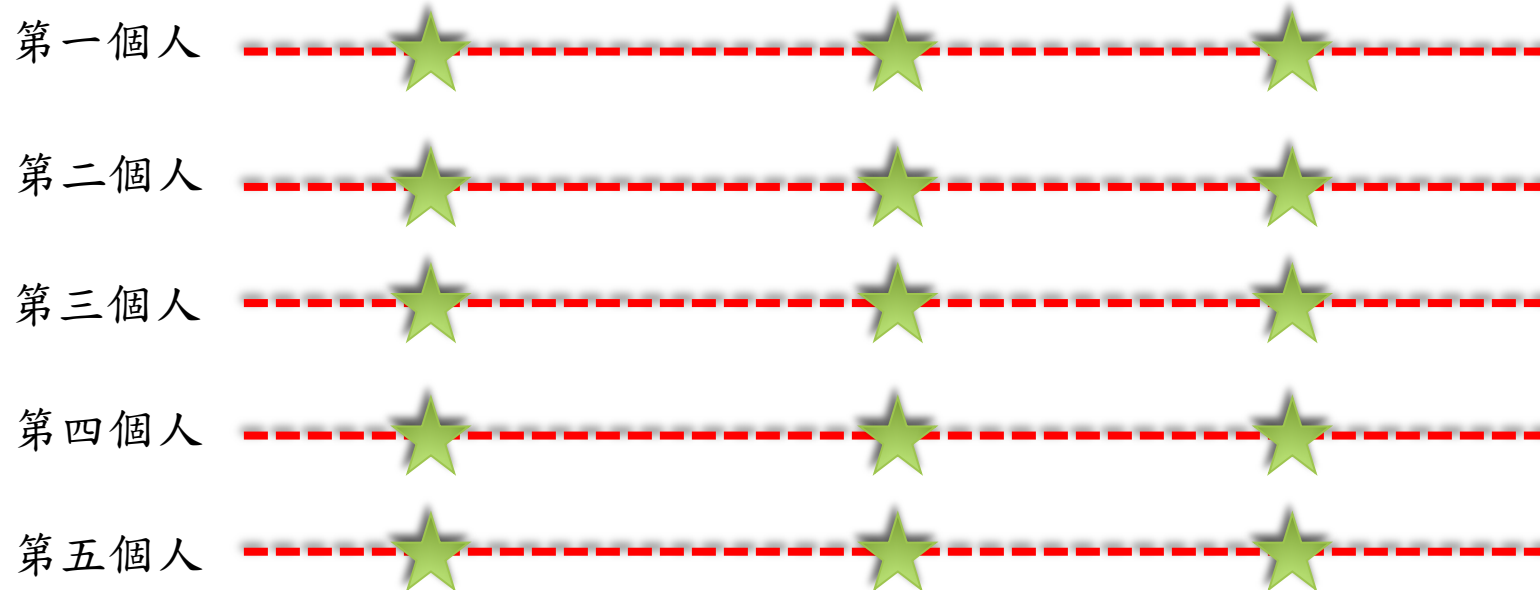
DNA 間的差異 (自閉症患者 vs. 無自閉症健康者)

自閉症患者 (18,381人)

健康者 (27,969人)

偵測 9,112,386 個位點

全基因體位點檢測率？



DNA和疾病間的關係

GWAS: Genome-Wide Association Study
全基因體關聯性分析研究

Grove et al. *Nature Genetics* (2019)

自閉症患者 (18,381人)

A ||| A A ||| A A ||| A A ||| A

A是主要等位基因 → ~100%

G是次要等位基因 → ~0%

健康者 (27,969人)

G ||| G G ||| G G ||| G G ||| G

G是主要等位基因 → ~100%

A是次要等位基因 → ~0%

DNA 間的差異 (自閉症患者 vs. 無自閉症健康者)

自閉症是一種複雜疾病。

自閉症譜系障礙的成因目前仍未有定論，普遍認為與遺傳因素有相當大的關係。

然而，患者間**基因變異的異質性**很大，絕大部分找到的基因變異對致病性的解釋能力都不高。除了基因變異之外，許多研究也發現環境以及環境與基因間的交互影響對自閉症的形成扮演一定的角色。雖然近年來陸續有一些研究利用腦組織來探討自閉症的轉錄體，目前人們對自閉症的致病分子機制仍所知甚少。

DNA 間的差異 (自閉症患者 vs. 無自閉症健康者)

GWAS: Genome-Wide Association Study
全基因體關聯性分析研究

Grove et al. *Nature Genetics* (2019)

自閉症患者 (18,381人)

健康者 (27,969人)

偵測 9,112,386 個位點

$P < 0.05 \rightarrow \#620,887$

$P < 1e-5 \rightarrow \#4,027$

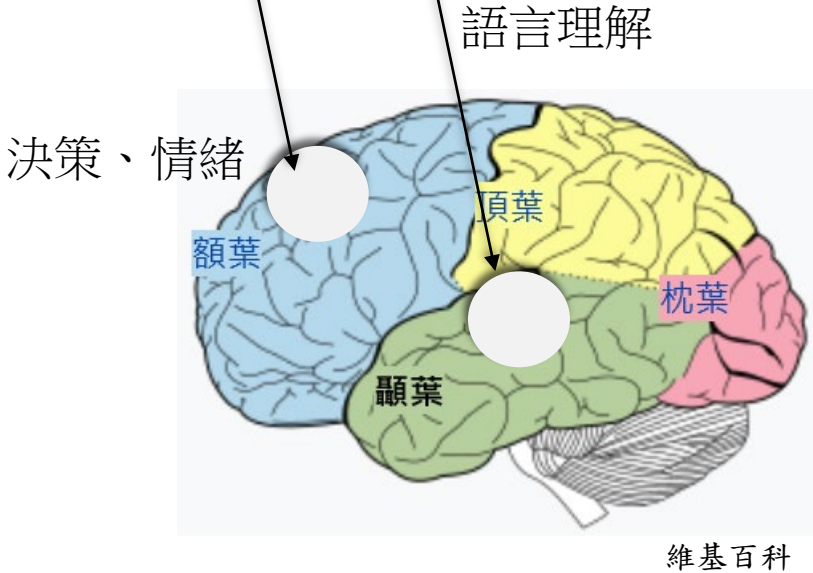
$P < 5e-8 \rightarrow \#93$

RNA和疾病間的關係

RNA 間的差異和疾病 (表現型差異)

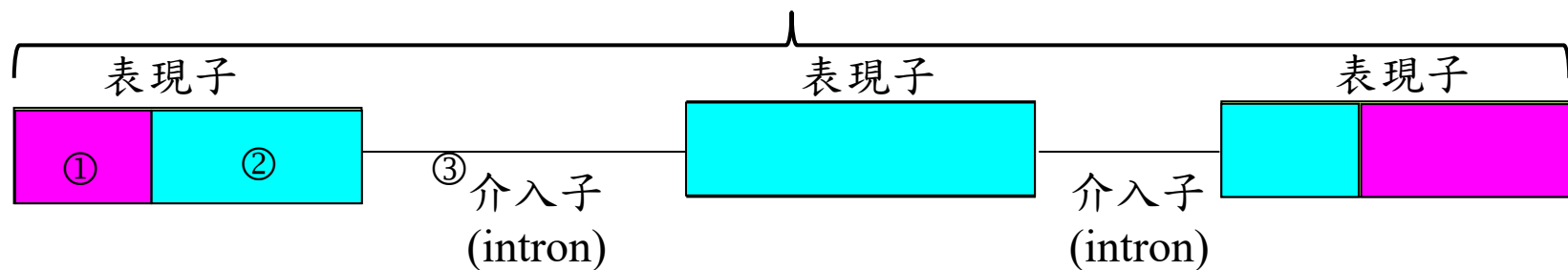
RNA-seq		45 ASD cases (自閉症)	43 CTL cases (健康者)	Sum
額葉	Frontal cortex (FC)	37	36	73
顳葉	Temporal cortex (TC)	36	25	61
Total		73	61	134

RNA-seq data of post-mortem brain tissue samples from the Synapse database: syn458609 (*Nature*, v540, p423, 2016)



RNA和疾病間的關係

基因區



■: 表現子之轉譯區

■: 表現子之非轉譯區

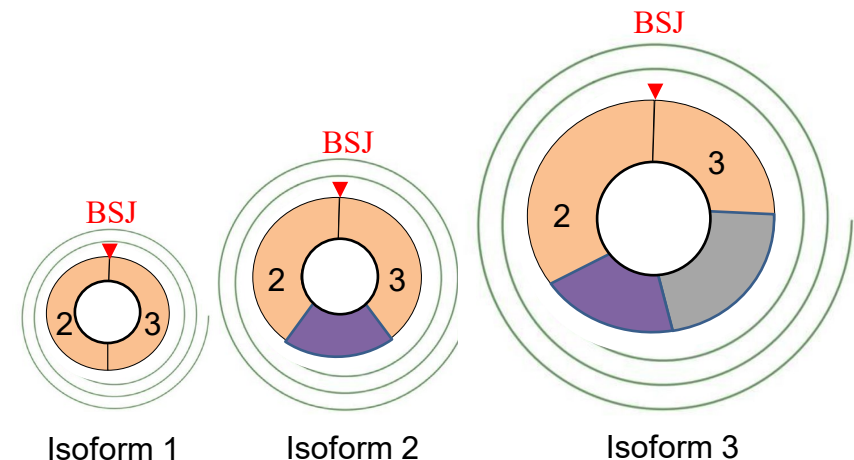
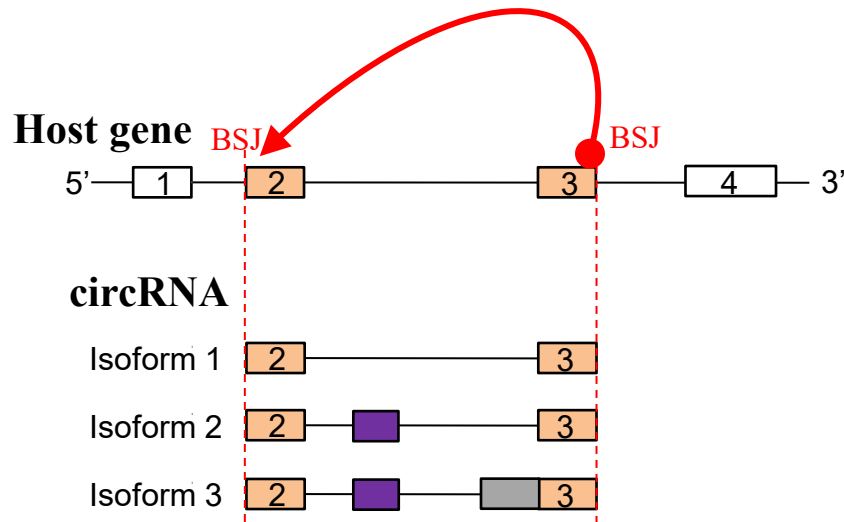
基因區 (genic region)

- Exon (表現子; 外顯子)
 - ✓ ORF (open reading frame; 轉譯區)
 - ✓ UTR (untranslated region; 非轉譯區)
- Intron (介入子; 內含子)

RNA和疾病間的關係

FL-circAS → <https://cosbi.ee.ncku.edu.tw/FL-circAS/>

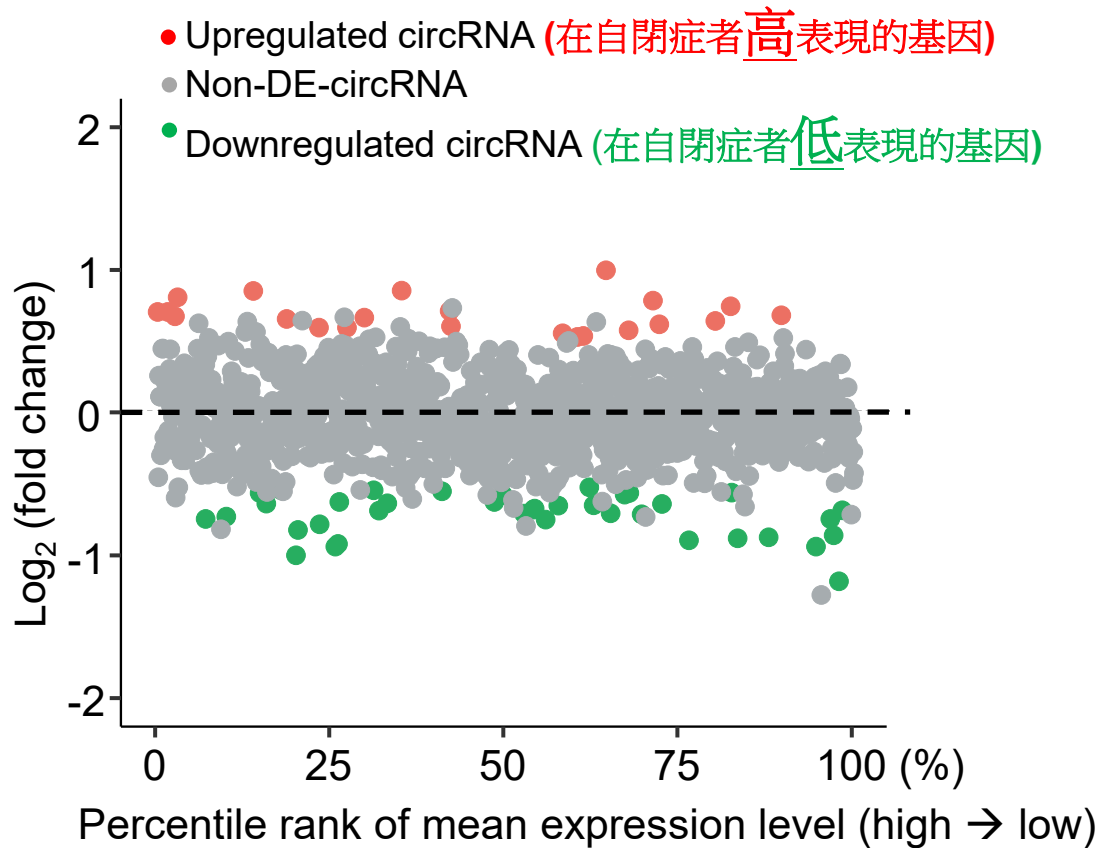
Nanopore long-read sequencing with circRNA enrichment bring an unprecedented opportunity for investigating **full-length alternatively spliced (AS) circRNA isoforms**.



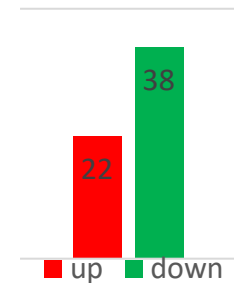
- Yu-Chen Chen, Chia-Ying Chen, Tai-Wei Chiang, Ming-Hsien Chan, Michael Hsiao, Huei-Mien Ke, Isheng Jason Tsai, and **Trees-Juen Chuang*** (2023) Detecting intragenic *trans*-splicing events from non-co-linearly spliced junctions by hybrid sequencing. *Nucleic Acids Research*, 51(15): 7777-7797.
- Tai-Wei Chiang, Song-En Jhong, Yu-Chen Chen, Chia-Ying Chen, Wei-Sheng Wu*, **Trees-Juen Chuang*** (2024) FL-circAS: an integrative resource and analysis for full-length sequences and alternative splicing of circular RNAs with nanopore sequencing. *Nucleic Acids Research*, 52: D115-D123.

RNA和疾病間的關係

A linear mixed effects (LME) model with controlling for sex, age, brain region (PC or TC), RNA quality (RIN), host gene expression, sequencing batch, and brain bank batch (性別、年齡、腦區、定序品質、定序單位、樣本來源.....)

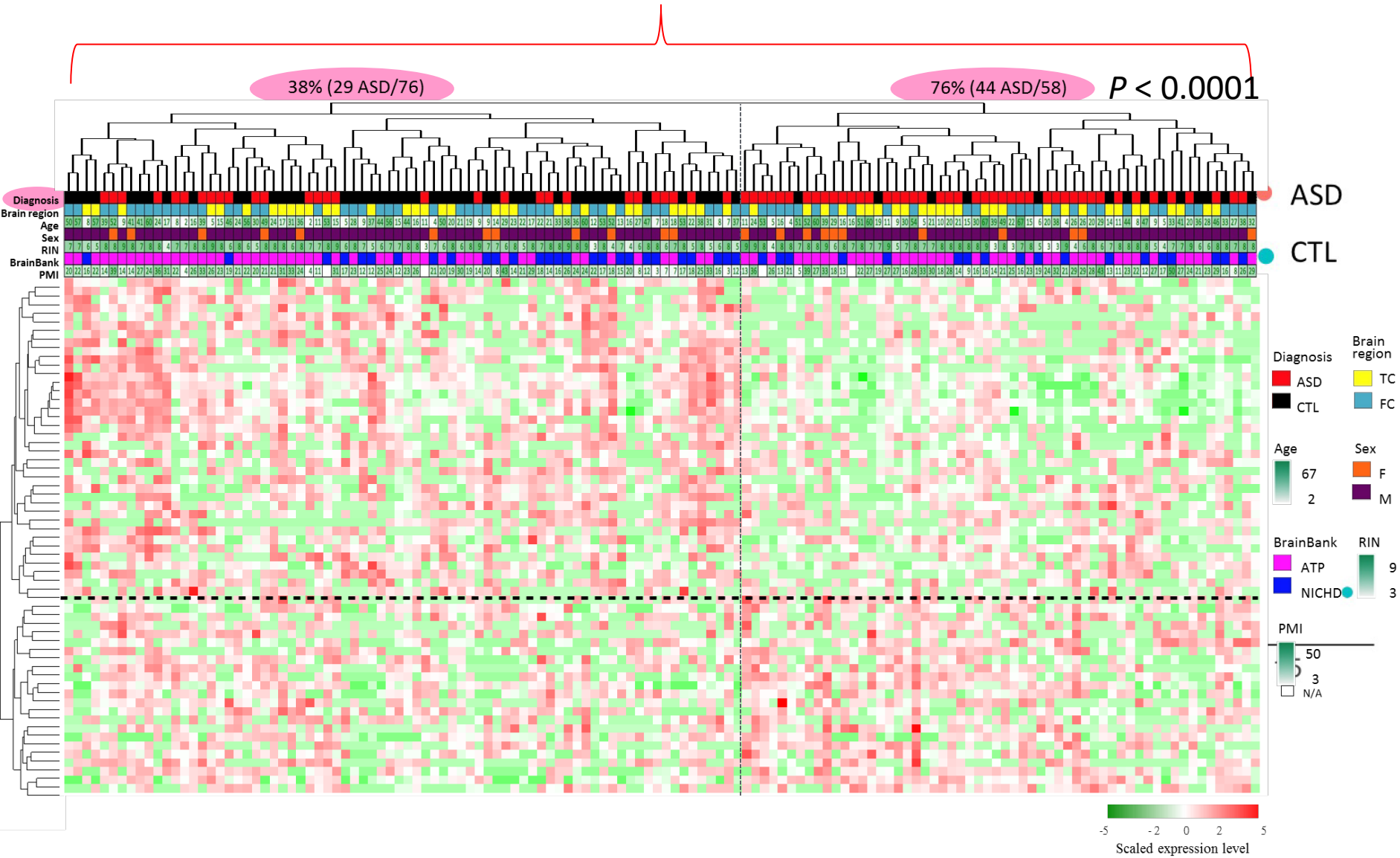


DE-circRNA (60)



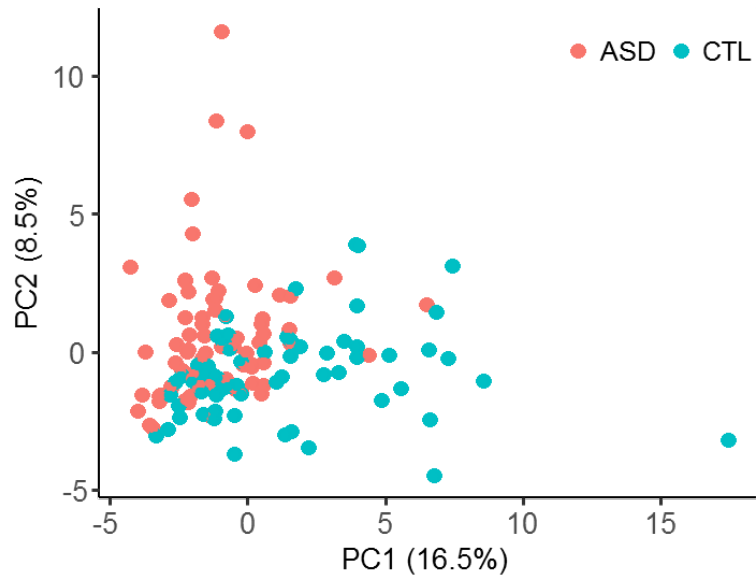
RNA和疾病間的關係

134 cortex samples



RNA和疾病間的關係

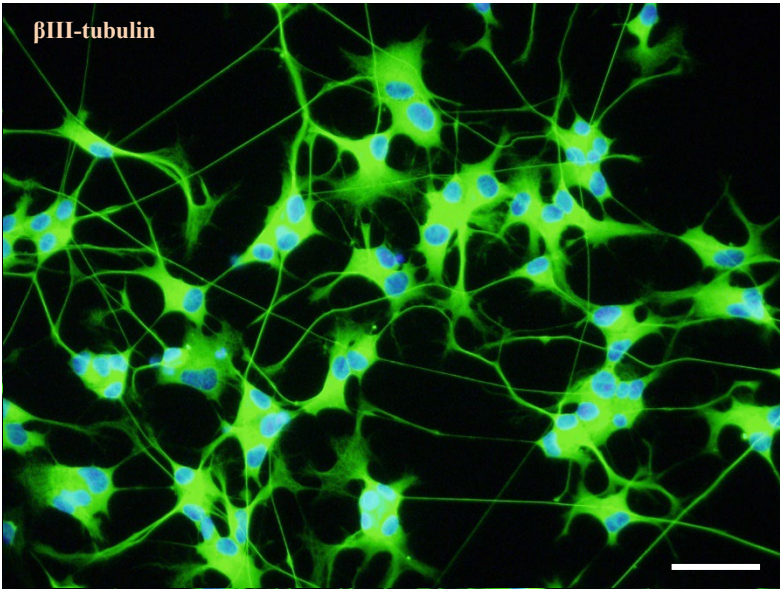
60 DE-circRNAs in ASD & CTL



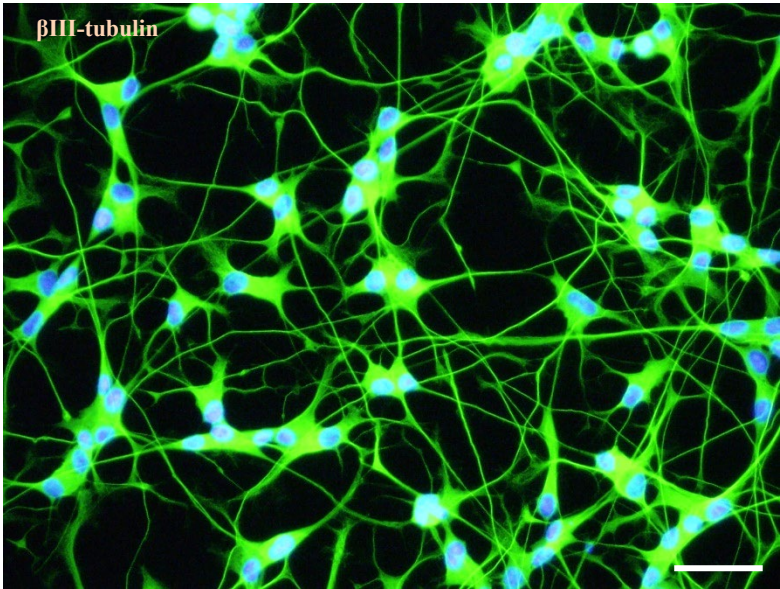
Lenti-virus
M.O.I.=3
(Puromycin selection)

(40x)

CT

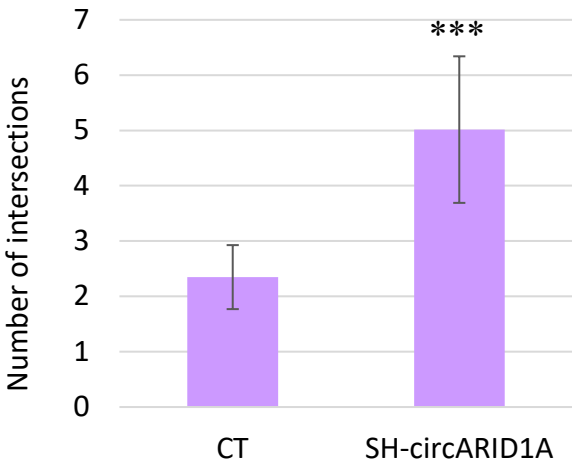
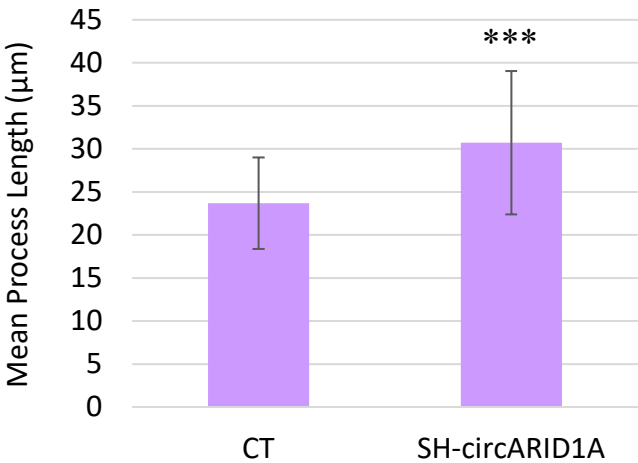
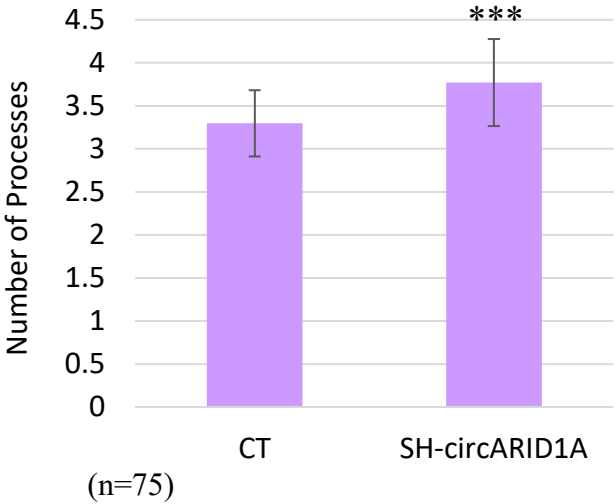


circARID1A KD



Scale bar = 50 μ m

Metamorph analysis of neurite



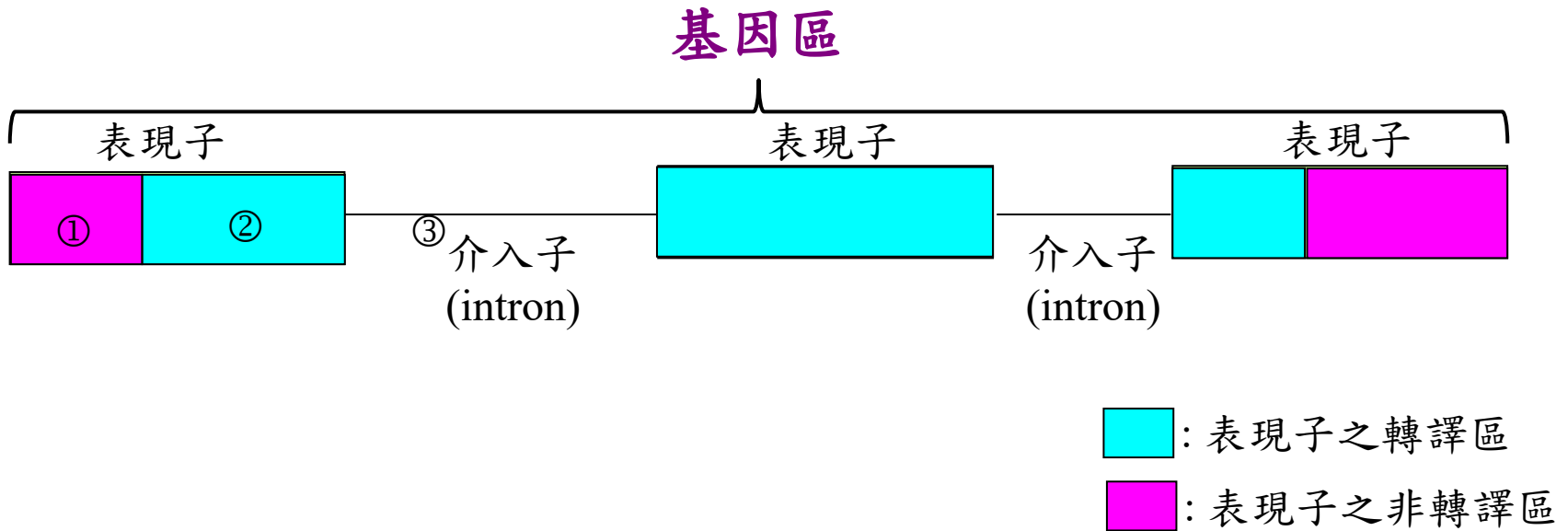


Google 關鍵字: 被圈住的心靈

PART I: 首度建構自閉症腦組織的環狀RNA與基因間調控網路圖譜

PART II: 探討基因型、環狀RNA與自閉症間的因果關係網路

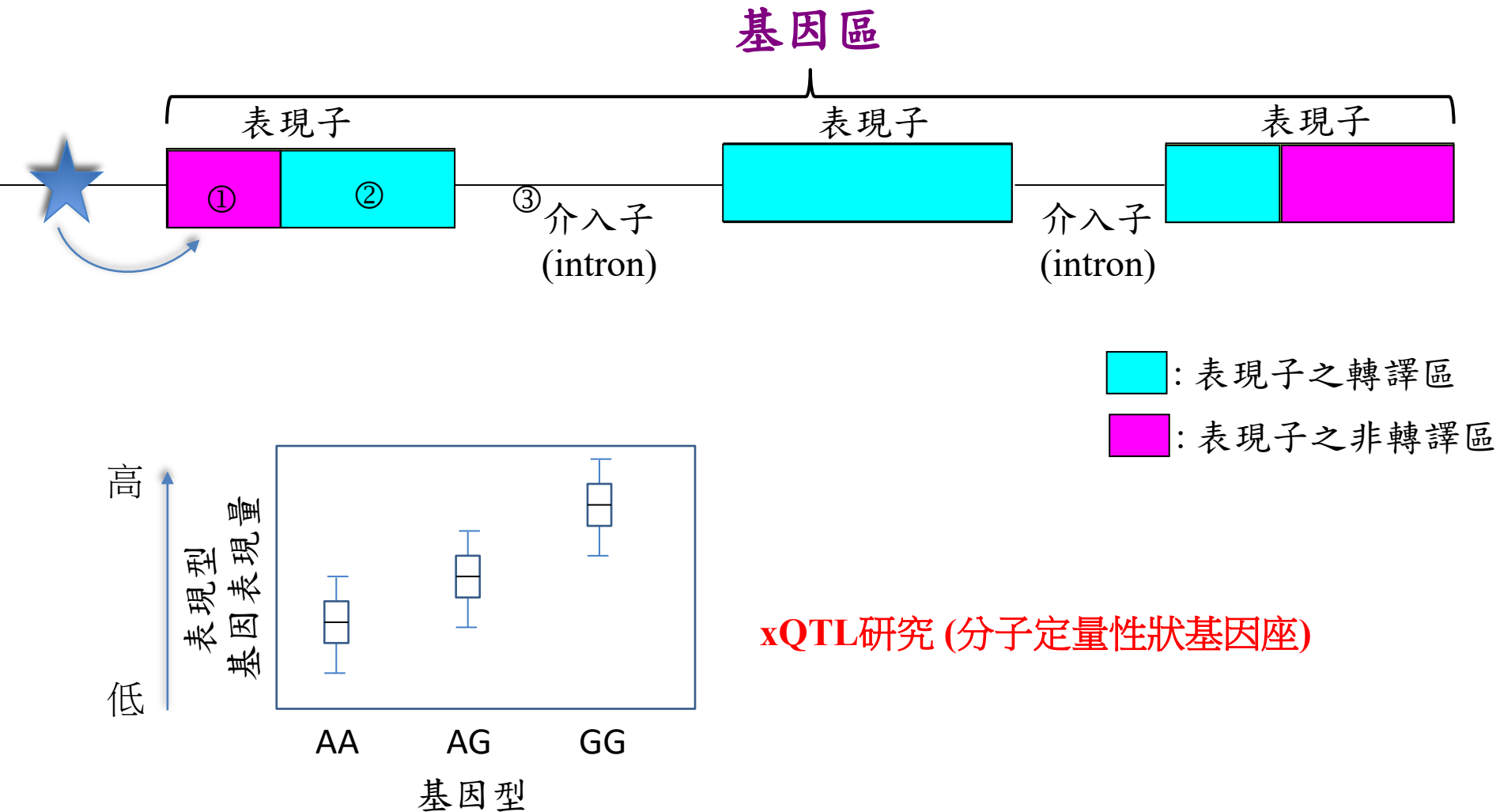
DNA和RNA間的關係



基因區 (genic region)

- Exon (表現子; 外顯子)
 - ✓ ORF (open reading frame; 轉譯區)
 - ✓ UTR (untranslated region; 非轉譯區)
- Intron (介入子; 內含子)

DNA和RNA間的關係



單核苷酸多態性 (Single Nucleotide Polymorphism; SNP)

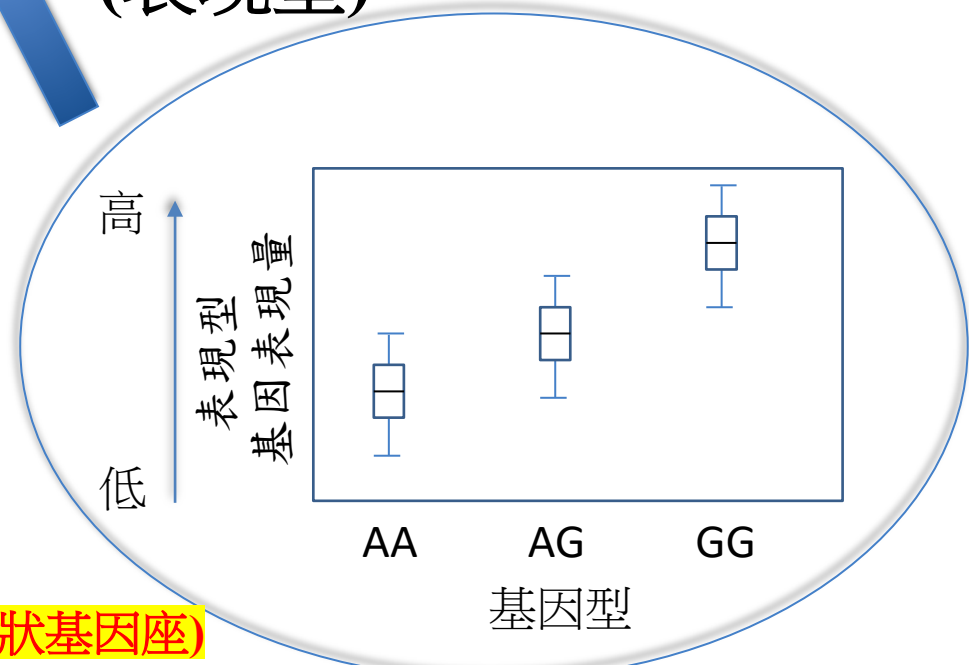
DNA、RNA和疾病間的關係

**GWAS: Genome-Wide
Association Study**
全基因體關聯性分析研究

ASD (疾病; 表現型)

基因型
(SNP)

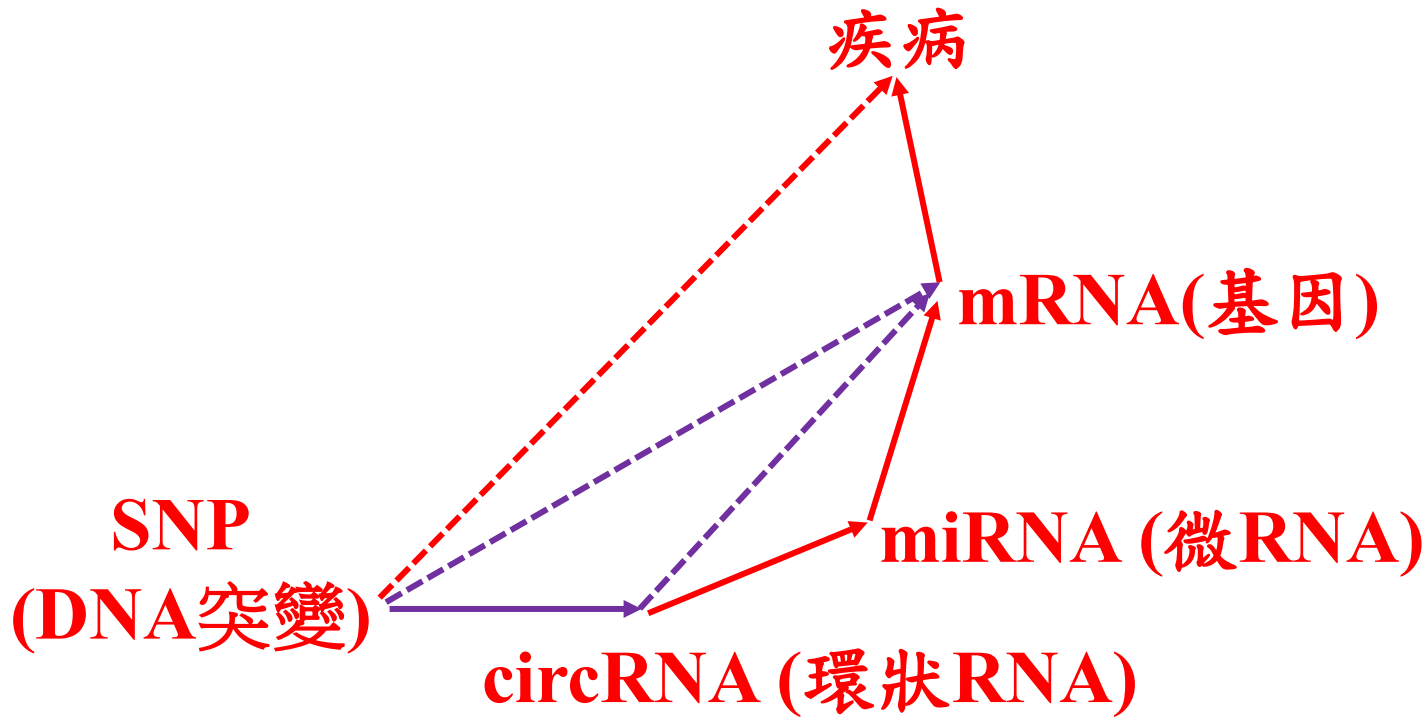
基因表現量
(表現型)

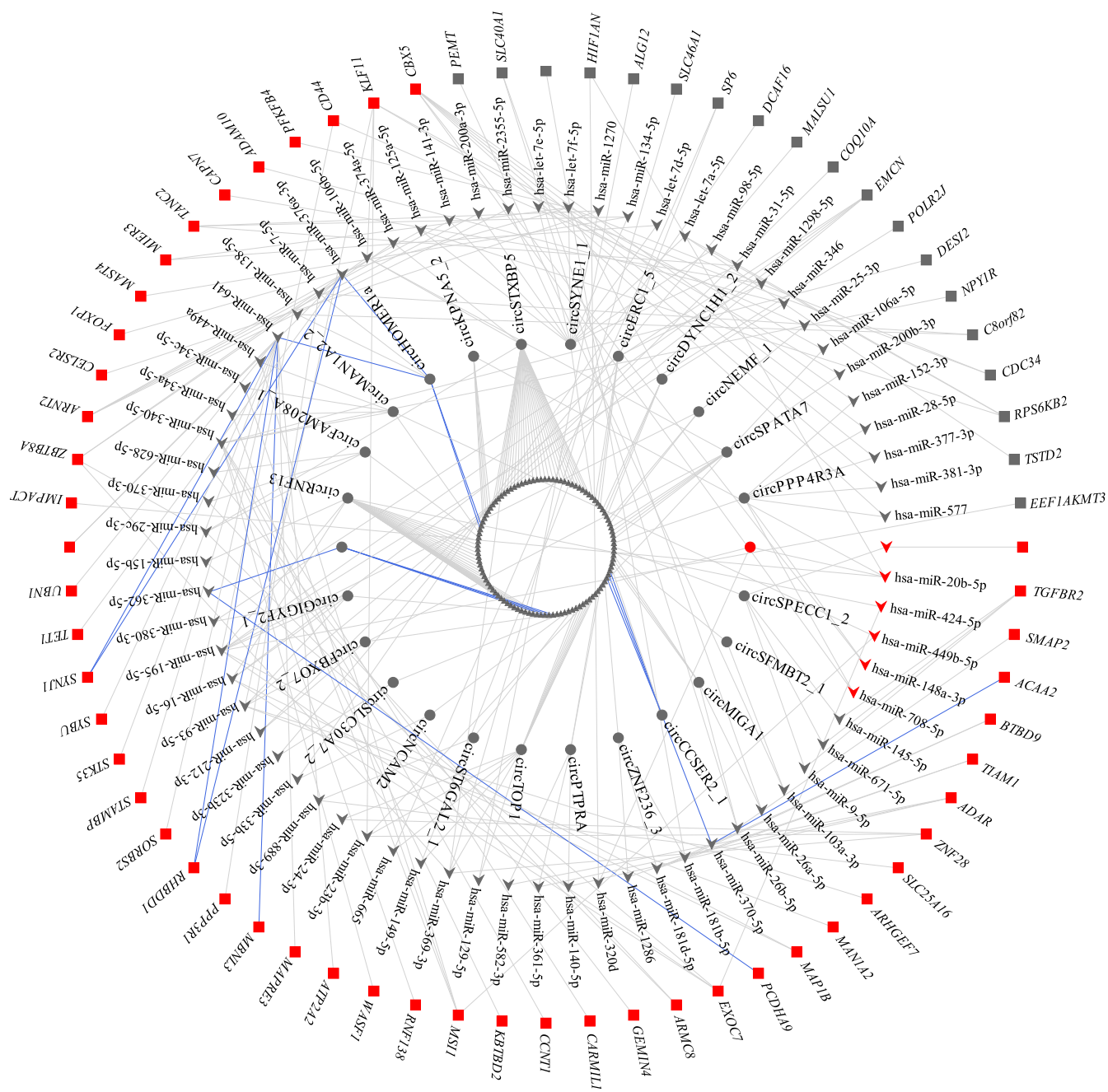


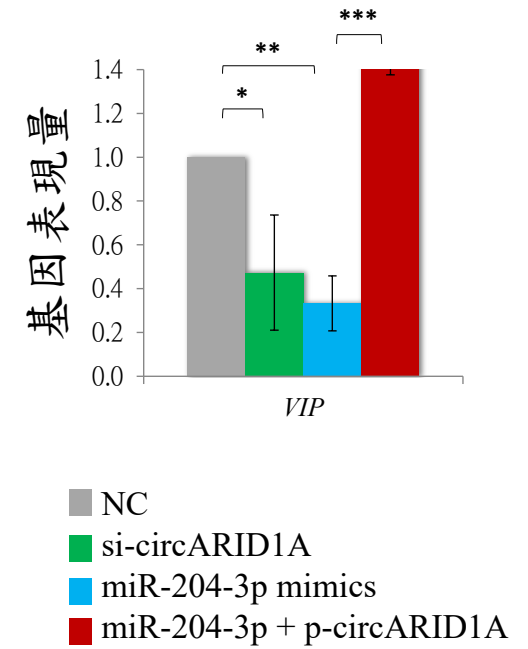
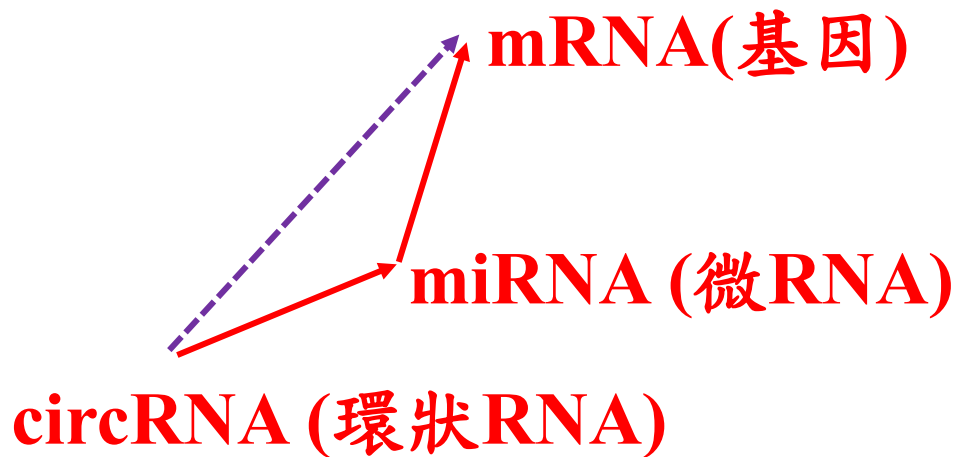
xQTL研究 (分子定量性狀基因座)

DNA、RNA和疾病間的關係

Causative Biology (因果生物學)







- 比較與演化基因體學（資訊統計）
- 調控因果生物學（資訊統計/分生實驗）
- 系統生物學（資訊統計/分生實驗）
- 神經精神疾病（分生實驗）
- 機器學習與多體學資料分析（資訊統計）

生活的樂趣
在於創造希望

Homepage:

<http://idv.sinica.edu.tw/trees/>



生物資訊與因果生物學實驗室 Bioinformatics & Bio-Causality (BBC)

大數據分析、演化、神經科學
Trees-Juen Chuang 莊樹諄 研究員/教授

不必痛的領悟
變映主果因：



Email: trees@gate.sinica.edu.tw

<https://www.genomics.sinica.edu.tw/tw/chuang-trees-juen->

August 29, 2024

