

國立中央大學生醫理工學院
生醫工程與數據科學工作坊

這幾年在系生所的一些工作

李弘謙

2019年9月5日
桃園市中壢區中大路300號
中央大學研究中心二期
R3-103室

- 今天只談我的實驗室分析基因組相關的工作
- 除了第一個議題之外，其它的都涉及分析與人類疾病相關的基因組

簡介基因組學

- 推薦大家看這個網頁（15分鐘）
<https://zh.wikipedia.org/wiki/基因組>

網頁的第一段話：在生物學中，一个生物体的基因組是指包含在該生物的DNA中的全部遺傳信息，又稱基因體 (genome)



A BRIEF GUIDE TO GENOMICS

基因組學簡介



脫氧核糖核酸 DNA

組蛋白

核小體

鹼基對

只有四種鹼基，以四個羅馬字母代表：A, C, G, T

基因組的（最重要的）遺傳資訊，由四個字母編成的一維序列代表

每個物種的每一個細胞裡都有一組相同的基因組

細菌的基因組序列長約數百萬；人類的長三十億

基因

染色體

細胞前期

細胞

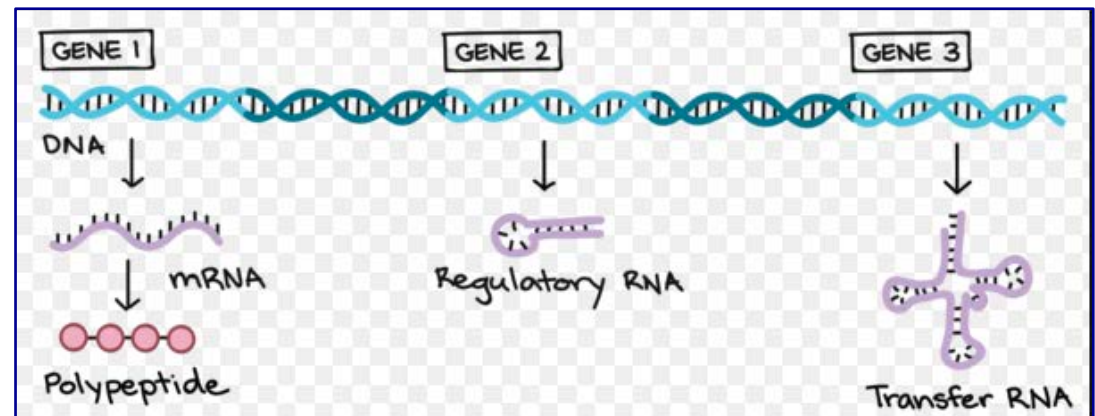


National Human Genome Research Institute

基因、蛋白質、奈米微機器

- 基因組鹼基序列裡編了很多碼，包括基因
- 部份的基因在適當的時機，依細胞的需要，經過被細胞中的機器讀、譯後成為兩大類機器：核糖體與蛋白質。
- 這些奈米微機器負責執行細胞內的各種工作，包括讀、譯基因。
- 人類的基因組有大約兩萬個基因，可以製造出大約一百萬種蛋白質

❖ 細胞與手機



細胞是全自動化的工廠

- 細胞好比一個很大的工廠，工廠裡的一切物件、設備細胞裡都有，並且一一對應：建築、機器、能源、原料、傳輸系統、工人、產品等等。
- 人類細胞體積大小可以差5個數量級。細胞直徑約5-150微米，但是它比我們最先進、最大的工廠複雜千、萬倍。
- 細胞和人類工廠最大的差別：細胞100%自我調控

蛋白質（與核糖體）是細胞裡的機器和工人

- 蛋白質製造：基因表達 > 訊息RNA > 蛋白質
- 不同的組織由不同的、有特定功能的細胞構成
- 有特定功能的細胞裡會有很多執行該功能的蛋白質群（以及維持細胞運作的基本蛋白質群）
- 細胞裡蛋白質群的多寡及活性反應細胞功能狀況
- 因此，細胞裡的 訊息RNA群 大致反應細胞功能

以測量細胞內的訊息RNA群內涵 探測細胞功能

- 測量訊息RNA（簡稱測序）比測量蛋白質容易許多
- 分析細胞內的訊息RNA群內涵（那一種、有多少），可以猜測細胞當下的功能。
- 對比健康細胞與疾病細胞的訊息RNA群，可以概估疾病細胞哪些功能正常/不正常
- 二十一世紀後測序科技快速發展，已經高度商業化

小結/重點

- 細胞內的訊息RNA群內涵反應細胞當下的功能、運作
- 快速測序科技已經高度發展及商業化
 - ❖ 建議搜尋：基因晶片、二代定序儀
- 由分析細胞測序數據獲得對細胞功能/失能的了解是一種生物資訊工作，也是一種前驅大數據分析
- 取得的結論是有數據支撐的猜測，最後還需要實驗驗證

簡介幾件工作

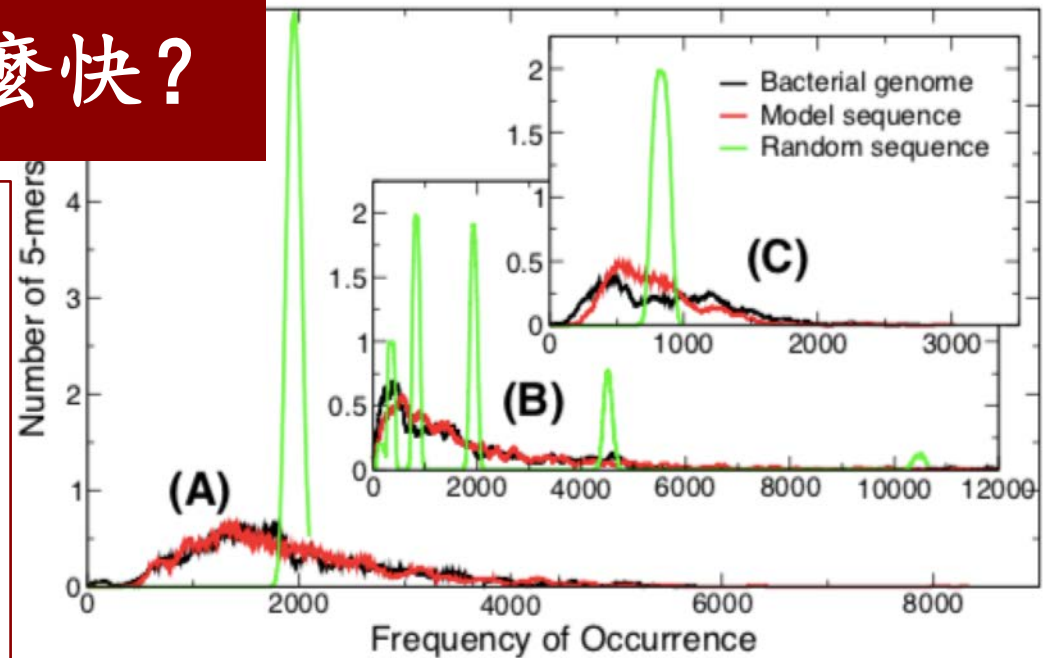
1. 生物演化為什麼那麼快？
2. 用疾病進展趨勢鑑定癌症靶基因
3. 舊藥新用、西藥中用
4. 鑑定 有生物活性複合物的藥性
5. 癌症桿細胞的基因型
6. miRDRN: 網路資料庫與工具 -
有疾病特异性的微RNA调控网络

物種基因組 與 生物演化

- 基因組決定物種的標準（平均）型態和生命模式
- 血緣逾相近的物種基因組逾相似
- 除生命起源之外，所有物種的演化過程儲藏在其基因組裡
- 基因組演化的動力：隨機突變、自然選擇|淘汰
- 地球成於45億年前，38億年前細菌出現，9億年前多細胞生物出現
- 問：從簡單的化學分子到細菌再到多細胞，依靠隨機突變的基因組演化為什麼這麼快？

基因組演化為什麼這麼快？

- 基因組演化來自隨機突變、自然選擇|淘汰
- 我們預期基因組-四個字母的序列-有很強的隨機性
- 事實不然



Genome **Biology** 2004, 5:P7

- 基本自然現象：大隨機系統的分佈曲線很窄。分佈曲線與 $\sqrt{N}$ 成反比； N 為系統大小
- 細菌基因組， $N \sim$ 數百萬；動物基因組， $N \sim$ 數、十億
- 但是基因組的分佈曲線像一個 $N \sim$ 數千的系統
- 自同一個基因組隨意截取兩長段，內涵相似度很高
- 考慮：將一條 $N \sim$ 五千的隨機序列複製一千或十萬次，它的分佈曲線維持 $N \sim$ 五千的特性

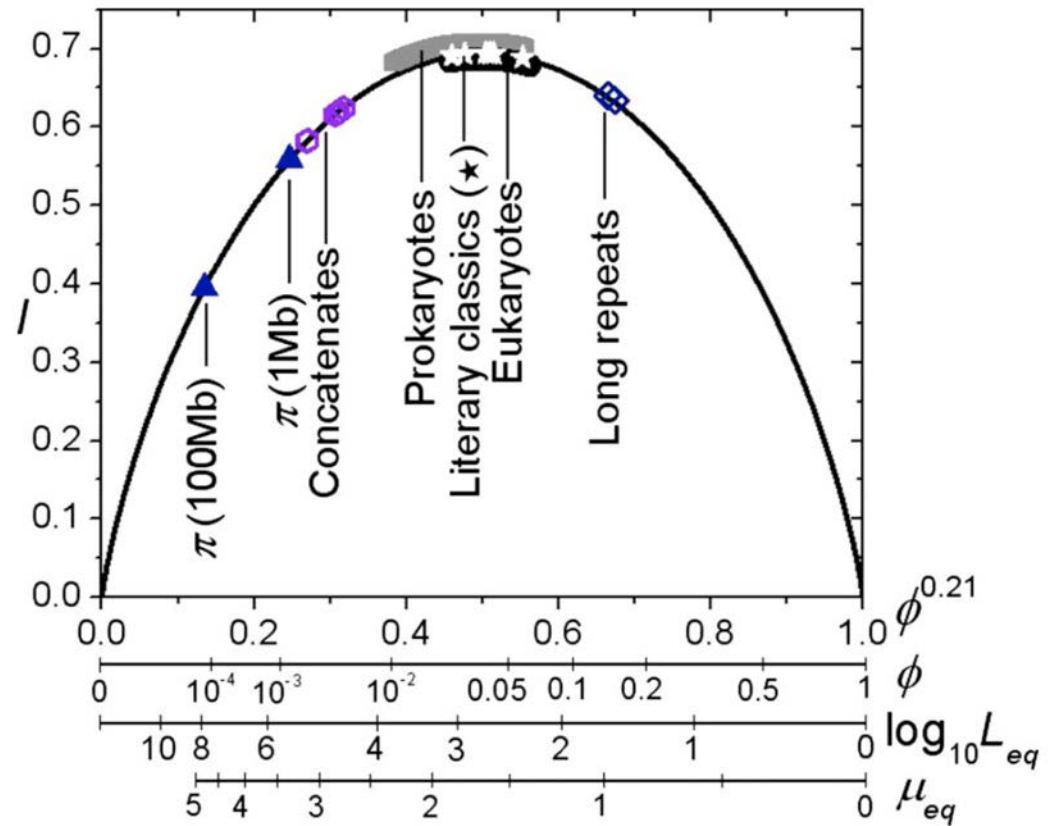
資訊載體的亂度與(次)序

PHYSICAL REVIEW E 79, 061911 (2009)

- 討論一維的資訊載體，如一個二元文本
- 兩個極端，完全有序/隨機，都不是有效的資訊載體
- 高質量的資訊載體的序/無序特徵是什麼？
- 基因組是經由進化演繹出的「最佳化」資訊載體，它有什麼特徵？

考慮：

- 高序文本可以經由隨機點突變增加亂度，把它向無序、隨機方向推
- 重複一個文本：資訊、序、亂度不變，長度增加一倍



基因組特徵

- 細菌基因組長度約百萬至千萬鹼基，它的亂度等價於將完全有序的序列每個位點平均點突變1.8次
- 或等價於將長約數千鹼基的隨機序列複製一千至一萬次

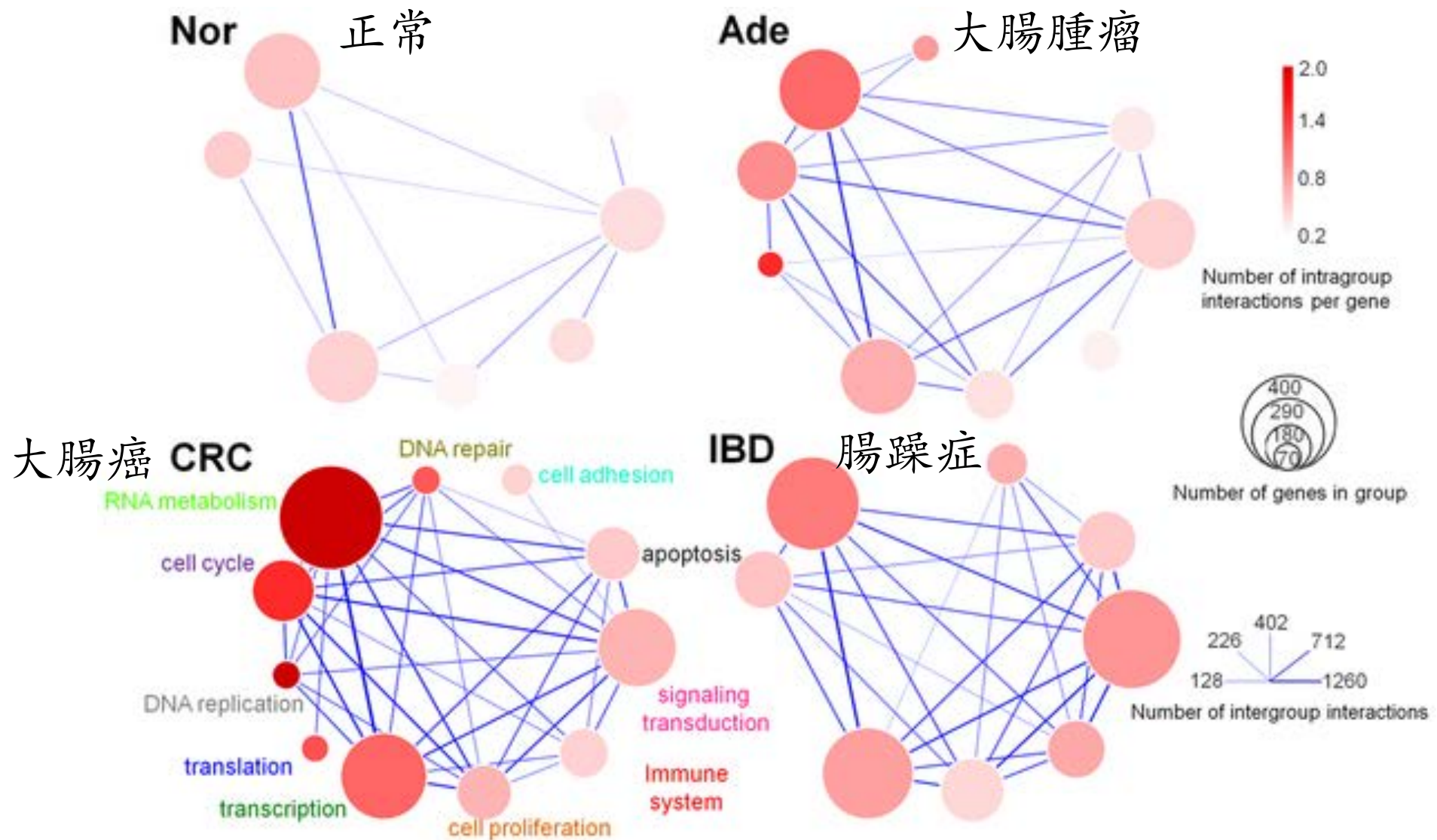
基因組告訴我們：（自我）抄襲 是累積資訊最有效的方法

- 隨機片段複製是基因組生長的主要模式-為什麼？
- 機制隨機（「盲製錶者」R. Dawkins), 而片段複製是增加資訊最有效的生長模式
 - 有最大的不被淘汰的機率
 - 指數性的加快了演化速度
- 因為是片段複製，基因組各處的內涵相似度很高，所以分佈曲線類似一個小系統

用疾病進展趨勢鑑定癌症基因

- 抗癌最有效的方法是儘早發現
- 在細胞層次，癌症是「過動」、「燥熱」型疾病：許多正常情況下無活性或活性底的功能/基因，癌症時活性會增加。
- 問題：在癌病症不明顯時，如何用活性上升的功能/基因預測癌症？
- 一個可能的方法：利用疾病進展趨勢中功能/基因活性變化

Function-function networks 功能-功能網絡.

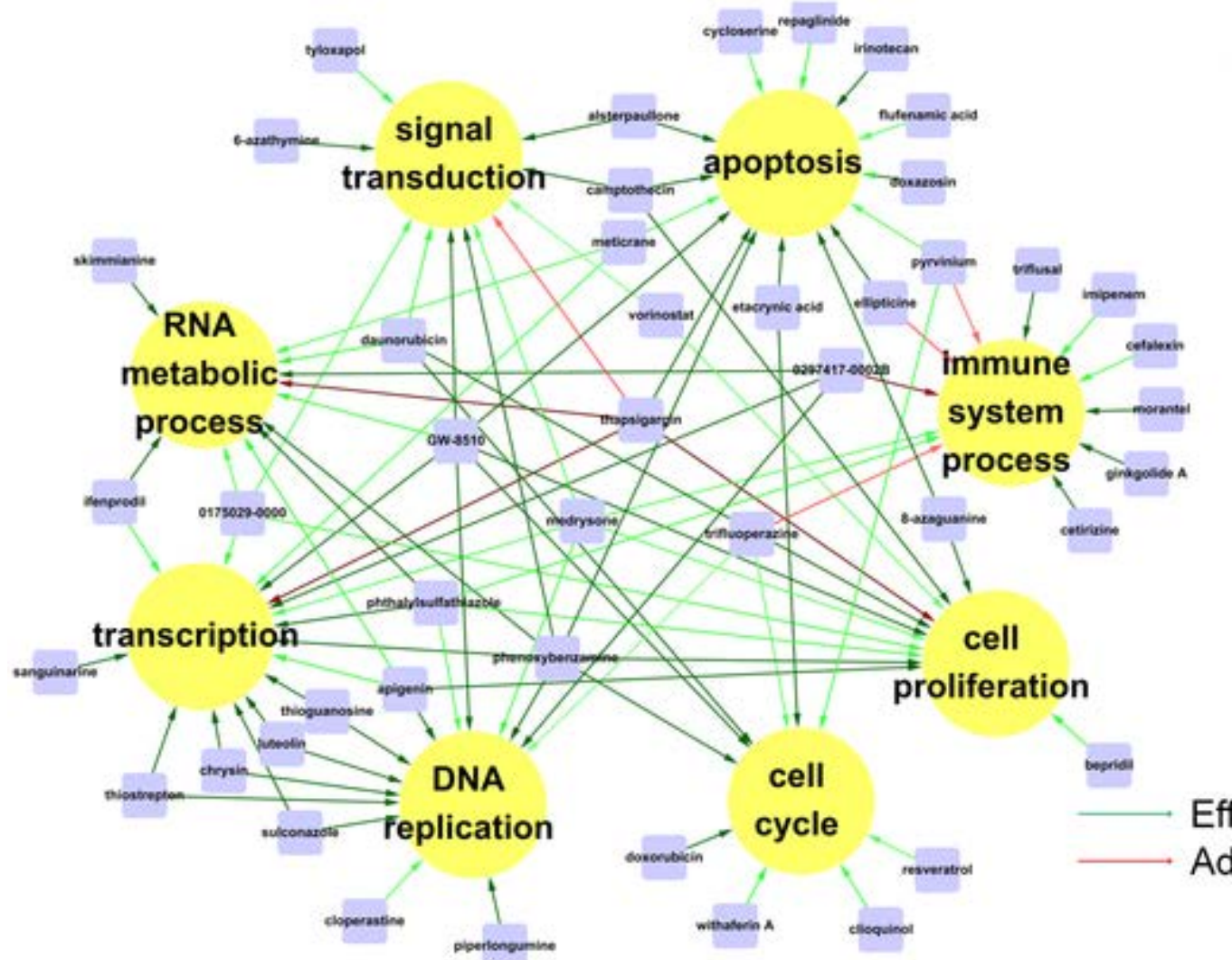


Chung FH, Lee HHC, Lee HC (2013) ToP: A Trend-of-Disease-Progression Procedure Works Well for Identifying Cancer Genes from Multi-State Cohort Gene Expression Data for Human Colorectal Cancer. PLOS ONE 8(6): e65683.
<https://doi.org/10.1371/journal.pone.0065683>
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0065683>

舊藥新用、西藥中用

- 針對治療某種功能障礙研發的藥物可能對它種功能障礙也有效；有些藥因不良副作用而被禁用
- 新藥物/功能關係可以經由對比藥物/疾病對基因組的影響認定
 - 有效的藥：使 失效/過度活化 的功能 活化/被抑制
 - 有害的藥：與以上相反
- 西醫理念 - 針對性。針對一個基因或一種功能治療 | 有效的西藥常常因不良副作用被禁用
- 中醫理念 - 整體性。強調多種功能的平衡、協調

舊藥新用、西藥中用 | 大腸腫瘤功能-藥物關聯圖與 Function-drug association map (FDAM) for colorectal adenoma



黃色圓形, 有障礙功能
紫色方形, 藥物
綠色箭頭對功能有效
紅色箭頭對功能有害

舊藥新用、西藥中用
認定一組藥物,
總體對所有障礙功能有效, 且不對任何功能有害

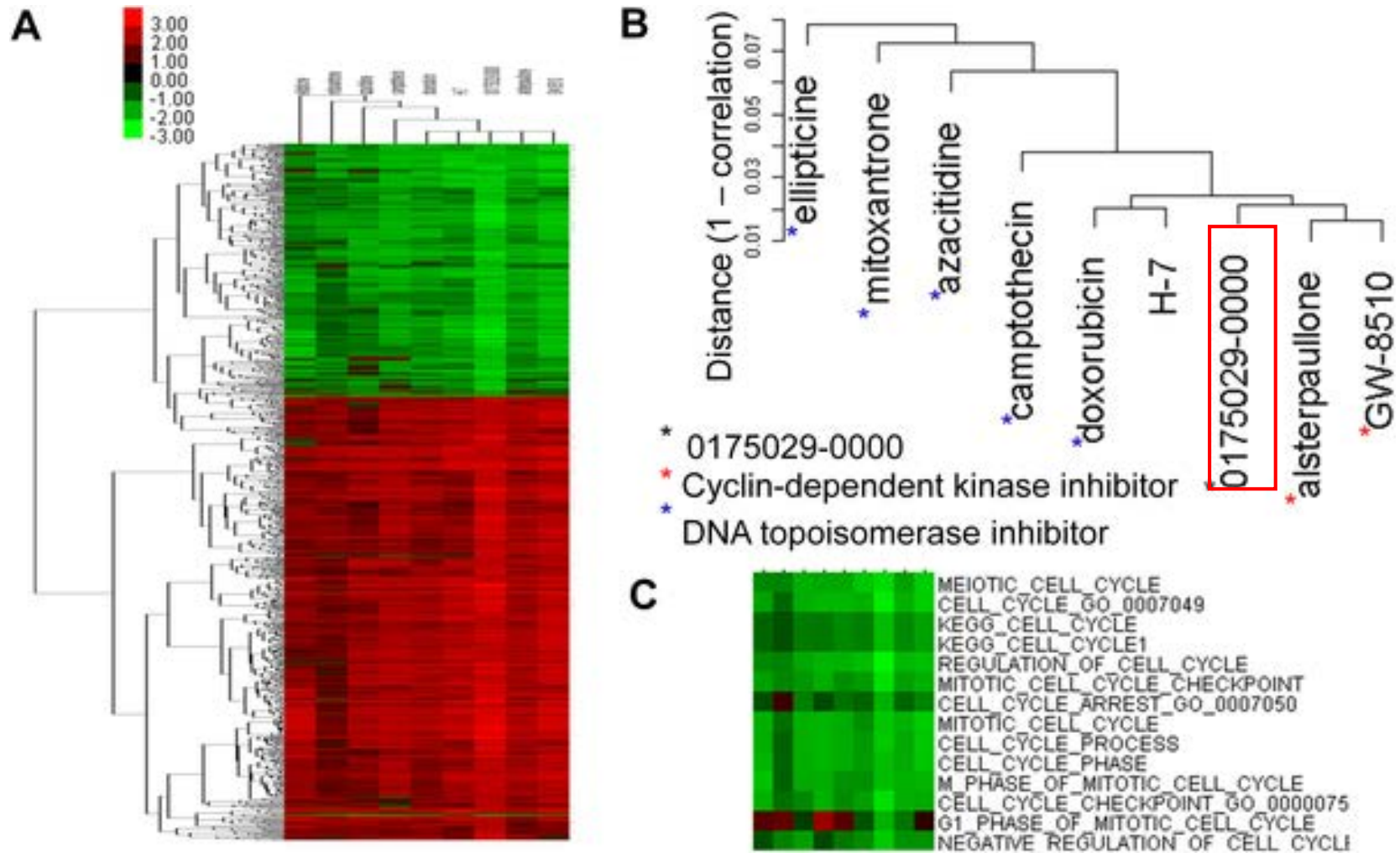
Chung FH, Chiang YR, Tseng AL, Sung YC, Lu J, et al. (2014) Functional Module Connectivity Map (FMCMap): A Framework for Searching Repurposed Drug Compounds for Systems Treatment of Cancer and an Application to Colorectal Adenocarcinoma. PLOS ONE 9(1): e86299. <https://doi.org/10.1371/journal.pone.0086299>
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0086299>

鑑定 有生物活性複合物的藥性

- 哈佛-麻省理工聯合研發的「連結圖輿」Connectivity Map (簡稱 Cmap) 有許多藥物和活性複合物的基因組譜，其中106 種活性複合物的藥性不明
- 我們用「基於基因集的局部分層聚類法」鑑定了其中18 種活性複合物非常可能具有的藥性。

The small molecule 0175029-0000 is identified as a novel cyclin-dependent kinase inhibitor (CDKi)

小分子0175029-0000 被鑑定為一種依賴細胞周期蛋白 激酶抑制劑



Chung FH, Jin ZH, Hsu TT, Hsu CL, Liu HC, et al. (2015) Gene-Set Local Hierarchical Clustering (GSLHC)—A Gene Set-Based Approach for Characterizing Bioactive Compounds in Terms of Biological Functional Groups. PLOS ONE 10(10): e0139889.

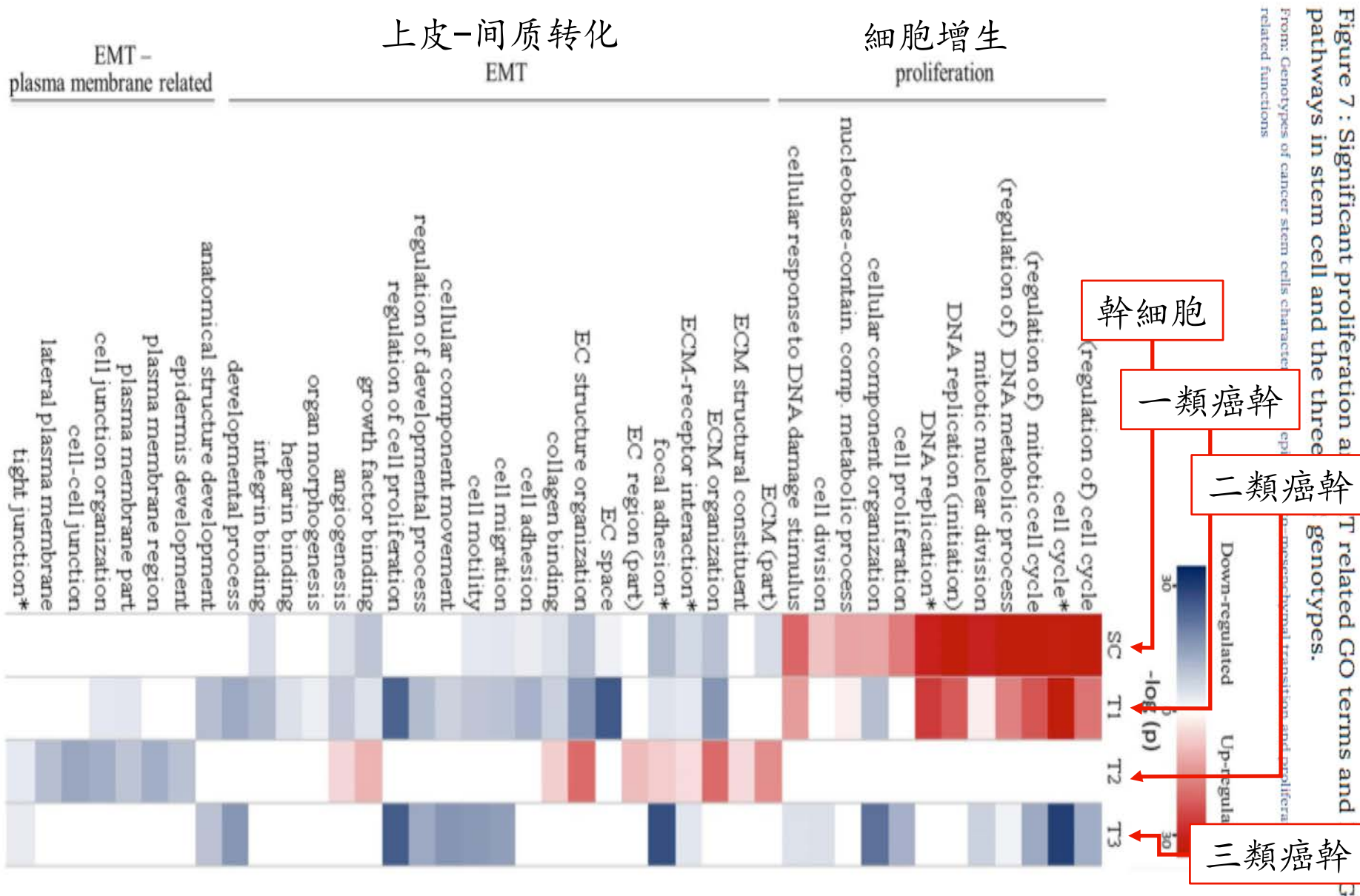
<https://doi.org/10.1371/journal.pone.0139889>

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0139889>

癌症幹細胞的基因型（一）

- 有一派學說認為癌症起因與復發由癌幹細胞主導
 - 癌幹細胞特性：自我更新、腫瘤啟動、侵入性、轉移性、抗治療性
 - 治療癌症需要把致癌的癌幹細胞殺除
- 癌幹細胞的基因型是什麼？
- 對卵巢癌、前列腺癌、神經膠質瘤、乳癌、肺癌、大腸癌等癌幹細胞以及數種（非癌）幹細胞做功能基因組對比分析

Figure 7 : Significant proliferation a pathways in stem cell and the three
From: Genotypes of cancer stem cells character related functions



癌症幹細胞的基因型（二）

- 癌幹細胞有三種基因型，主要由上皮-間質轉化（EMT）和細胞增生兩類功能的活化程度決定
 - 一類癌幹細胞-神經膠質瘤、肺癌：細胞增生高度活化、EMT中度抑制
 - 二類癌幹細胞-乳癌：EMT高度活化
 - 三類癌幹細胞-卵巢癌、前列腺癌、大腸癌：細胞增生及EMT都中度抑制
- 用途：藥物治療癌症的有用資訊。比如說：
 - 治療神經膠質瘤和肺癌不建議用激化細胞增生的藥
 - 治療乳癌不建議用激化 EMT 的藥

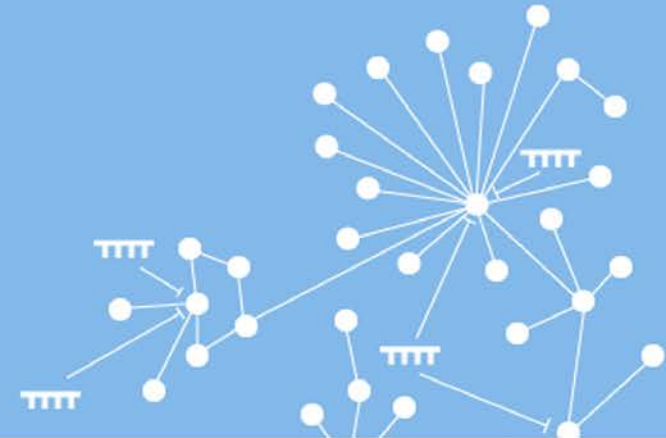
miRDRN:網路資料庫與工具- 有疾病特异性的微RNA调控网络

- 細胞中的生物功能由特定的基因/蛋白質群執行
- 蛋白質形成一個有層次調控網絡，上游的蛋白質調控下游的
- 調控多個下游蛋白質的稱軸蛋白質
- 軸蛋白質往往被微RNA (miRNA) 調控
- 微RNA與蛋白質的關係可以一對多或多對一
- 特定的疾病由特定的功能障礙，因此也有特定的相關蛋白質群和微RNA群
- miRDRN 是提供探索疾病|微RNA群|蛋白質調控網絡的資料庫和網上工作平台
<http://mirdrn.ncu.edu.tw/mirdrn/>

miRDRN (miRNA Disease Regulatory Network) is a useful platform for discovering potential novel disease networks and comorbidity genes by disease-associated miRNAs targeting pathways. Meanwhile, we provide a powerful visualization tool to present information.

We offer all possible sub-pathways from miRNA target protein by using protein-protein interactions. Our database integrated various data sources such as Disease-associated miRNAs, miRNA-associated targeted gene, Tumor-associated gene, Gene annotation and expression data.

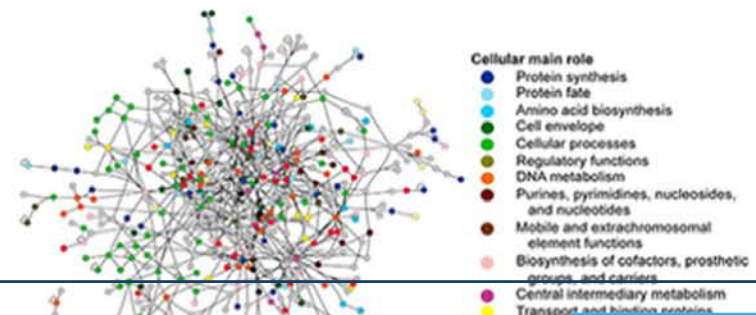
version 1.0



Statistics

Protein Entries	15,901
Protein-Protein Interactions	165,502
Disease Association	380
KEGG Pathways	274
miRNA Entries	932
siRNA Entries	11,221

Visualization



✓ PEER-REVIEWED Bioinformatics and Genomics section >

miRDRN—miRNA disease regulatory network: a tool for exploring disease and tissue-specific microRNA regulatory networks

Hsueh-Chuan Liu¹, Yi-Shian Peng¹, Hoong-Chien Lee^{1,2}

August 6, 2019

PeerJ
Life & Environment

The Journal of Life and Environmental Sciences

rights reserved. 

miRDRN 一個小應用：為什麼過去十年抗阿茲海默的抑制全軍覆沒？

- 阿茲海默的主要致病假設： β 型類澱粉蛋白的聚合和累積致神經細胞死亡繼而導致大腦失能
- 抗阿茲海默的兩個策略
 1. 抑制BACE1基因打斷 β 型類澱粉蛋白的製造
 2. 免疫治療-用抗體清除 β 型類澱粉蛋白
- 過去近十年所有7-8種抗阿茲海默藥在人體試驗階段全部失敗
- BACE1藥試驗組病患的認知能力比控制組的能力更差

信息傳遞關鍵基因 🤔

BACE1: 藥物抑制標靶基因
調控下游的 NCSTN,
PSEN1, MMP2, RANBP9,
PLSCR1 等基因

製造類澱粉蛋白關鍵基因 😊

Figure 12: A miRNA regulatory sub-network centered on *BACE1*.

The genes *PSEN1*, *NCSTN*, *RANBP9*, *PLSCR1*, *MMP2* are immediately downstream to, that is, have level 1 PF

miRDRN—miRNA disease regulatory network: a tool for exploring disease and tissue-specific microRNA regulatory networks

Hsueh-Chuan Liu¹, Yi-Shian Peng¹, Hoong-Chien Lee^{1,2}
August 6, 2019

感謝 與這些工作的第一作者前博 生（現在的博士）

-  陳鴻大、 康星源、 范文郎 –
基因組演化
-  鍾豐翔 – 致癌基因認定、舊藥新用/.
西藥中用、藥性認定、癌幹細胞
-  許爵麟 – 癌幹細胞
-  劉學銓 – miRDRN 資料庫和網上平台
-  彭奕憲 – 阿茲海默症

謝謝大家

祝大家在生醫系有快樂的時光