

科學實驗寫作工作坊： 實驗設計與數據分析研究

阮雪芬

臺灣大學生命科學系

Sep. 12, 2013

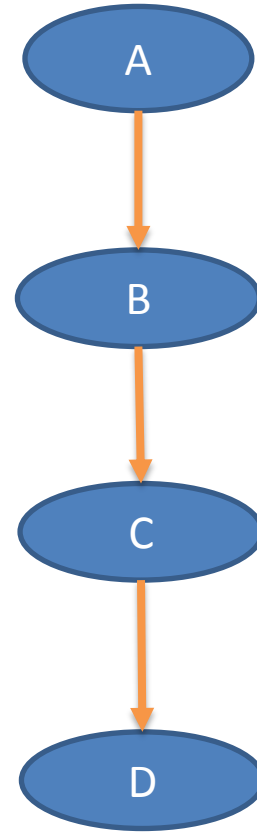
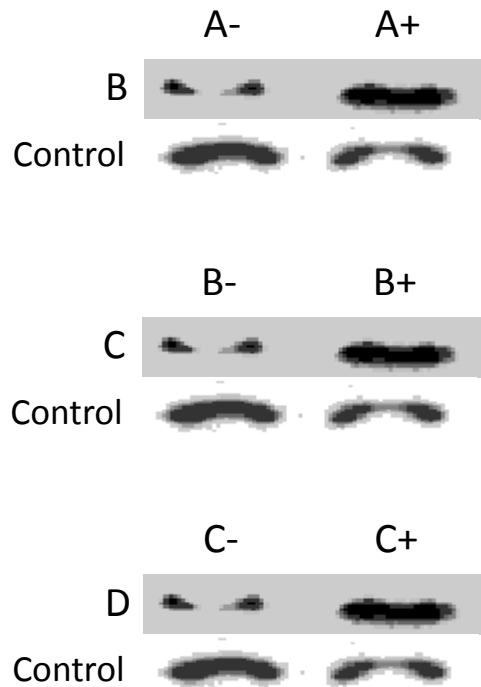
主題

- 實驗設計
- 數據分析

實驗設計

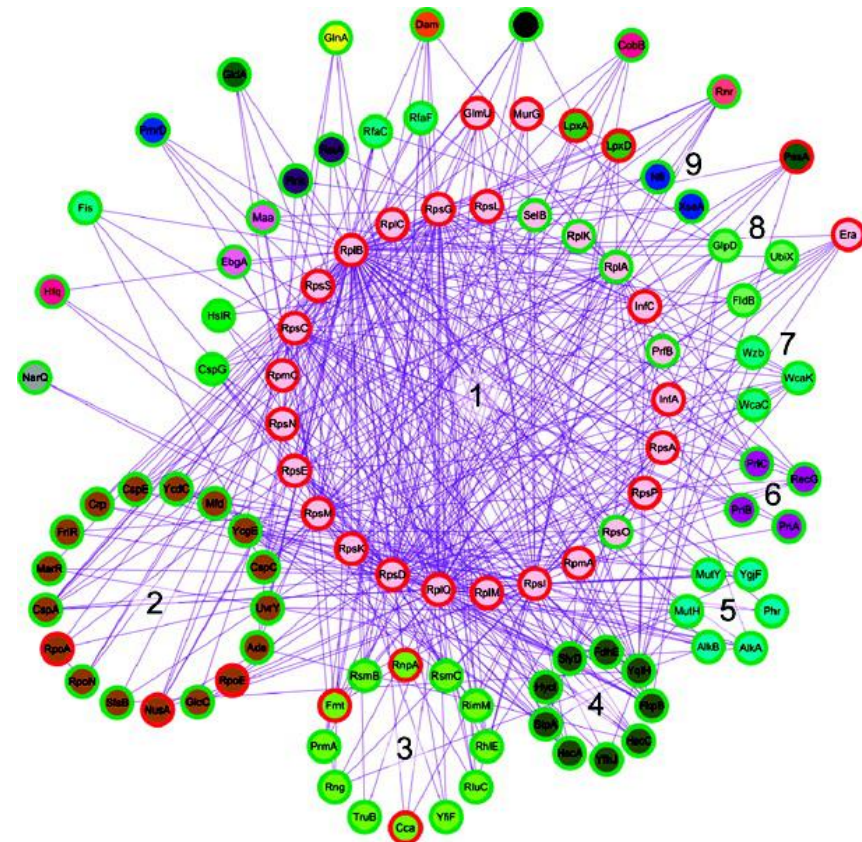
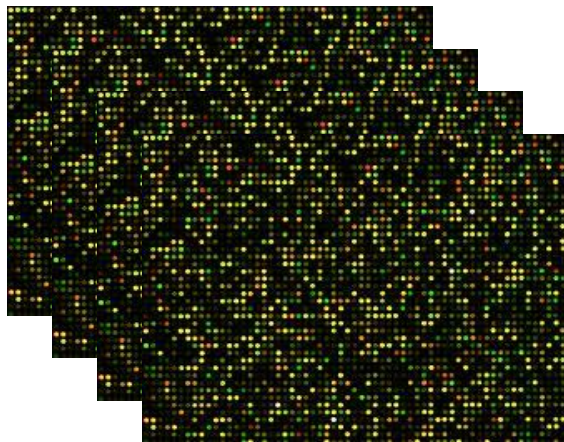
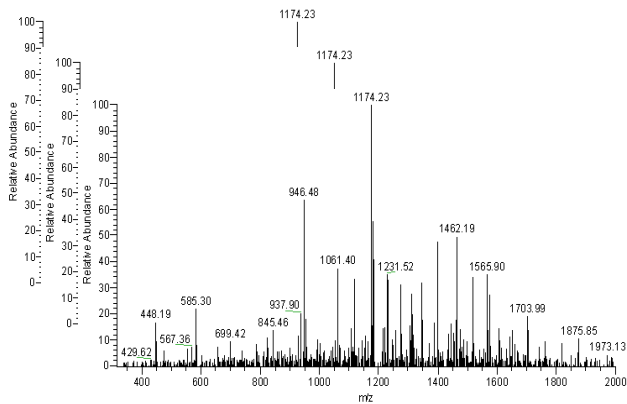
- 假設驅動的實驗 (hypothesis-driven experiment)
- 數據驅動的實驗 (data-driven experiment)

假設驅動的實驗



數據驅動的實驗

高通量實驗 (High-throughput experiments)



假設驅動的實驗

實驗設計

- 控制組
 - 一個標準品用來作為比較
 - 除了要研究的變因外, 其餘皆與實驗組相同
- 試驗組
 - 要進行測試的組 (一次只改變一個變因)
- 取樣誤差 (Sampling error)
 - 不好的取樣會扭曲結果
 - 藉由大量取樣來降低誤差

假設驅動的實驗範例: 生物治療實驗

是否我們可以利用病毒(噬菌體)來攻擊細菌以
對抗細菌感染?

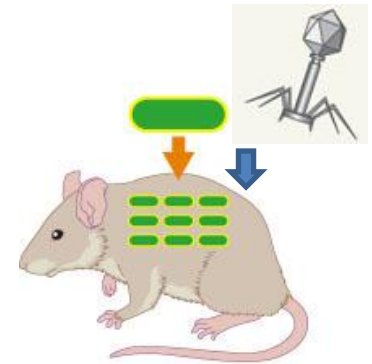
實驗一

- 假說: 噬菌體可以保護老鼠對抗細菌感染
- 預測: 如果用噬菌體先感染老鼠, 則感染細菌的老鼠將不會死

實驗一：測試

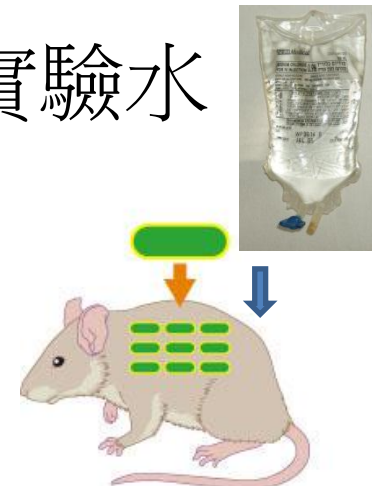
- 試驗組

老鼠感染細菌和噬菌體



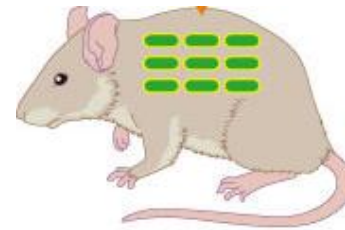
- 控制組

老鼠感染細菌和生理實驗水



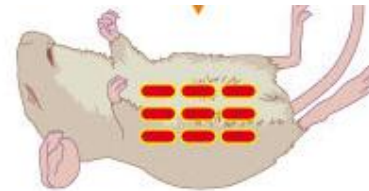
實驗一 結果和結論

- 試驗組
所有老鼠皆活著



© Brooks/Cole, Cengage Learning

- 控制組
所有老鼠皆死了

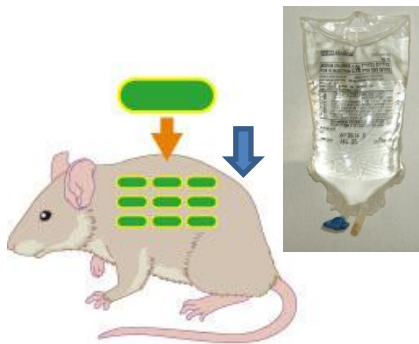


- 結論: **感染噬菌體將會保護老鼠避免被細菌
感染**

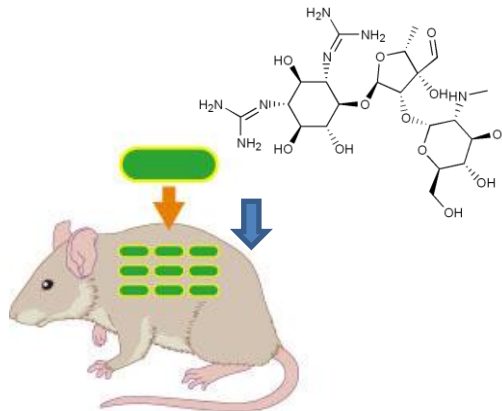
實驗二

- 預測: 噬菌體注射將比單一濃度的抗生素鏈黴素有效
- 測試: 老鼠注射細菌後, 將分三組分別注射

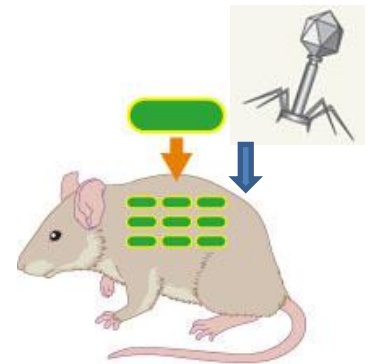
生理食鹽水



鏈黴素



噬菌體



實驗二： 結果

- 注射：
 - 噬菌體 - 12隻老鼠有11隻活著 (11/12)
 - 60 mg/gm 鏈黴素 - 12隻老鼠有5隻活著 (5/12)
 - 100 mg/gm鏈黴素- 12隻老鼠有3隻活著 (3/12)
 - 生理食鹽水 - 所有老鼠都死了
- 結論 - 噬菌體處理比抗生素有效

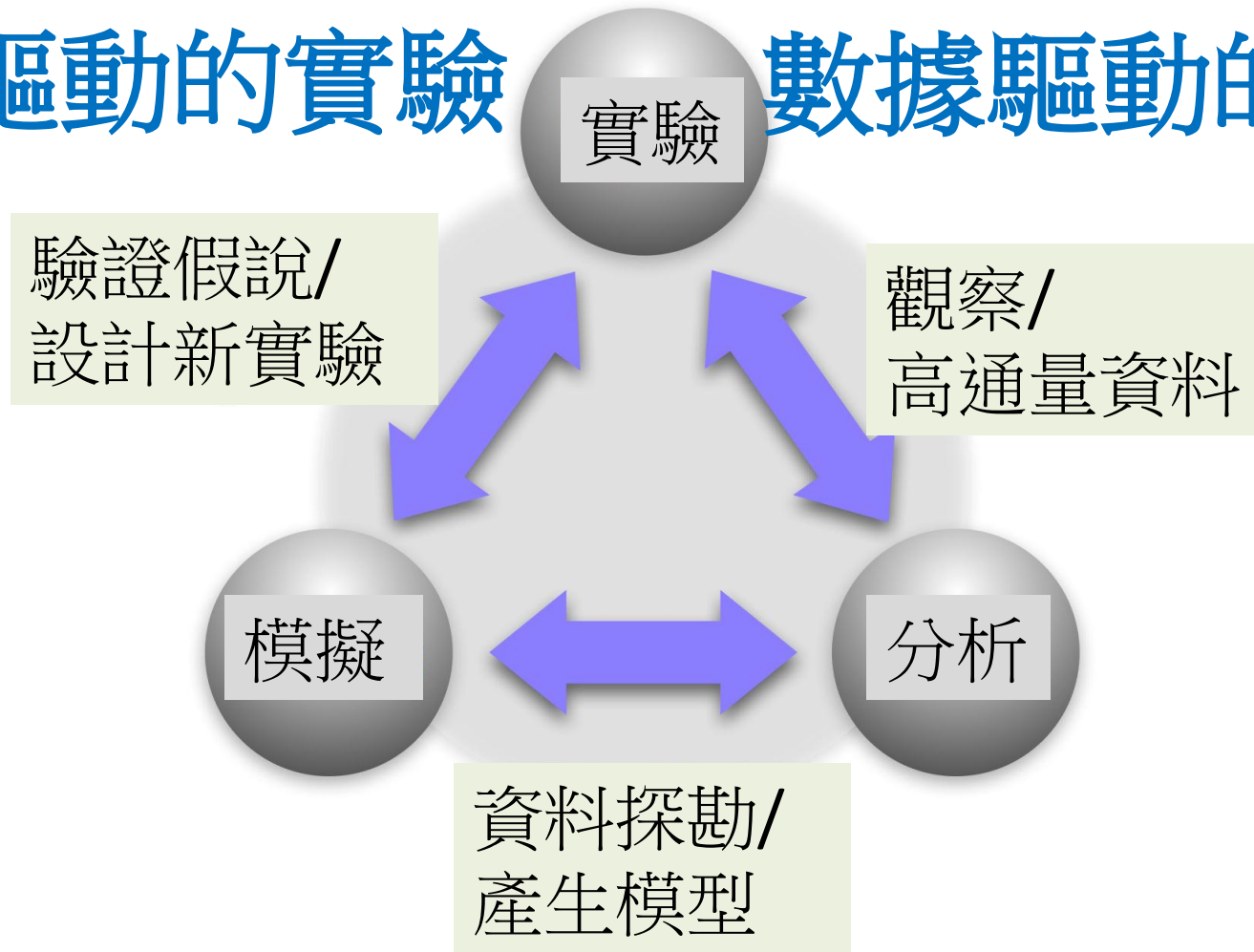
降低變因

- 所有老鼠接式相同年齡和性別, 以及相同培養方式
- 在每一個測試組中皆接受完全相同的處理
- 在控制組的老鼠皆接受相同量的生理實驗水
- 測試的變因有抗生素處理和噬菌體處理

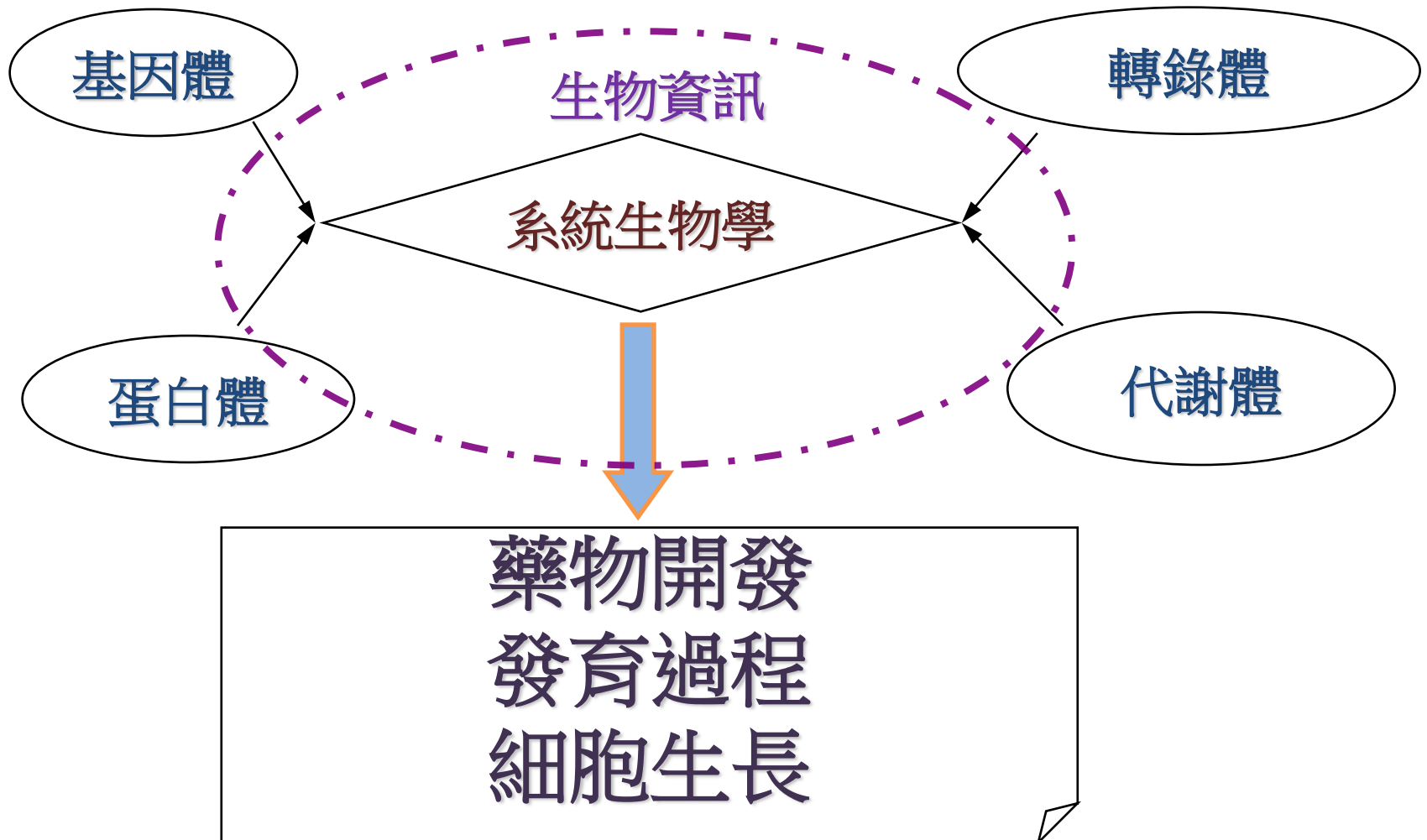
數據驅動的實驗

新興生物學: 系統生物學

假設驅動的實驗 數據驅動的實驗

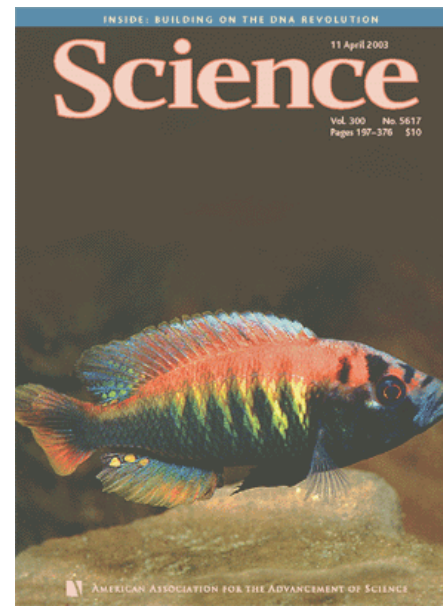
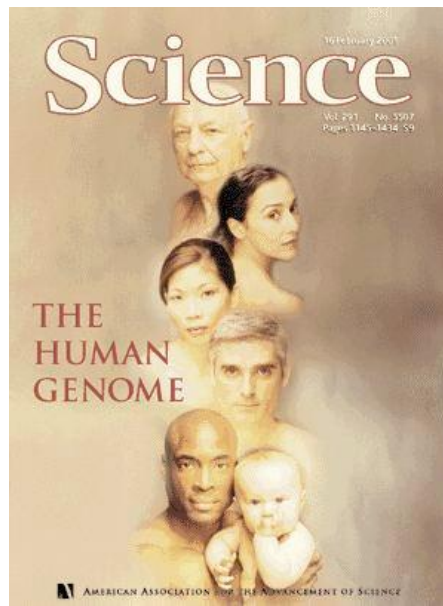
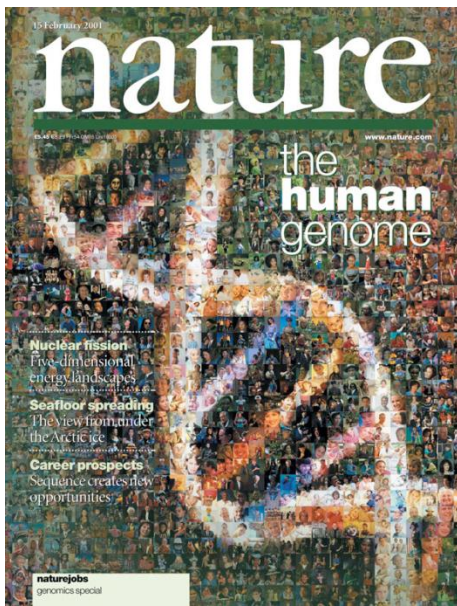


系統生物學 & 各種體學 (Omics)



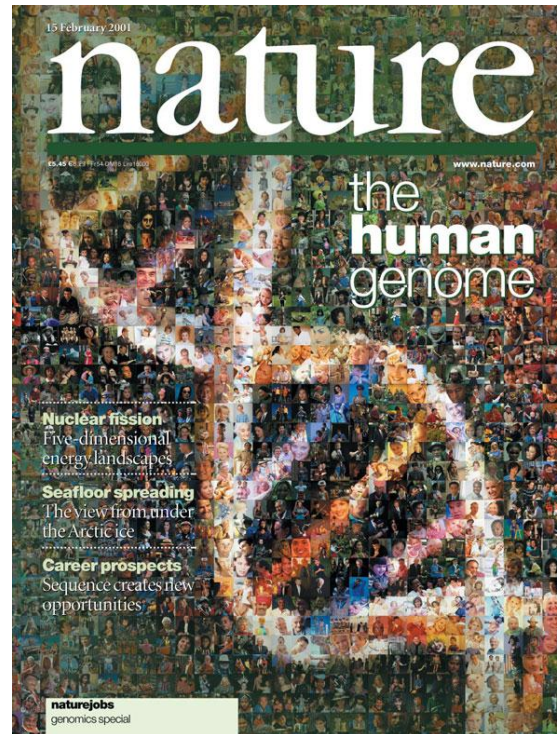
系統生物學與基因體

- 基因體圖譜就像**遊戲規則**，而系統生物學則是**球賽**
 - by Alexandra Stikeman, Technology Review, 2002 March

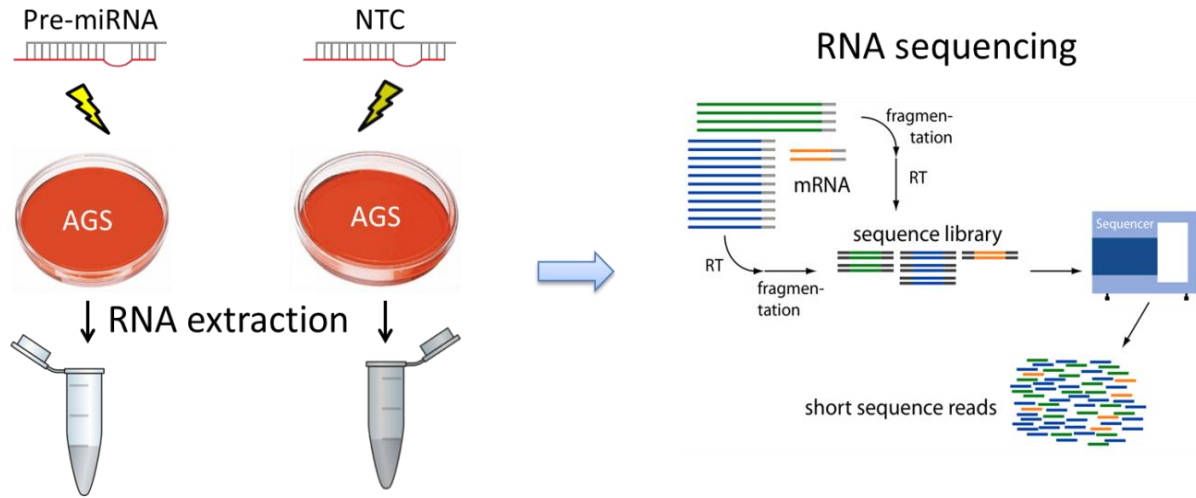


人類基因體計畫

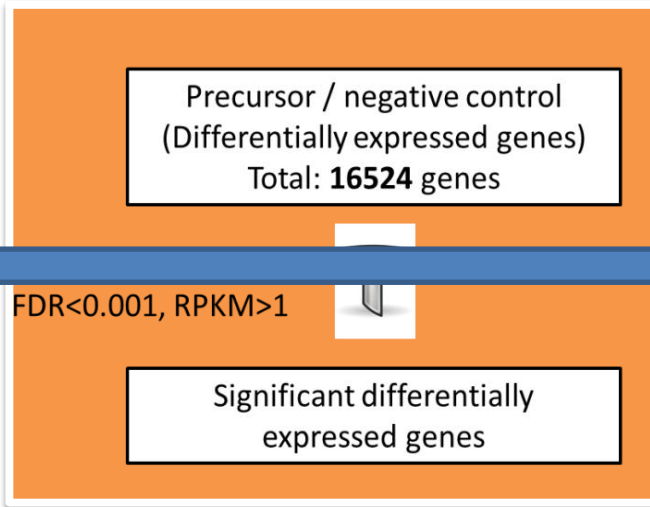
- 利用定序儀在2001年完成了第一份人類基因體草圖
- 在2003年完全解碼



基因體實驗設計



獲得大量的資料



進行數據分析

蛋白質體實驗設計範例一

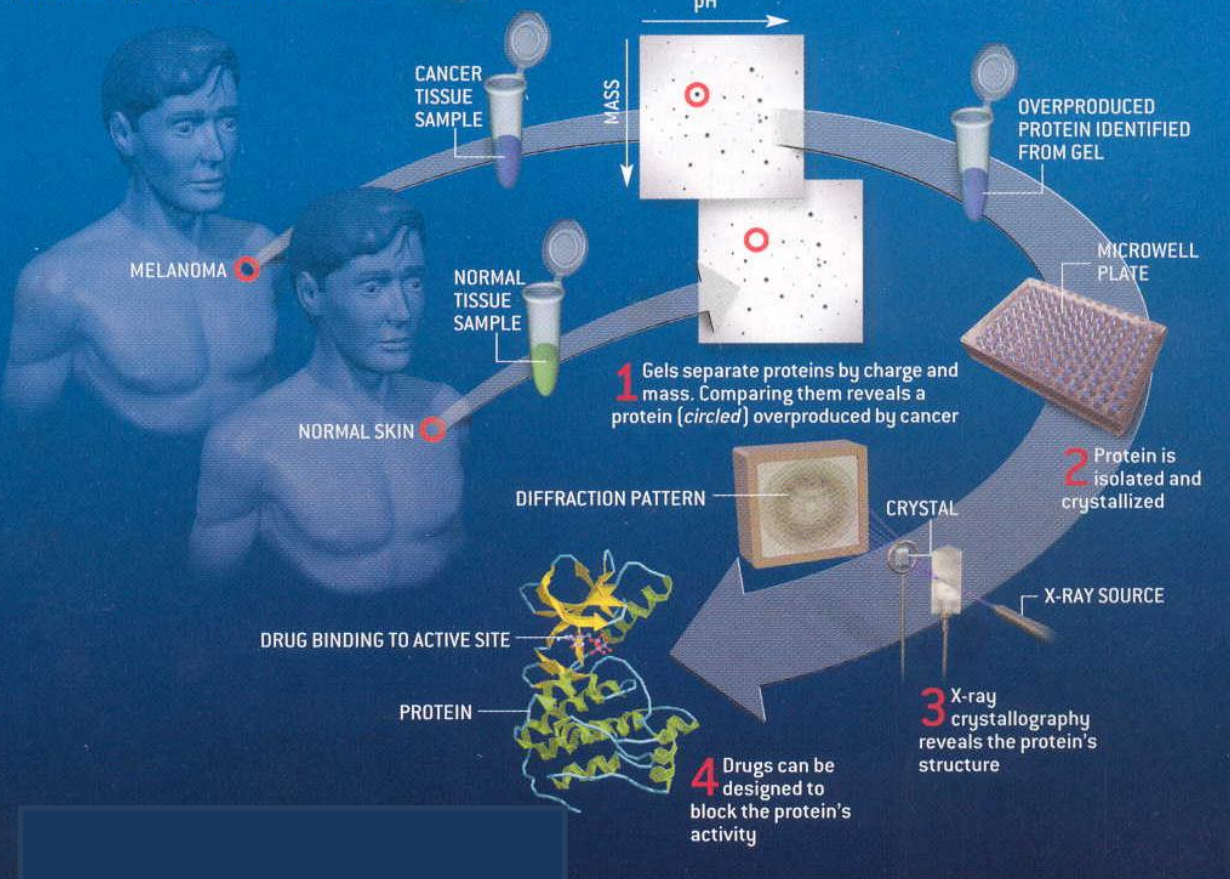
比較癌症病人的正常組織和癌組織

獲得大量的資料



進行數據分析

FINDING NEW DRUG TARGETS
(Here, devising a drug to kill the skin cancer melanoma)



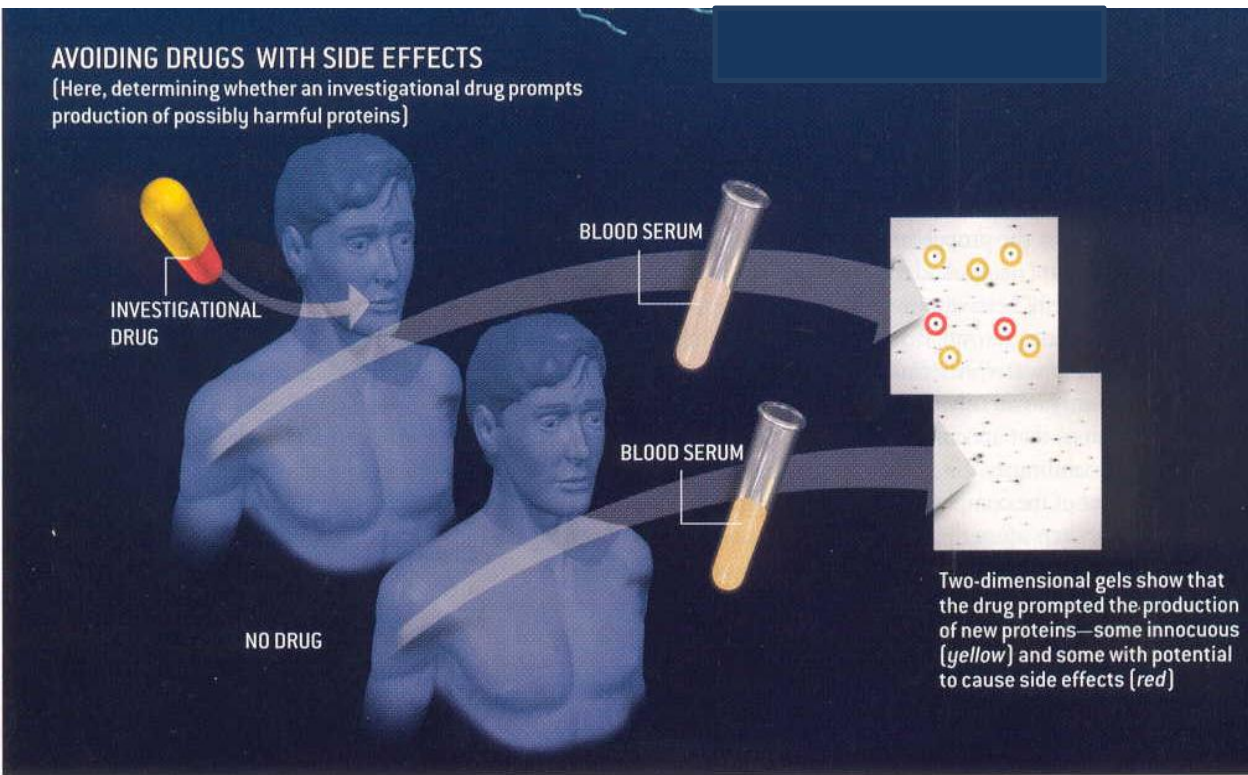
蛋白質體實驗範例二

比較投藥後的人和不投藥的人之血清

獲得大量的資料

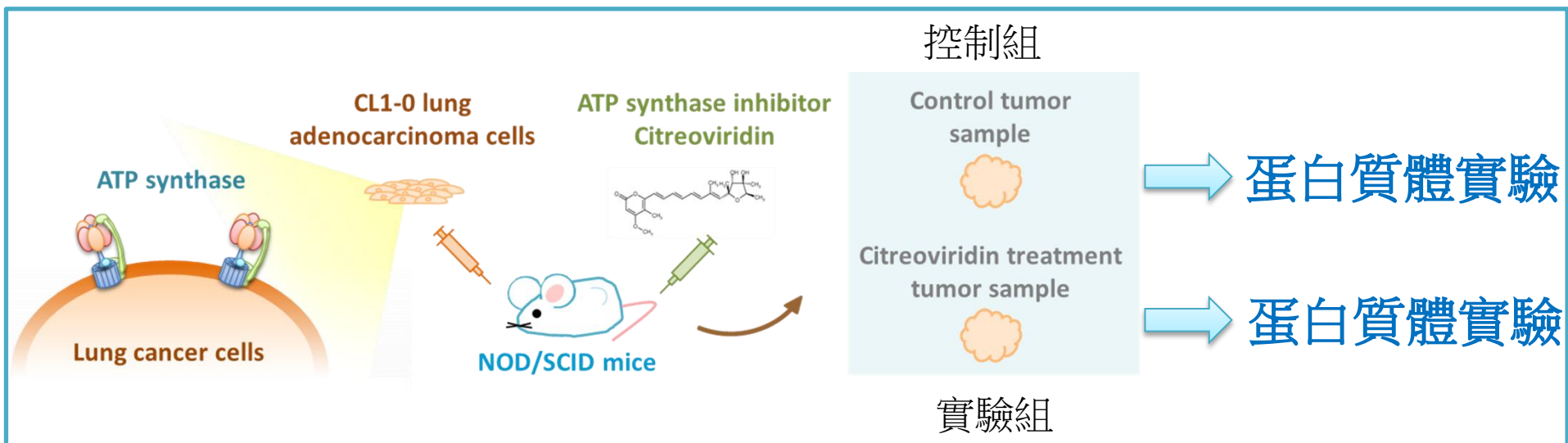


進行數據分析



蛋白質體實驗範例三

研究藥物對於老鼠腫瘤抑制是經由改變哪些蛋白表現



獲得大量的資料



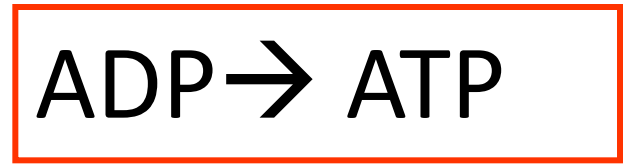
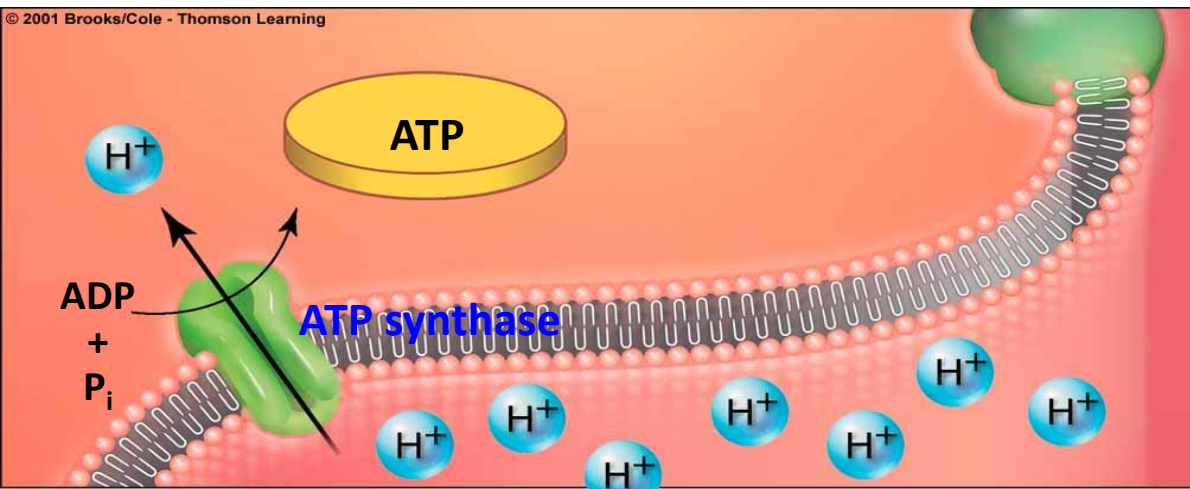
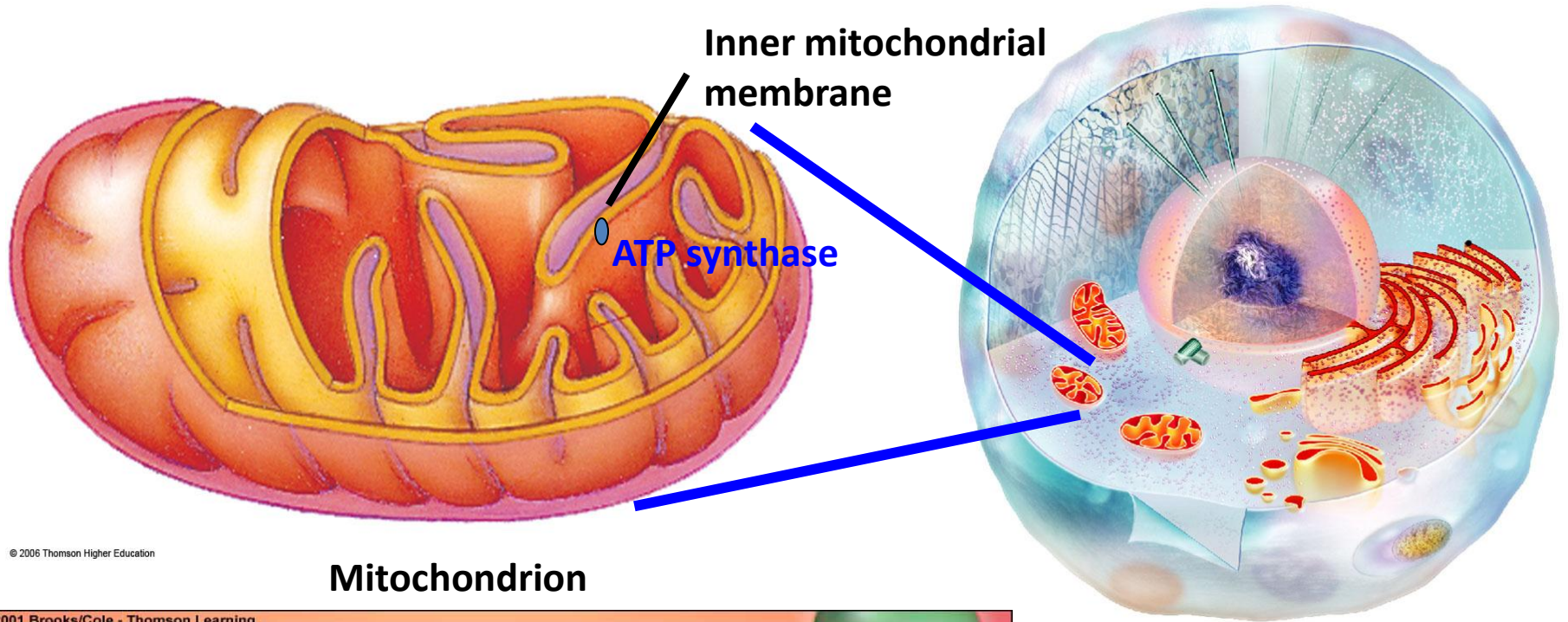
進行數據分析

數據分析

- 假設驅動的實驗數據分析
- 數據驅動的實驗數據分析

假設驅動的實驗數據分析

ATP合成酶

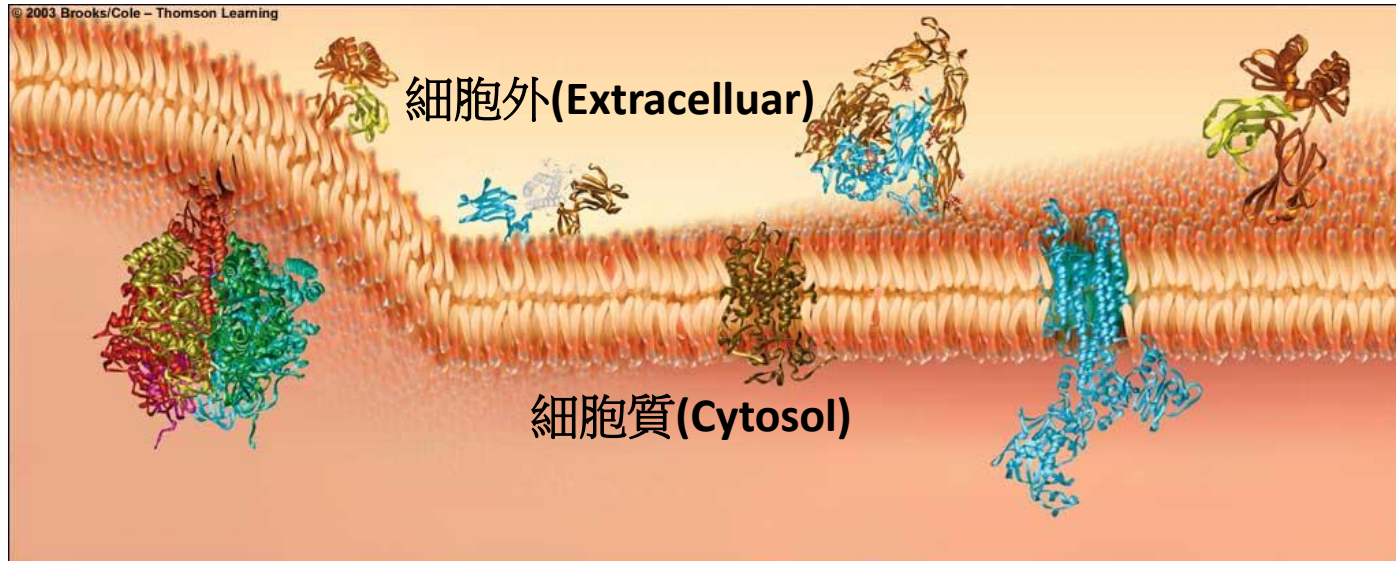


The flow of hydrogen ions through an ATP synthase drives the formation of ATP.

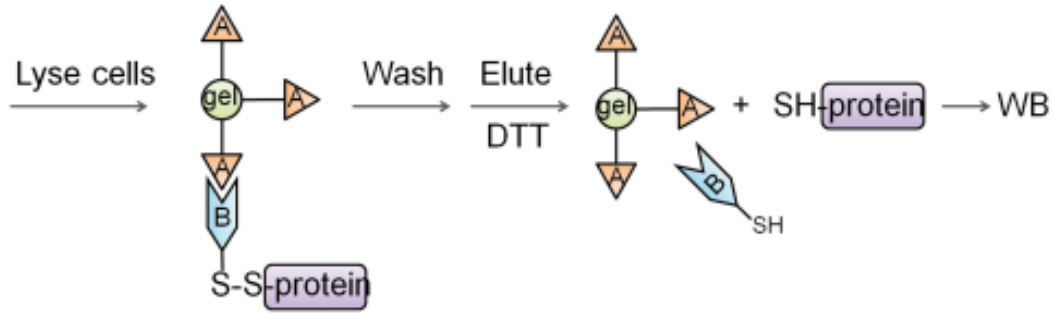
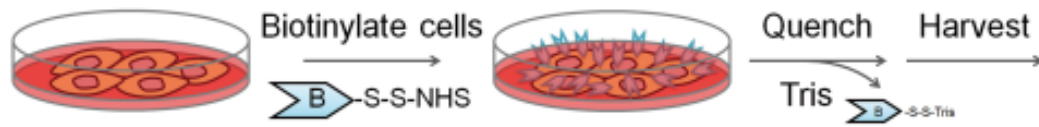
實驗一

假說: 是否ATP合成酶會表現在肺癌細胞膜上?

預期: 在肺癌細胞膜上量測到ATP合成酶



方法



分析和結果解讀

電子傳遞鏈
蛋白質



ATP合成酶



肺癌細胞

CL1-0 A549

Complex I
NDUFB4



Complex II
SDHA



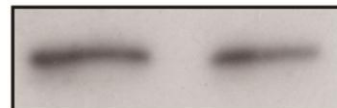
Complex III
Core2



Complex IV
Cox5A



Complex V
ATP5B



結論

- **ATP**合成酶會表現在肺癌細胞膜上

實驗二A

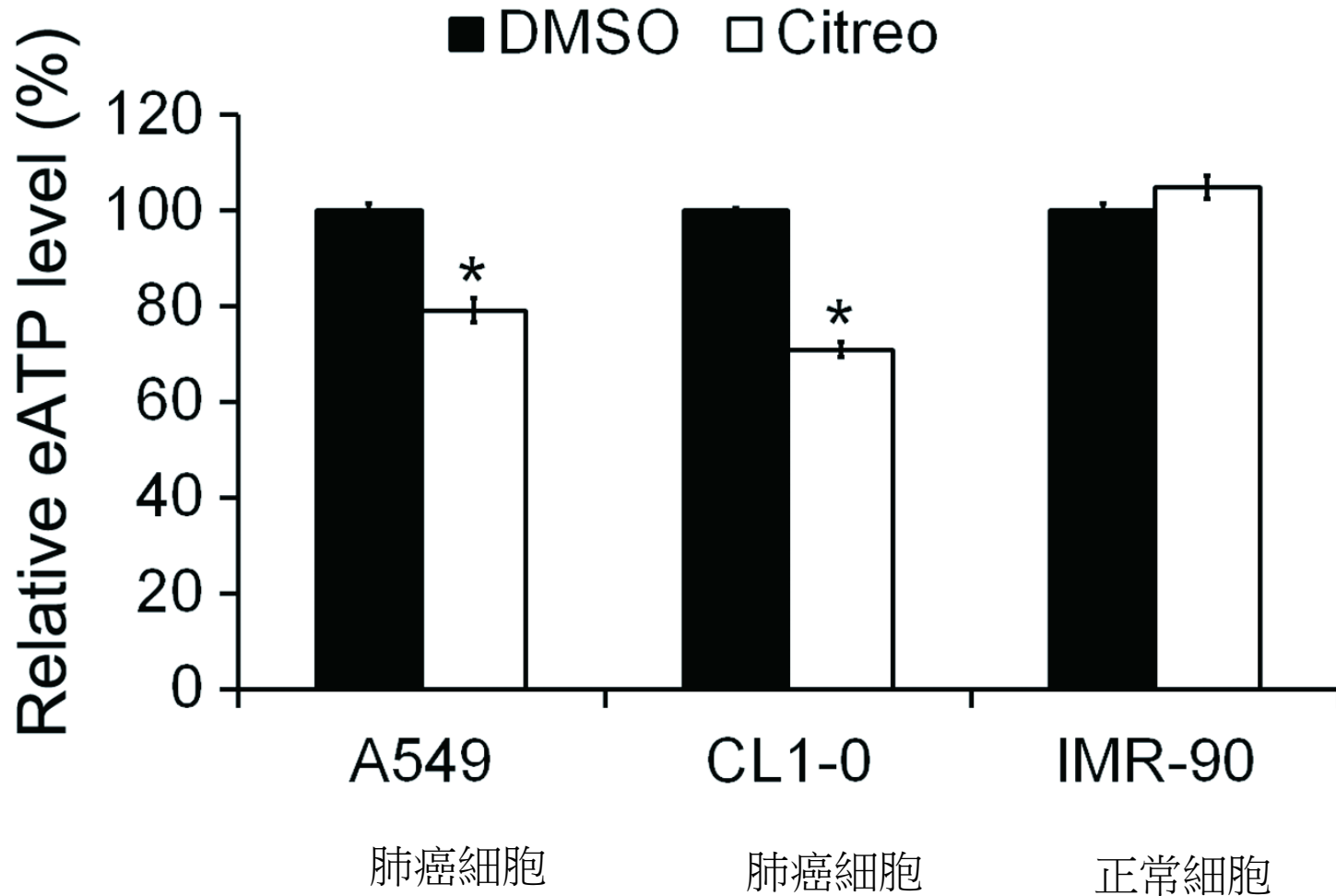
問題: 為何ATP酶出現在肺癌細胞上?

假說: 是否肺癌細胞膜上的ATP合成酶功能是產生ATP至細胞外?

預期: 加入ATP合成酶抑制劑會使肺癌細胞外ATP生成濃度下降

數據分析和解讀

*代表統計數據, $p < 0.05$



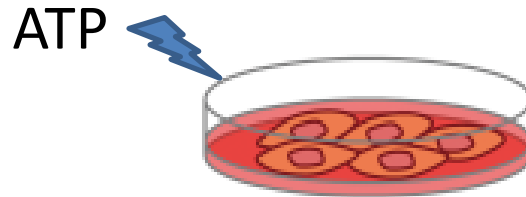
結論

- **ATP合成酶**功能是產生**ATP**至細胞外

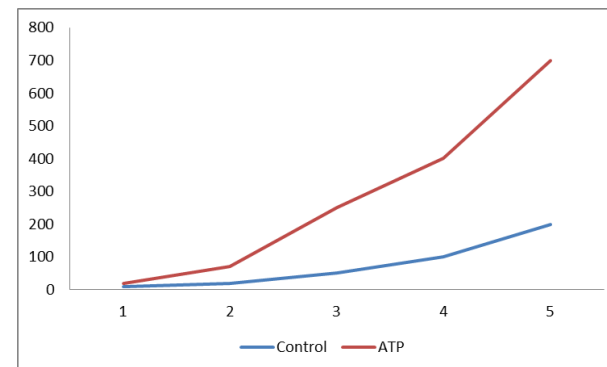
實驗二B

問題: 為何ATP酶出現在肺癌細胞上?

假說: 是否釋放出的ATP會促使肺癌細胞改變生長速率?

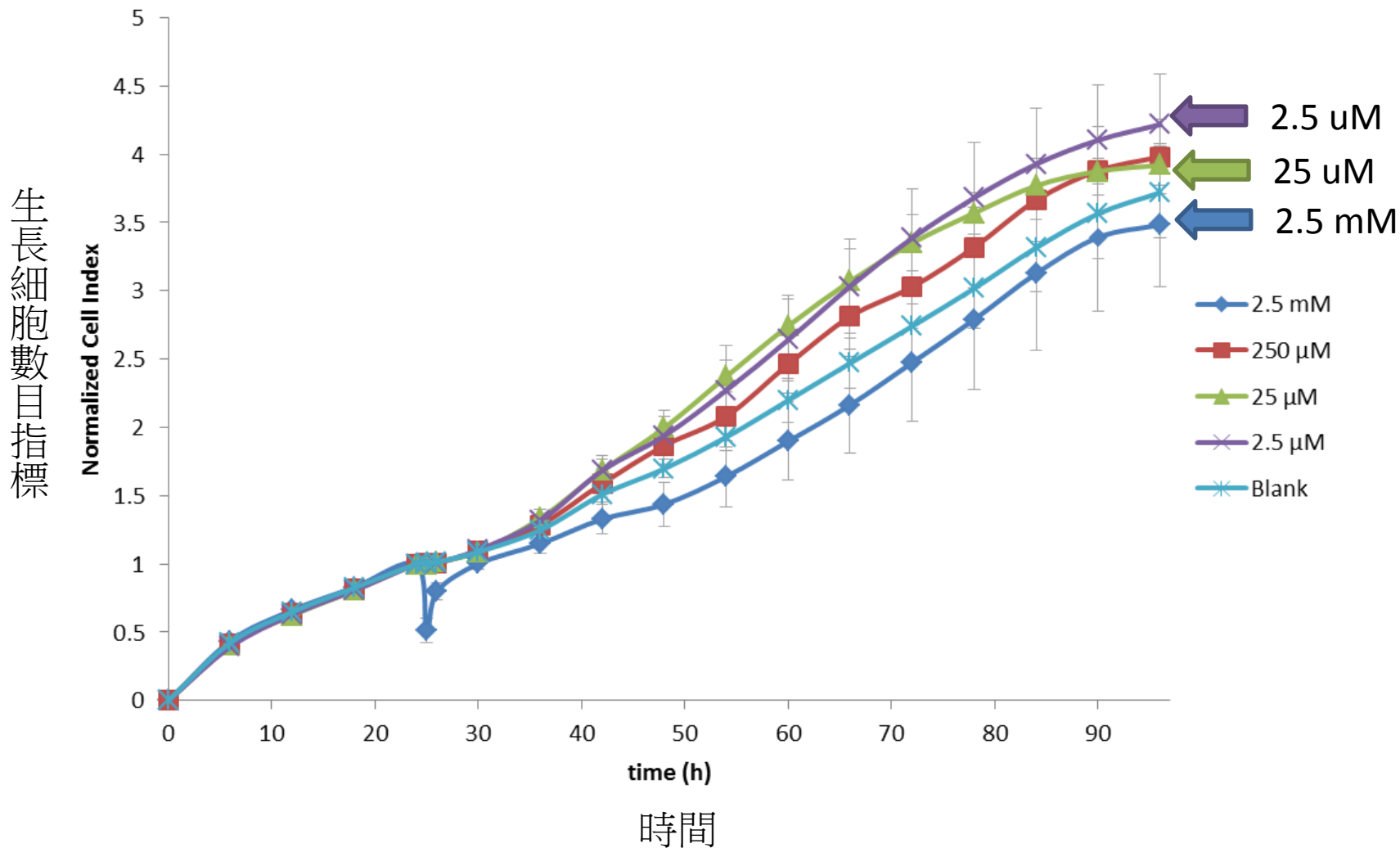


預期: 外加ATP到培養肺癌細胞的培養基中看到細胞生長曲線往上



數據分析和解讀

缺點: 尚未計算P值確定實驗數據是否具統計意義



結論

- 並不是只要外加**ATP**都會促使細胞生長
- 在某個特定的外加**ATP**濃度時，肺癌細胞的數目才會變多，生長曲線變快
- 在肺癌細胞膜上的**ATP**合成酶的功能是釋放**ATP**分子，並誘發其他肺癌細胞增生

數據驅動的實驗數據分析

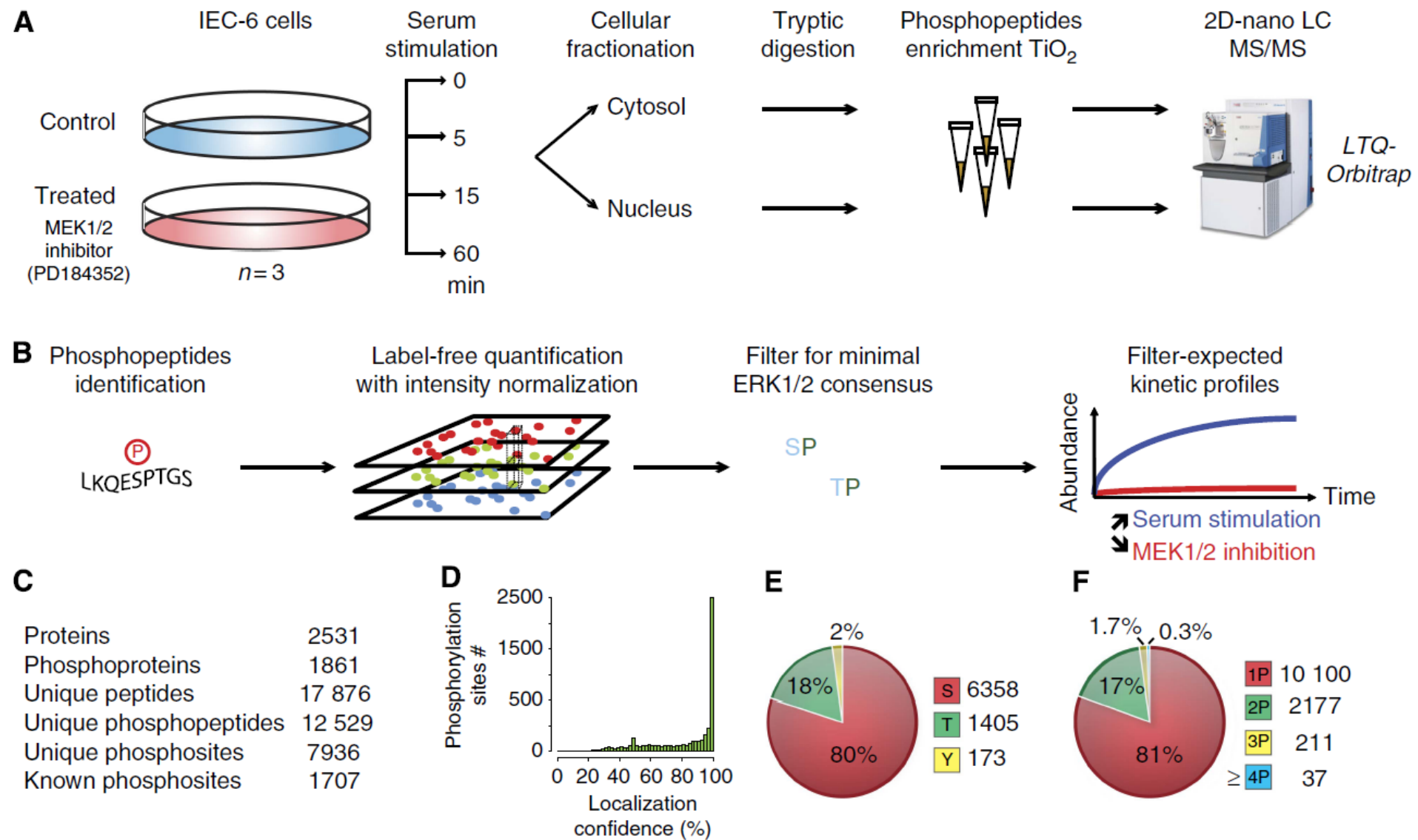
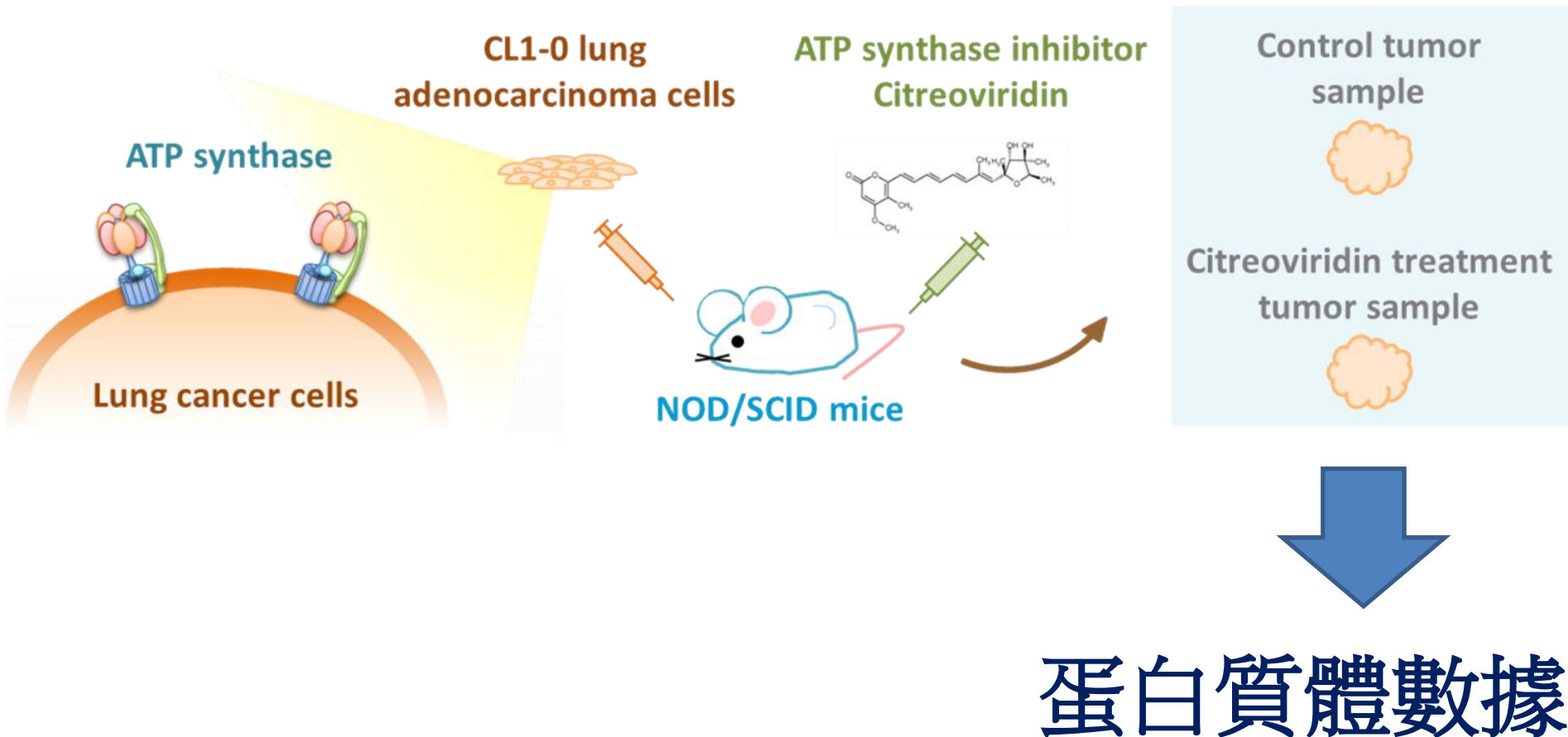


Figure 1 Experimental workflow and data processing for the identification of candidate ERK1/2 substrates. **(A)** Experimental workflow for sample processing and MS analysis. **(B)** Data analysis for the selection of candidate ERK1/2 substrates. **(C)** Statistics on the number of identified phosphopeptides. **(D)** Distribution of site-localization confidence data. **(E)** Distribution of phosphorylated amino acids. **(F)** Number of phosphorylation sites per peptide.

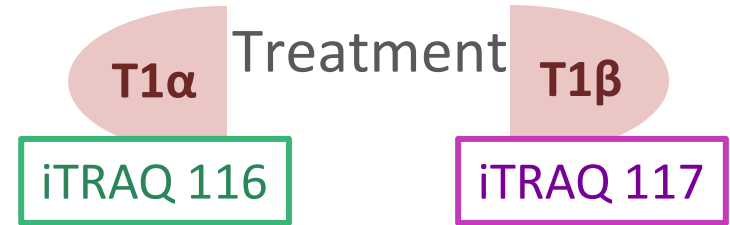
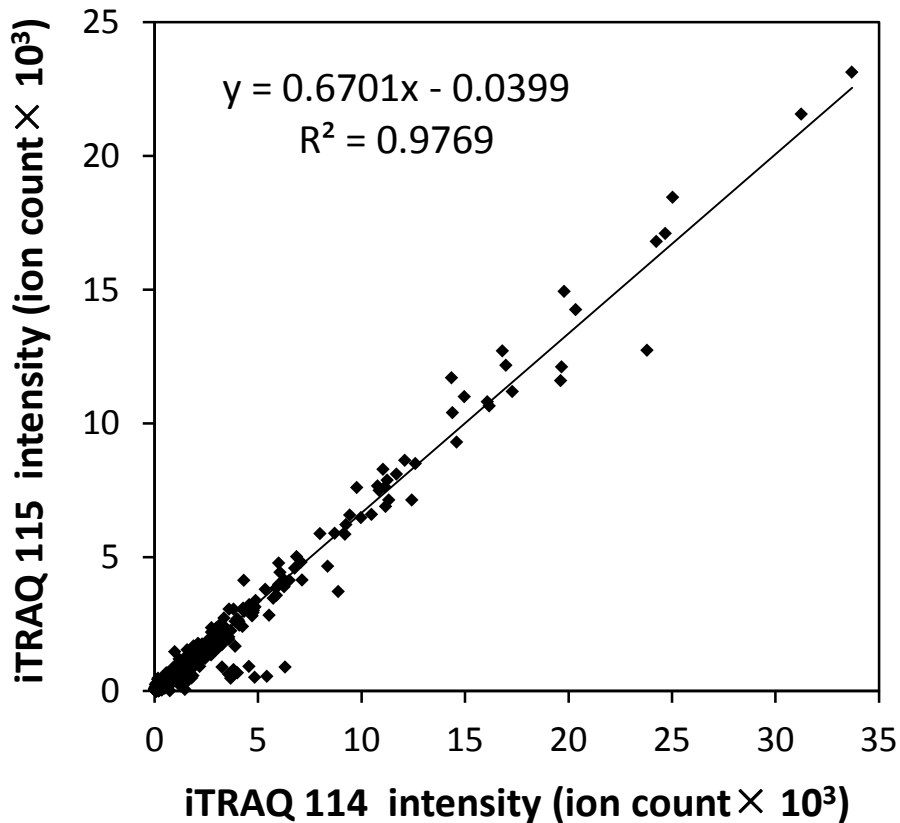
實驗一: ATP合成酶抑制劑抑制老鼠腫瘤生成是經由影響哪些蛋白質變化?



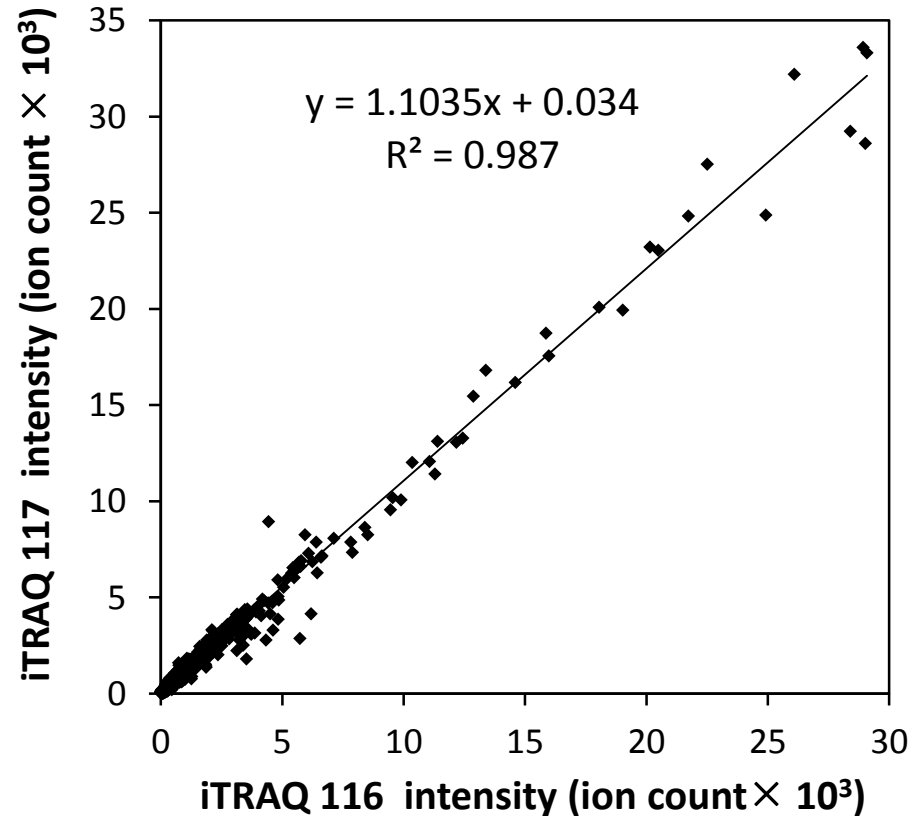
確定data重複性是否夠好?



N = 1152 $R^2 = 0.9769$



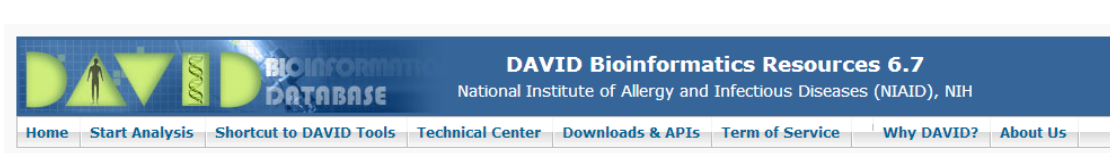
N = 1163 $R^2 = 0.9879$



R²

- R²是用來解釋線性迴歸模式的適配度(Goodness of Fit)
- 當R²=0時，代表反應變數(Y)與解釋變數(X)之間沒有線性關係存在
- 當R²≠0時，是反應變數(Y)被解釋變數(X)所解釋的比率值，計算公式如下 **R²= 1-(SSE/SST)**，其中SSE為隨機誤差的變異量，SST是總變異量。
- R²是迴歸模式可以解釋的變異量。

生物資訊分析



Human differential
proteomic profiling

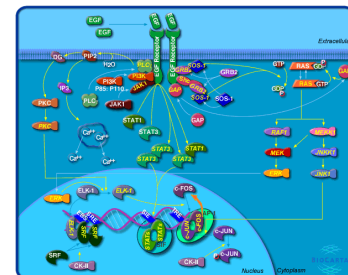


249

基因功能註解分群
(Gene Ontology enrichment)



路徑圖譜分析
(Pathway map analysis)



基因功能註解的三大群 (Three categories of GO)

- **Biological process (BP)** refers to a biological objective to which the gene or gene product contributes.
 - Cell growth and maintenance
 - Signal transduction
- **Molecular function (MF)** is defined as the biochemical activity (including specific binding to ligands or structures) of a gene product.
 - Enzyme
 - Transporter
 - Ligand
- **Cellular component (CC)** refers to the place in the cell where a gene product is active.
 - Nuclear membrane
 - Golgi apparatus

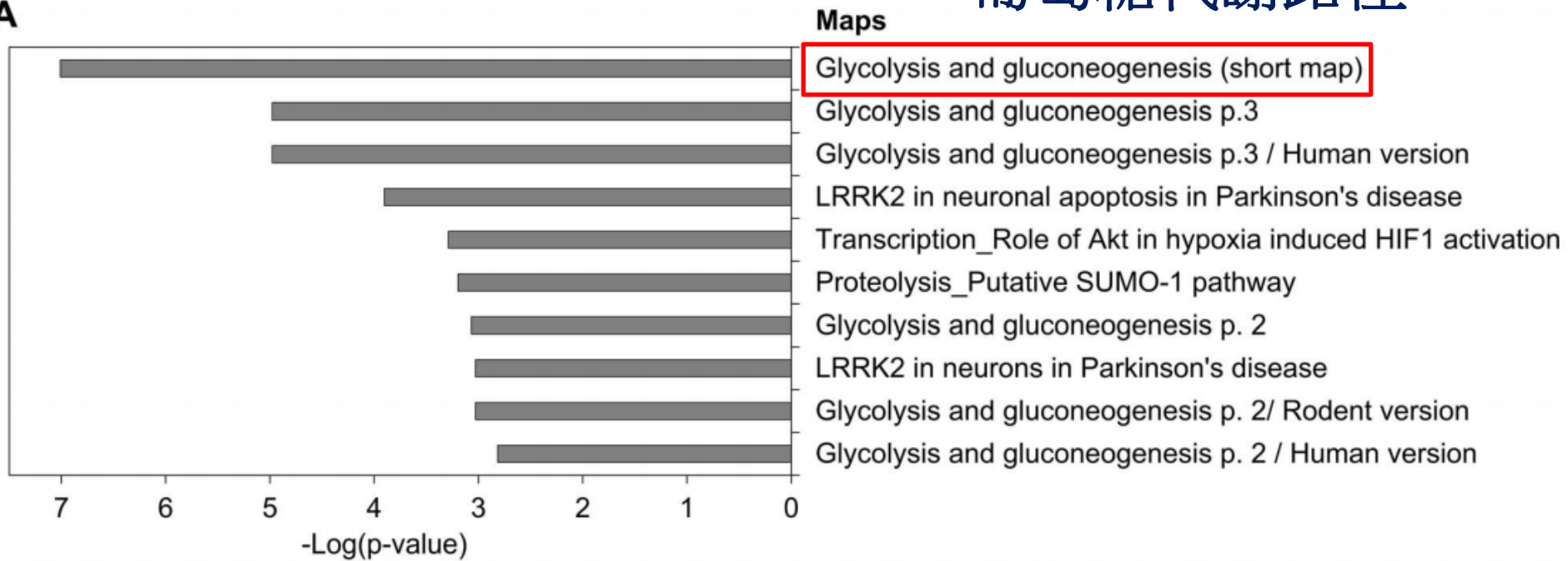
基因功能註解分群

No.	GO biological process cluster	Enrichment score
1	Glycolysis	6.52
2	Glucose metabolic process	6.41
3	Protein ubiquitination	5.91
4	Macromolecule catabolic process	4.47
5	Regulation of developmental growth	2.31
6	Regulation of programmed cell death	1.77
7	Tricarboxylic acid cycle	1.66
8	Cellular amide metabolic process	1.65
9	RNA stabilization	1.52
10	Regulation of cellular component biogenesis	1.52
11	Eicosanoid metabolic process	1.47
12	Cellular component disassembly	1.40
13	Negative regulation of apoptosis	1.40

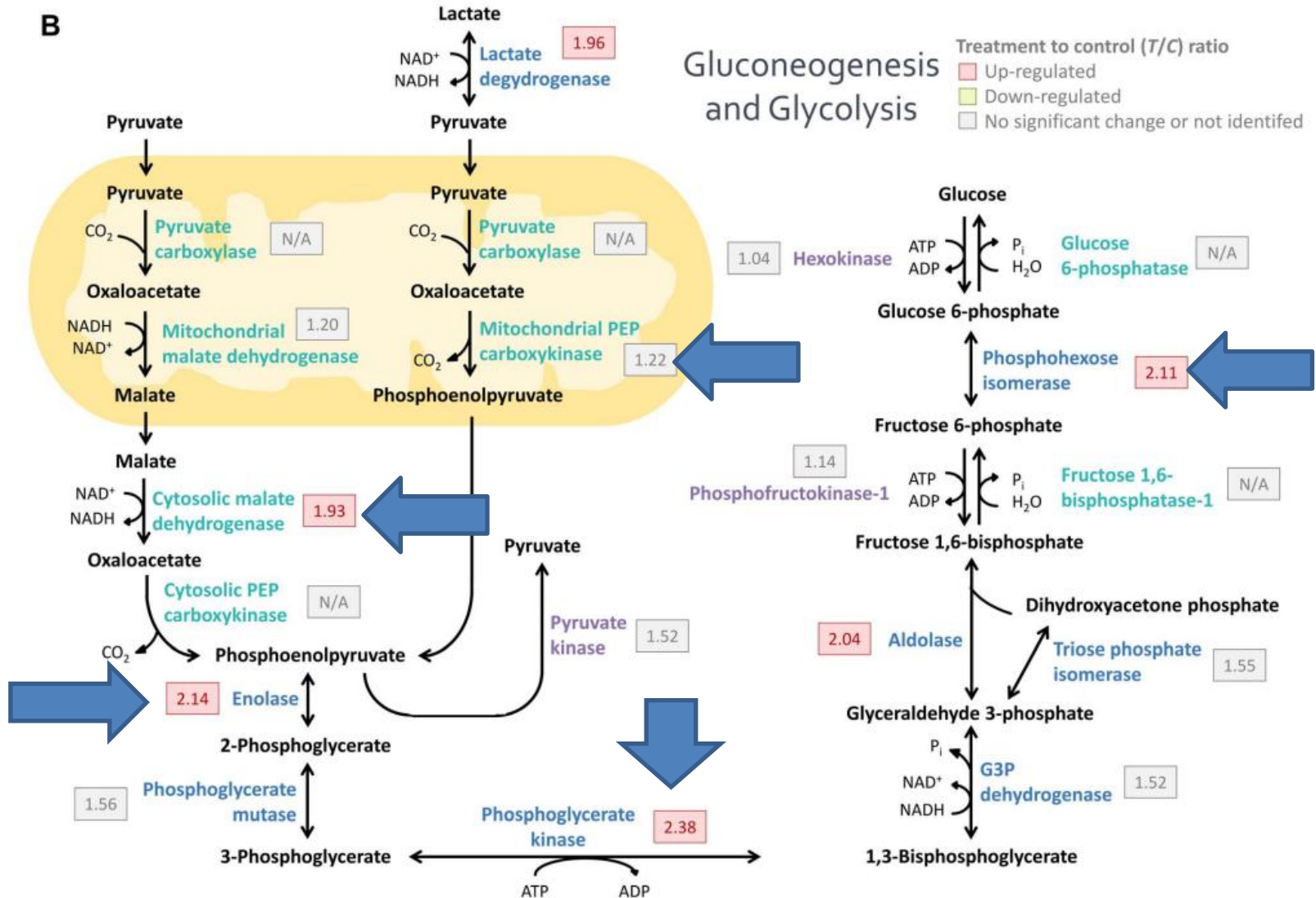
路徑圖譜分析

葡萄糖代謝路徑

A

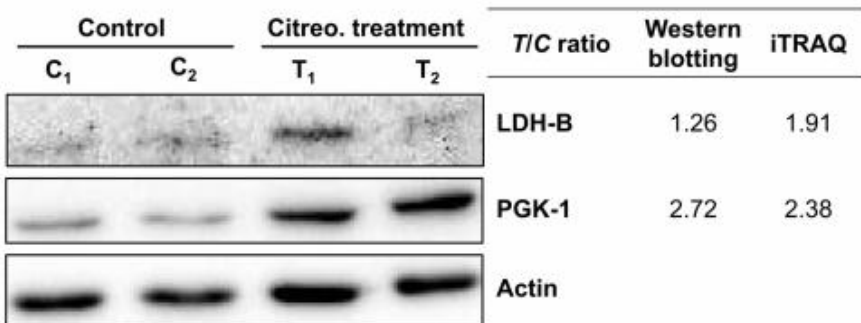
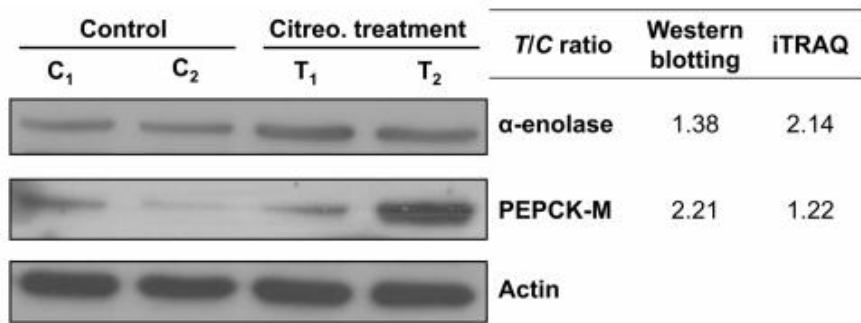


葡萄糖代謝路徑

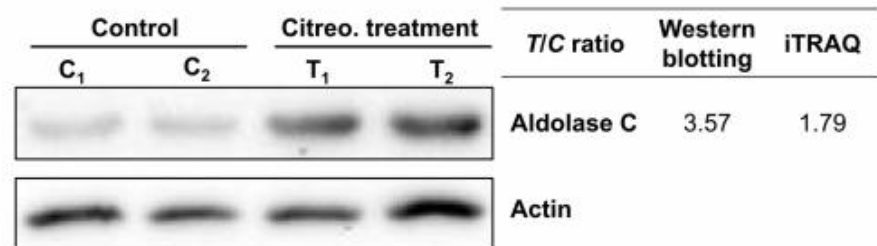
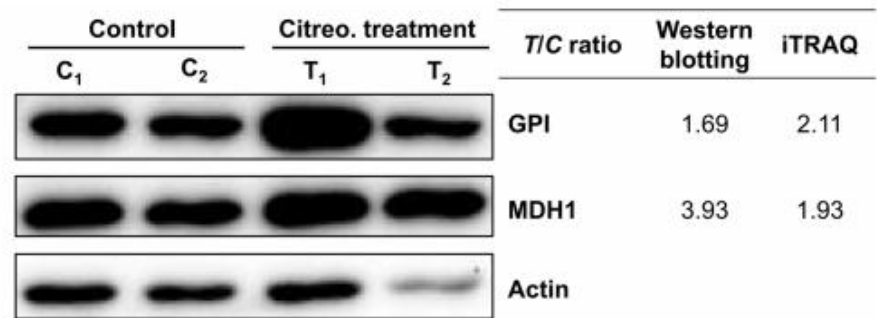


西方墨點法證明蛋白質體實驗中 葡萄糖代謝相關酵素蛋白質表現

三次實驗平均值



三次實驗平均值



總結

- 根據問的問題設計實驗
- 實驗重複性檢測
- 多用統計說服審查者及讀者數據的可信度
- 善用生物資訊工具分析高通量資料

Take Home Message

- 鼓勵對生物有興趣的學生學習統計
- 鼓勵對生物有興趣的學生學習資訊

2013年國際系統
生物學研討會

