

Εύρηκα

分析肺鱗狀細胞癌(LUSC)相關 長鏈非編碼核糖核酸(lncRNA) 調控之基因網路模組

顏頌翰
台大生科系 阮雪芬 教授
魏宏仁 老師



研究動機

Eύρηκα

- 長鏈非編碼核糖核酸(lncRNA, Long non-coding RNA) 在各種生物作用中扮演關鍵角色，但大多數的功能目前仍不是很清楚。
- 研究指出，lncRNA的異常表現會影響腫瘤，促進或抑制癌症的作用，所以lncRNA可能可以作為判斷肺癌的標準或是治療的標靶。
- 本研究著眼於非小細胞肺癌(NSCLC, non-small cell lung cancer)肺癌中的lncRNA，針對NSCLC當中比例佔第二的肺鱗狀細胞癌(squamous cell carcinoma)，利用TCGA資料庫的基因的表現量，建立基因網路模組，找出重要的基因，作為判斷肺癌的標準或是治療的標靶。

研究目的

- 一. 利用已經建立好的lncRNA-mRNA bipartite network，進而計算肺鱗狀細胞癌(LUSC)中長鏈非編碼核糖核酸對(lncRNA-lncRNA pair)的皮爾森相關係數(PCC)。
- 二. 將edge list (lncRNA-lncRNA pair)送入Cytoscape建立lncRNA association network。
- 三. 計算每個節點(node)的級數(degree)、中間性(betweenness centrality)和聚類係數(clustering coefficient)。



研究器材與設備

- 阮雪芬教授系統生物學實驗室中已建立好的
lncRNA association index 資料
- Python程式語言
- Cytoscape軟體



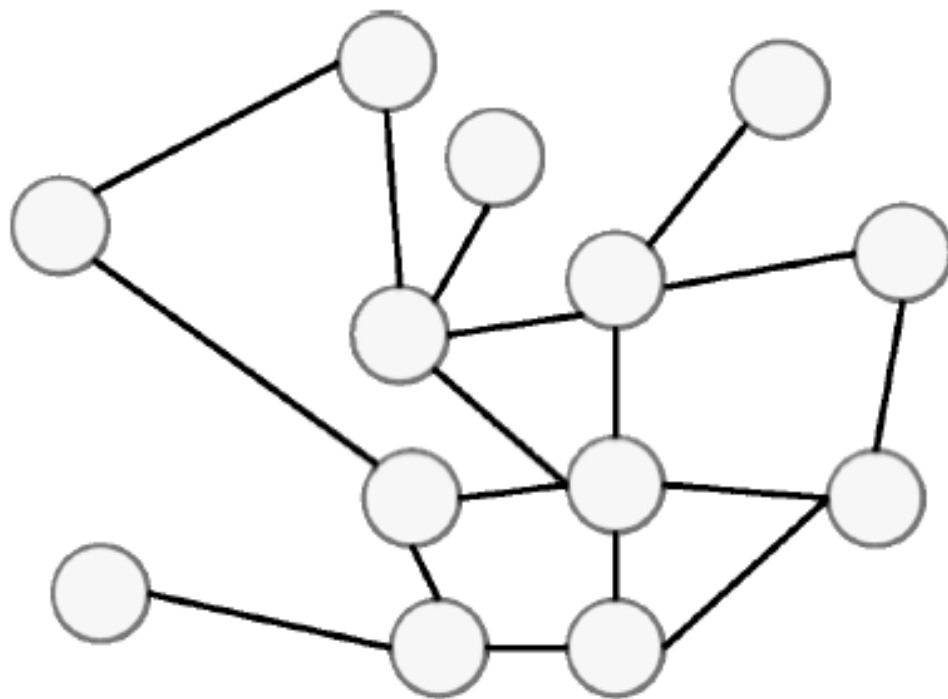
(PCC)(註5)

The **Pearson correlation coefficient** is the correlation between the interaction profiles of A and B.

$$PCC_{AB} = \frac{|N(A) \cap N(B)| \cdot n_y - |N(A)| \cdot |N(B)|}{\sqrt{|N(A)| \cdot |N(B)| \cdot (n_y - |N(A)|) \cdot (n_y - |N(B)|)}}$$

$$PCC_{AB} = \frac{|N(A) \cap N(B)| \cdot n_y - |N(A)| \cdot |N(B)|}{\sqrt{|N(A)| \cdot |N(B)| \cdot (n_y - |N(A)|) \cdot (n_y - |N(B)|)}}$$

背景知識(註6)



○ Node — Edge

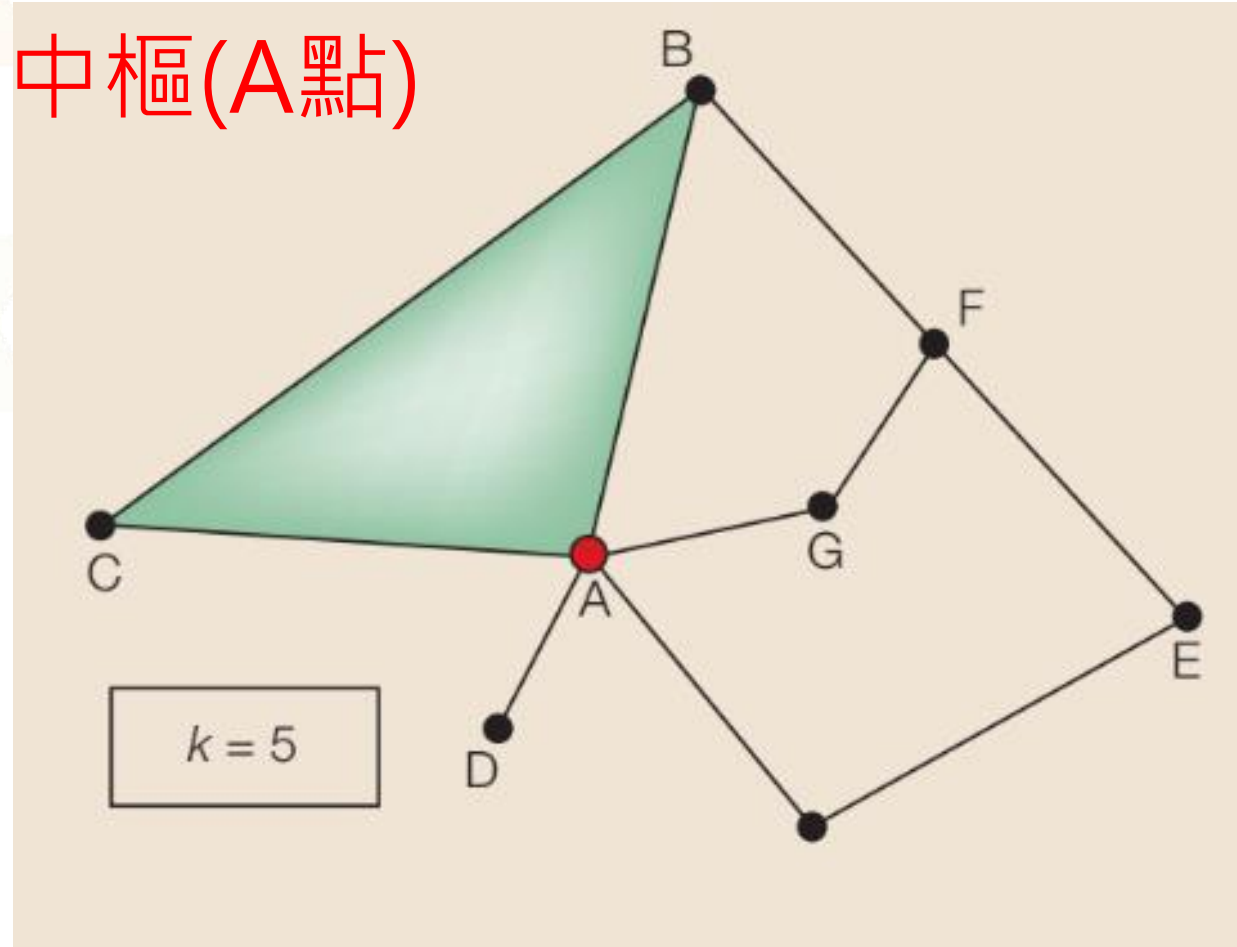
Node: 網路圖中的單位
Edge: 單位之間的連結



背景知識(註7)

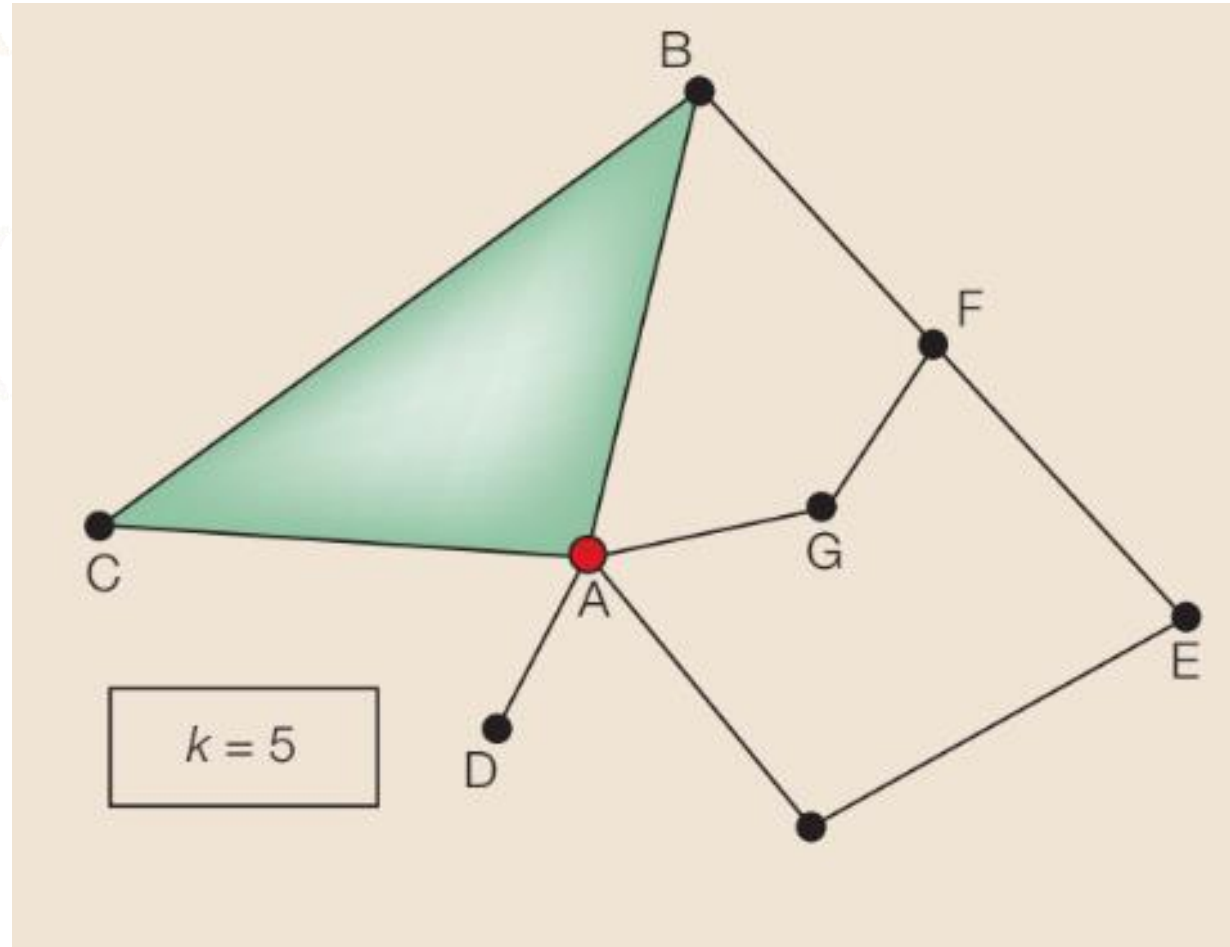
Hub:網路圖的中樞(A點)

Degree:一個
node連接的
edge個數



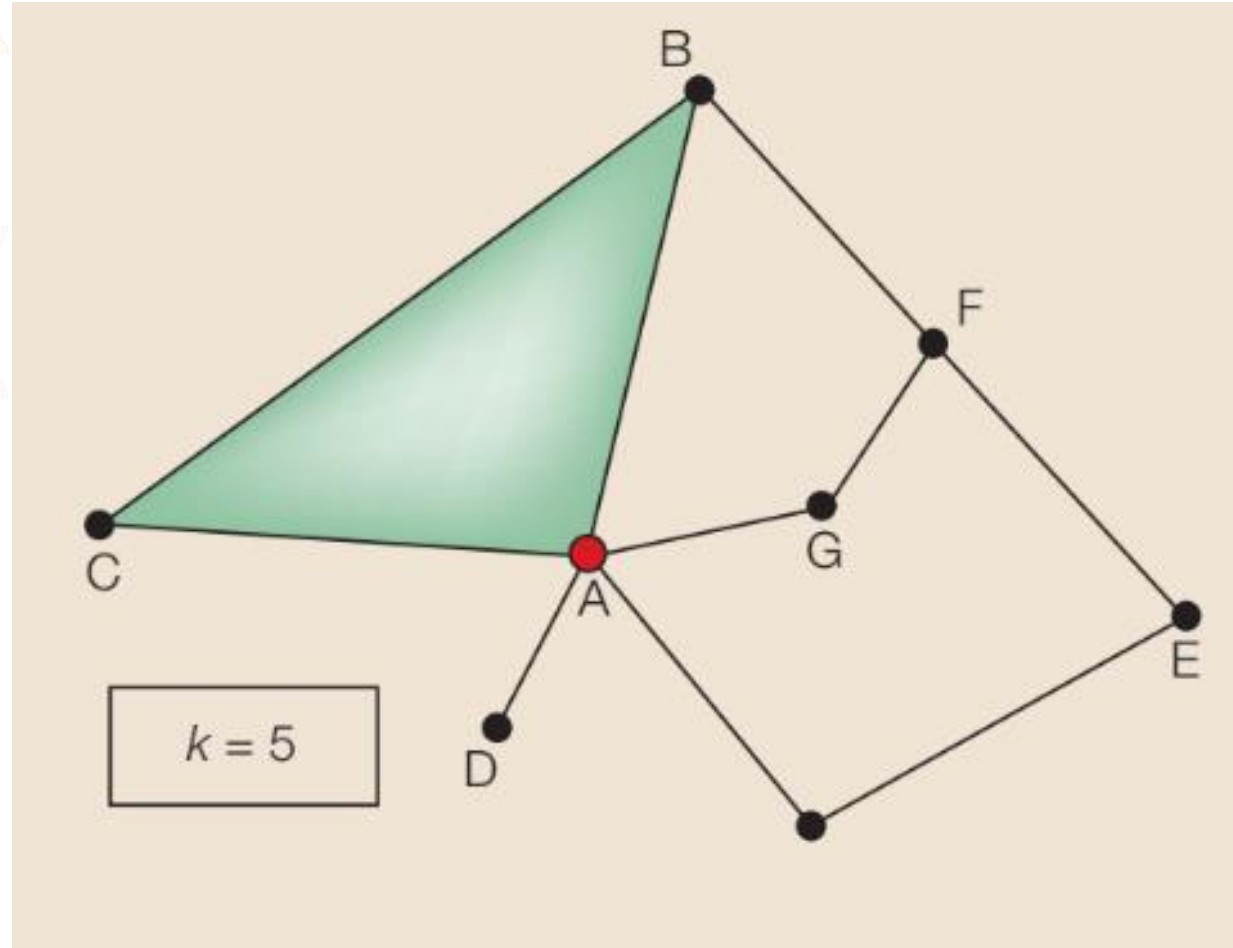
背景知識(註7):Betweenness Centrality

$$B = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$



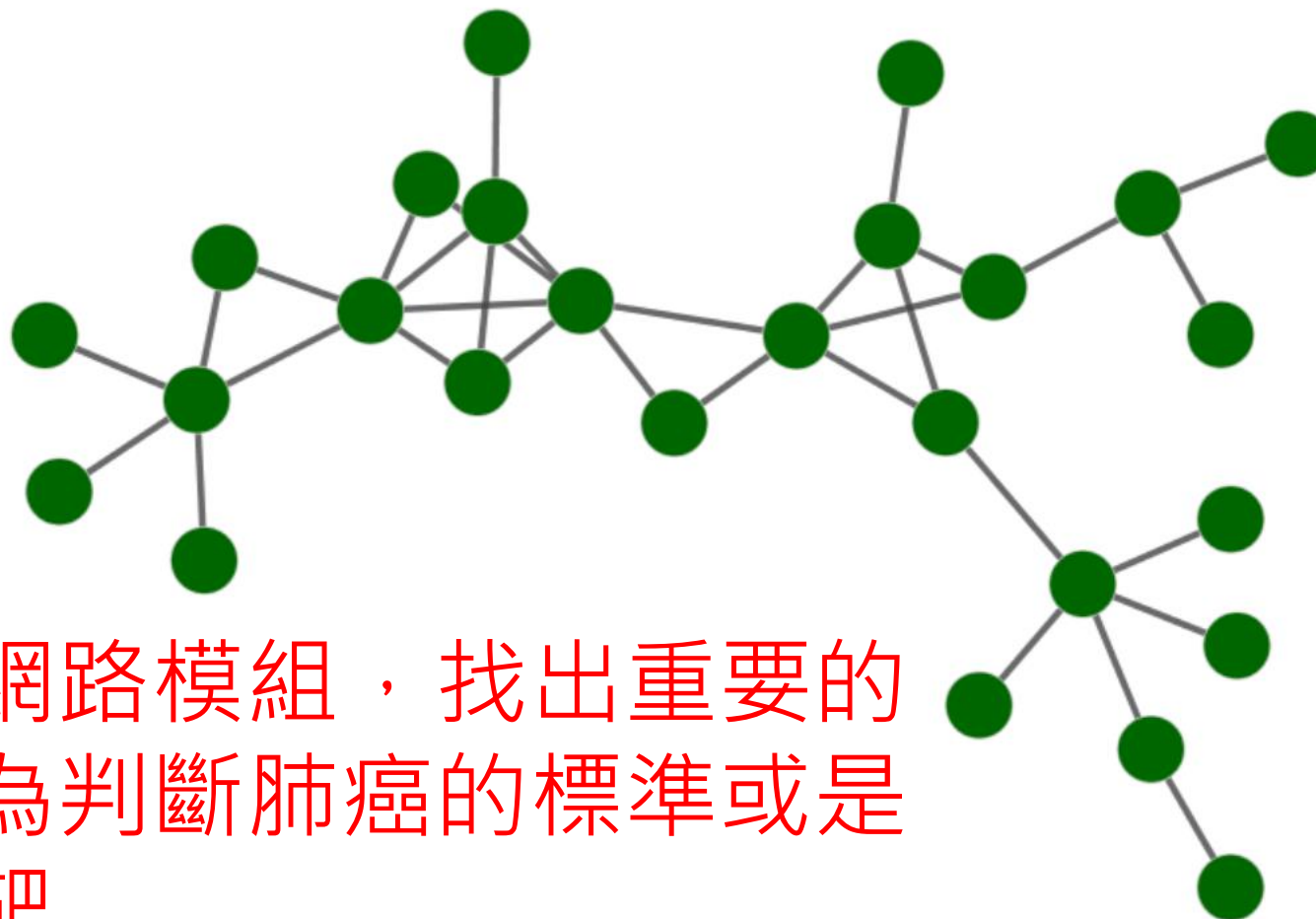
背景知識(註7):Clustering Coefficient

$$C_I = \frac{2n_I}{k(k-1)}$$



研究動機

Εύρηκα



利用基因網路模組，找出重要的基因，作為判斷肺癌的標準或是治療的標靶



研究方法一

一、計算肺鱗狀細胞癌(LUSC)中長鏈非編碼核糖核酸對(lncRNA-lncRNA pair)的皮爾森相關係數(PCC)。

1. 從The Cancer Genome Atlas(TCGA)資料庫中下載肺鱗狀細胞癌(LUSC)以及肺腺癌(LUAD)的長鏈非編碼核糖核酸對(lncRNA-lncRNA pair)的基因表現，並計算lncRNA-mRNA 的correlation後建立lncRNA-mRNA correlation bipartite network，接著利用association index (皮爾森相關係數(PCC))來分析lncRNA-lncRNA的association
2. 利用Python對數據讀檔並依大小排序，參考肺腺癌(LUAD) 中長鏈非編碼核糖核酸(lncRNA)的皮爾森相關係數(PCC)，計算要留下多少比例的lncRNA-lncRNA pair，將這個比例套用在LUSC上，得到最低的PCC。

研究方法二

Εύρηκα

二、畫出肺鱗狀細胞癌(LUSC)的長鏈非編碼核糖核酸 (lncRNA) 的基因網路模組。

1. 利用Python語言將LUSC中留下的lncRNA-lncRNA pair，匯出得到新檔案。
2. 將篩選過後的檔案，輸入軟體cytoscape並畫出LUSC的lncRNA 的基因網路模組



研究方法三

- 利用軟體cytoscape中的分析工具，計算各基因網路模組中，每個節點的級數(degree)、中間性(betweenness centrality)和聚類係數(clustering coefficient)



研究結果一

Εύρηκα

```
[1]: import pandas as pd
df_ad = pd.read_csv('C:/Users/user/Desktop/LUAD_lnc_mes_association_index_com.csv')
print(df_ad.shape)
ad = df_ad.shape[0]
```

(6011778, 3)

```
[2]: df = pd.read_csv('C:/Users/user/Desktop/LUAD_lnc_mes_association_index_com.csv')
print(df_ad[df_ad.PCC>=0.7].shape, df_ad[df_ad.PCC>=0.7].sort_values(by = ["PCC"], ascending=False))
pcc = df_ad[df_ad.PCC>=0.7].shape[0]
```

		lncA	lncB	PCC
(729, 3)				
7890	ENSG00000273010	ENSG00000261242	1.000000	
2871905	ENSG00000261766	ENSG00000233967	1.000000	
2994922	ENSG00000230565	ENSG00000273419	1.000000	
2983516	ENSG00000228748	ENSG00000260808	1.000000	
2981135	ENSG00000244055	ENSG00000270332	1.000000	

·
·
·

1552608	ENSG00000224505	ENSG00000228925	0.707022
1287877	ENSG00000273729	ENSG00000273148	0.707022
5673218	ENSG00000261069	ENSG00000257647	0.706979

[729 rows x 3 columns]

肺腺癌(LUAD)
的PCC(6011778筆)
篩選比例:729/6011778
($PCC \geq 0.7$)



研究結果一

Εύρηκα

```
[3]: df_sc = pd.read_csv(r'C:\Users\user\Desktop\LUSC_lnc_mes_association_index_com.csv')
      print(df_sc.shape)
      scc = df_sc.shape[0]

(6557631, 3)
```

```
[4]: df_sc = pd.read_csv(r'C:\Users\user\Desktop\LUSC_lnc_mes_association_index_com.csv')
      df_sc.sort_values(by = ["PCC"], ascending=False).head(int(scc*pcc/ad))
```

```
[4]:
```

	lncA	lncB	PCC
6531886	ENSG00000275367	ENSG00000273245	1.000000
5709330	ENSG00000248714	ENSG00000265975	0.707065
1534581	ENSG00000235026	ENSG00000269416	0.707065
5821189	ENSG00000233791	ENSG00000260464	0.707065

⋮

4938431	ENSG00000196951	ENSG00000273399	0.288401
2466463	ENSG00000271551	ENSG00000273523	0.288401
5576687	ENSG00000272720	ENSG00000236204	0.288401

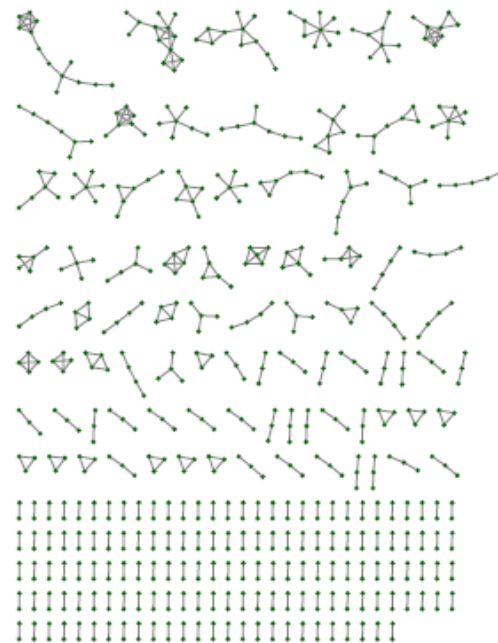
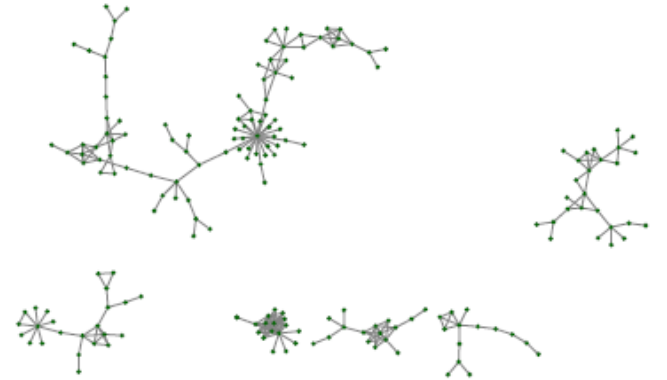
795 rows × 3 columns

肺鱗狀細胞癌(LUSC)
的PCC篩選比例:729/60116778

6557631 → 795
PCC=0.288401

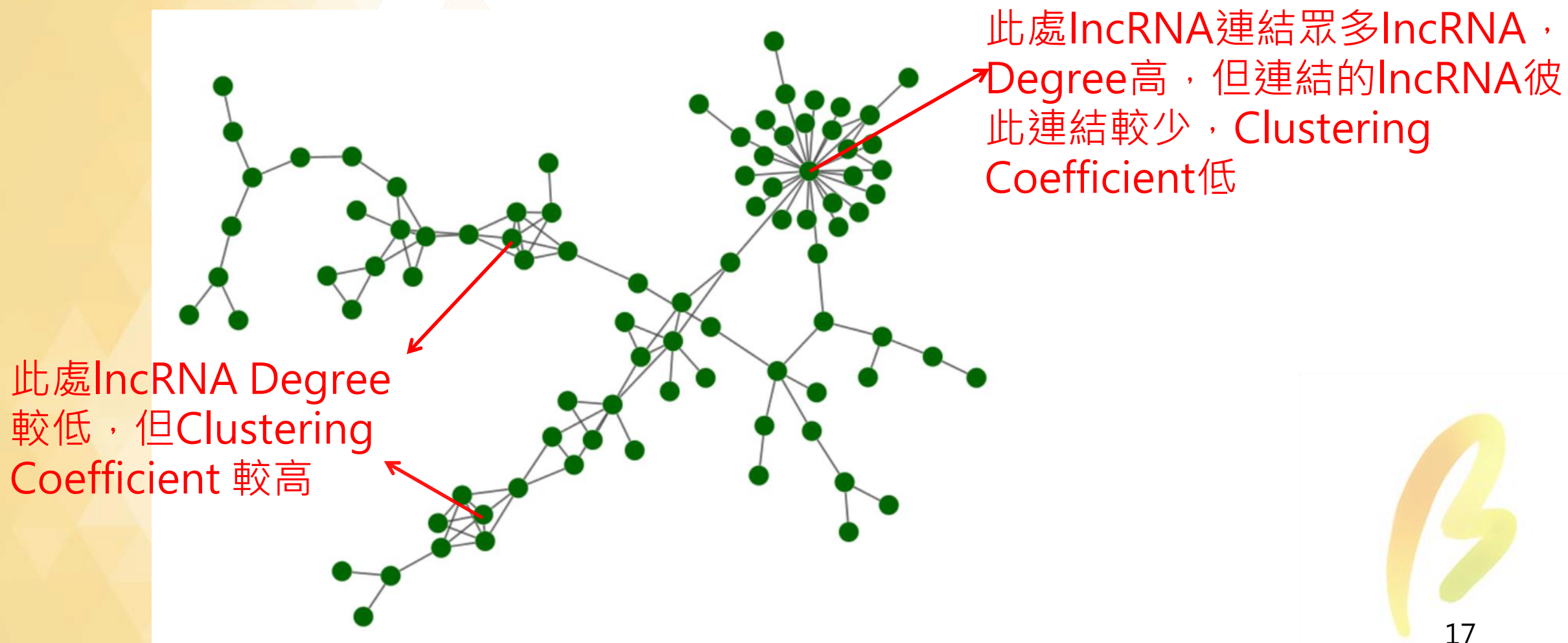


研究結果二



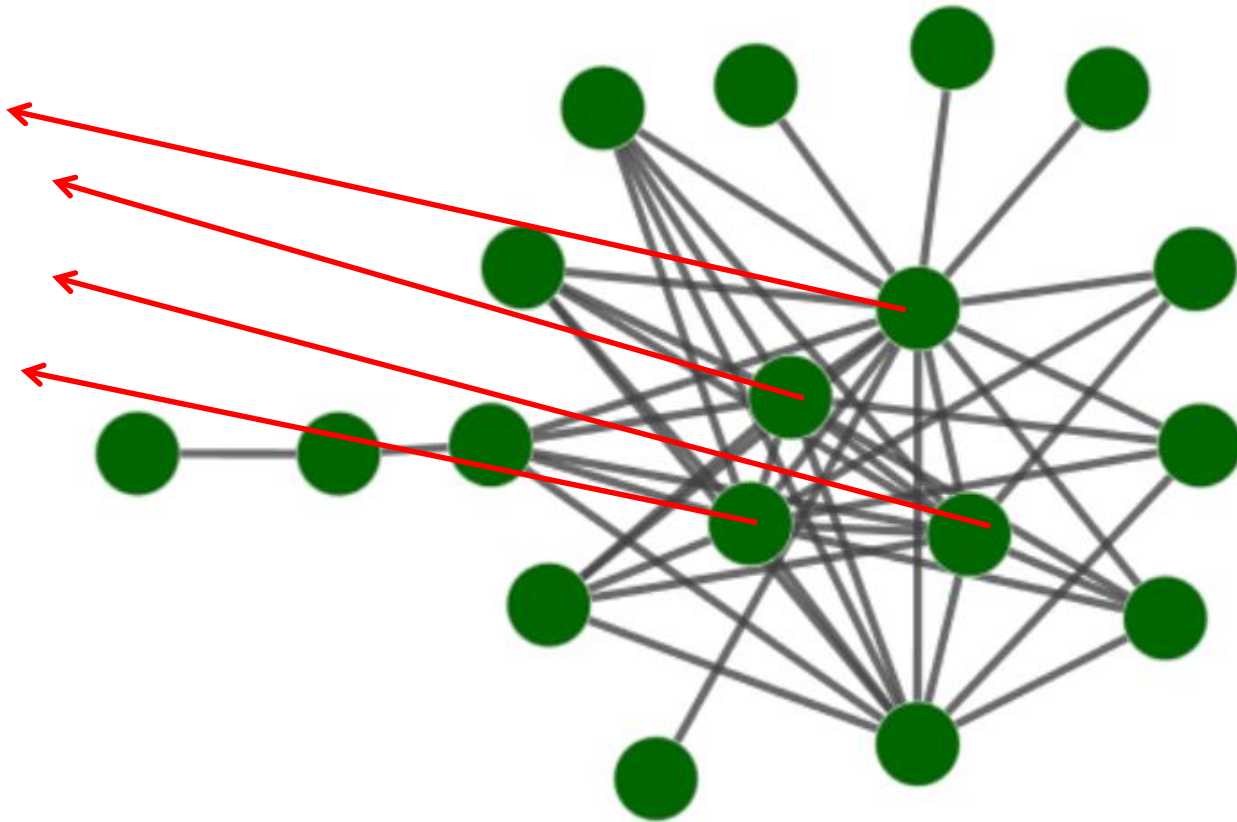
肺鱗狀細胞癌(LUSC)
的association
network

研究結果二:基因網路模組1

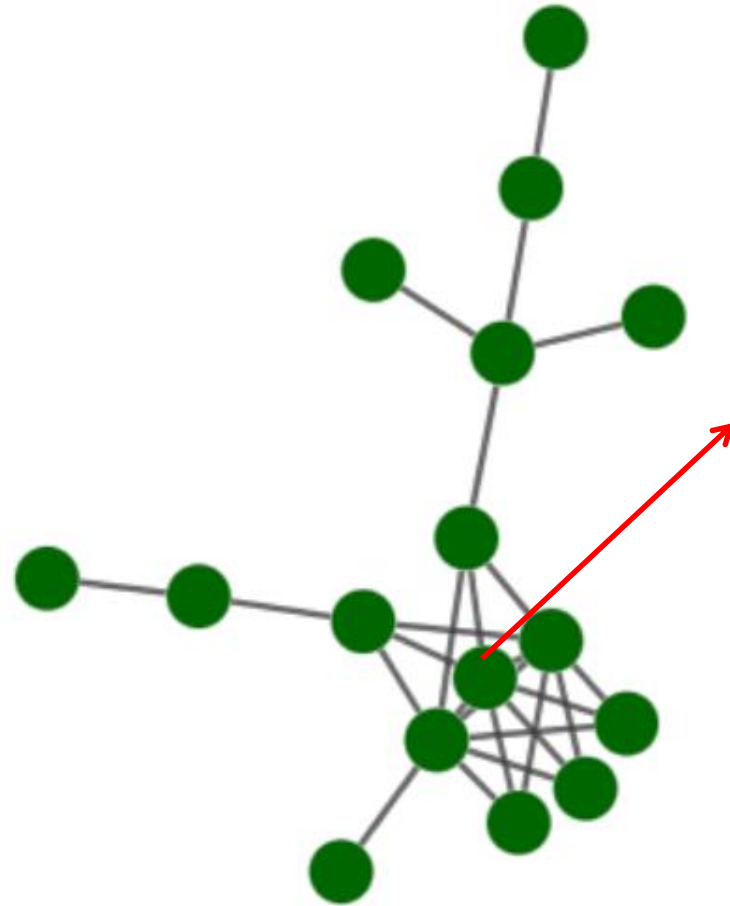


研究結果二:基因網路模組2

這四個lncRNA
Degree較高



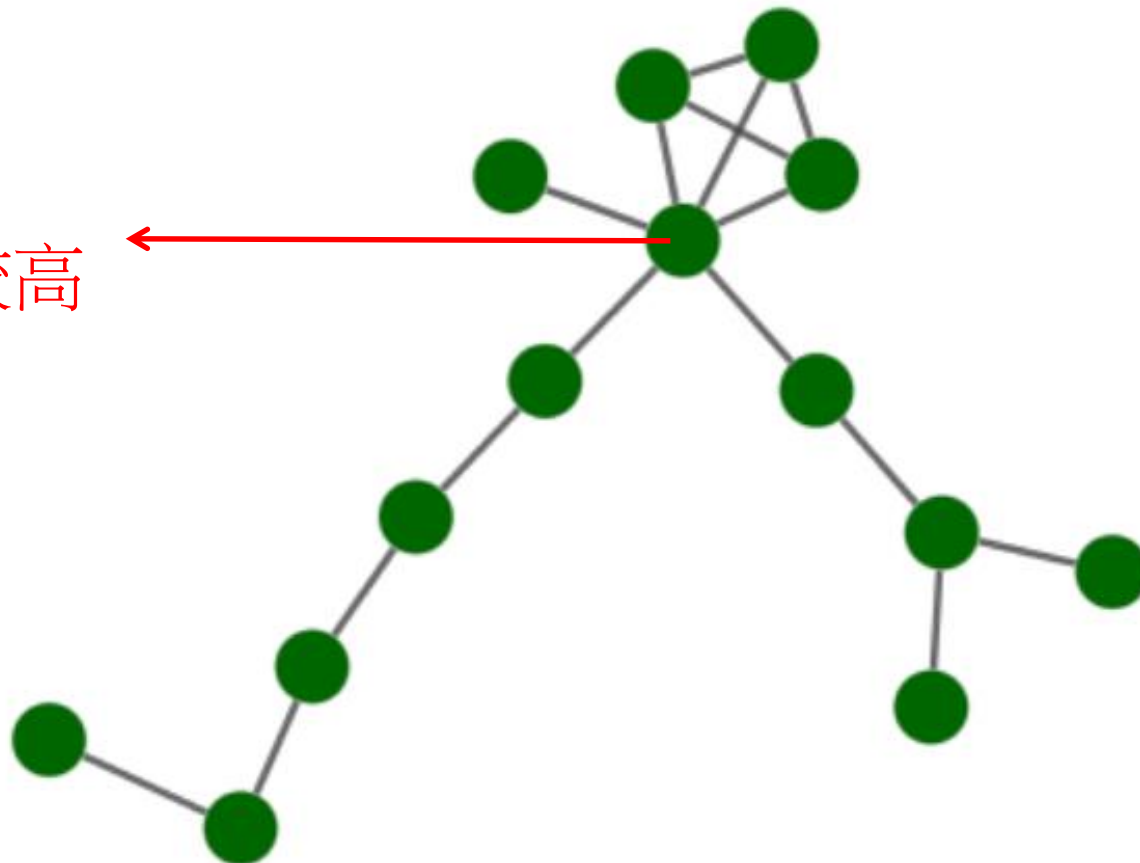
研究結果二:基因網路模組3



此lncRNA Degree及
Clustering Coefficient較高

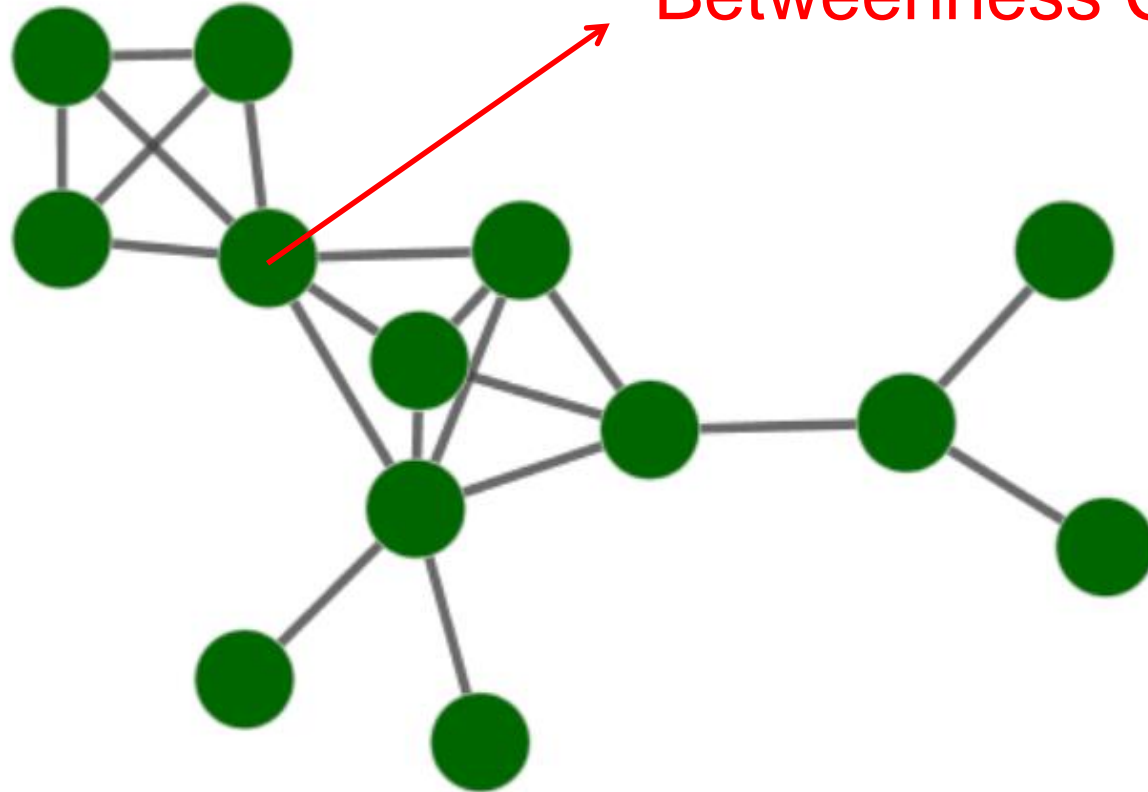
研究結果二:基因網路模組4

此lncRNA
Degree 較高

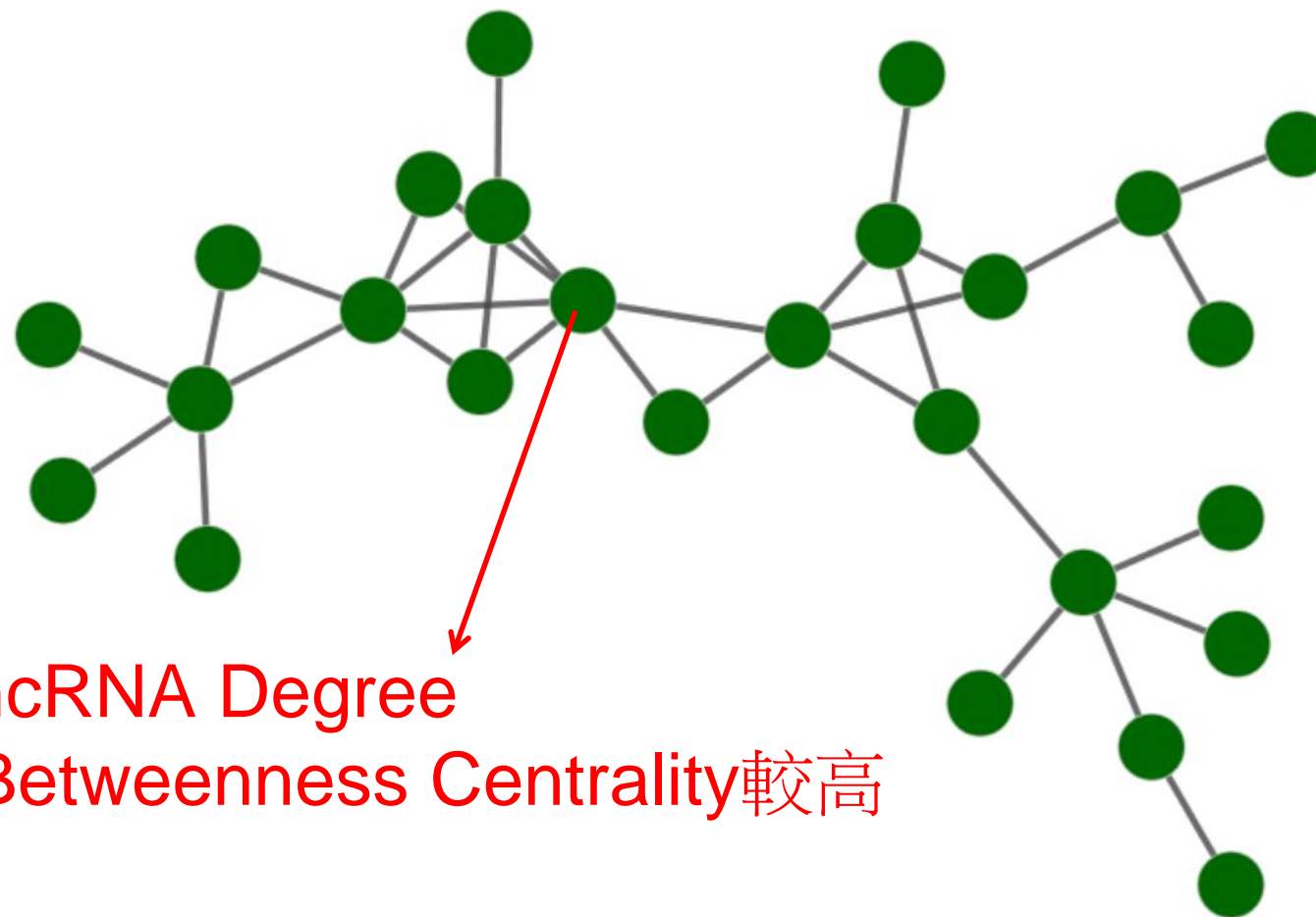


研究結果二:基因網路模組5

此lncRNA Degree及
Betweenness Centrality 較高



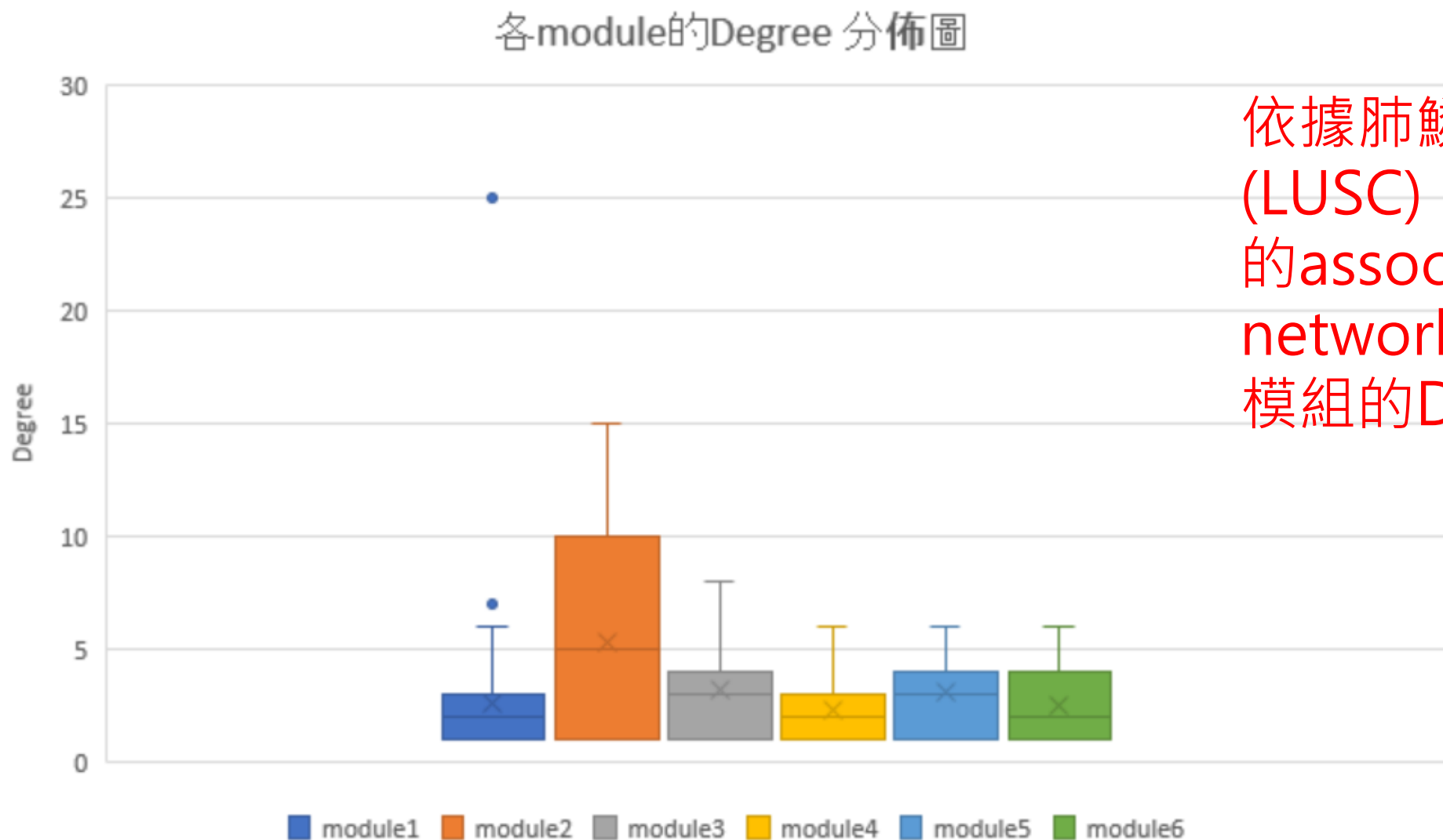
研究結果二:基因網路模組6



此IncRNA Degree
及 Betweenness Centrality較高



研究結果三:各基因網路模組Degree分佈圖

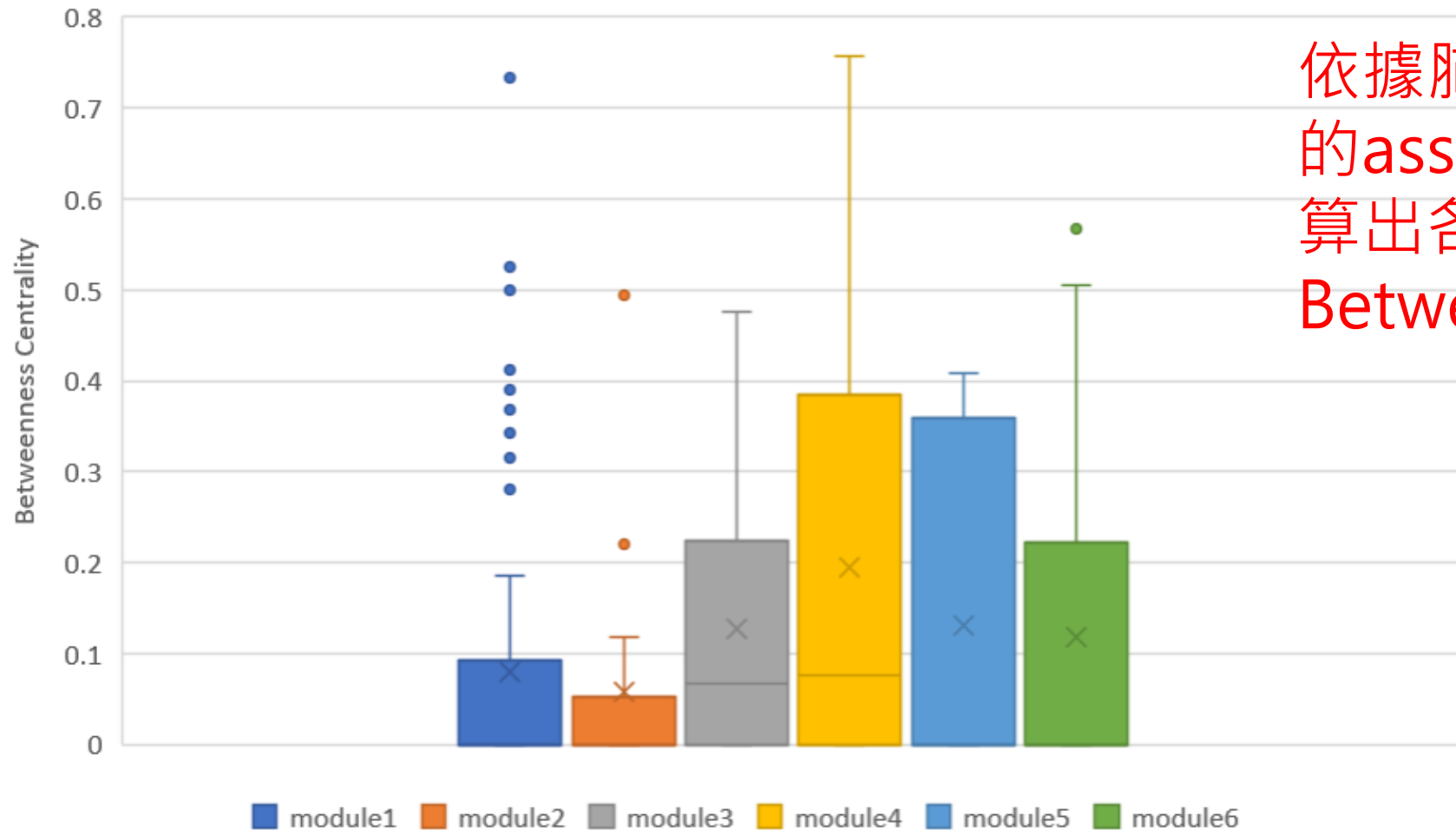


依據肺鱗狀細胞癌
(LUSC)
的association
network，算出各個
模組的Degree



研究結果三:各基因網路模組 Betweenness Centrality分佈圖

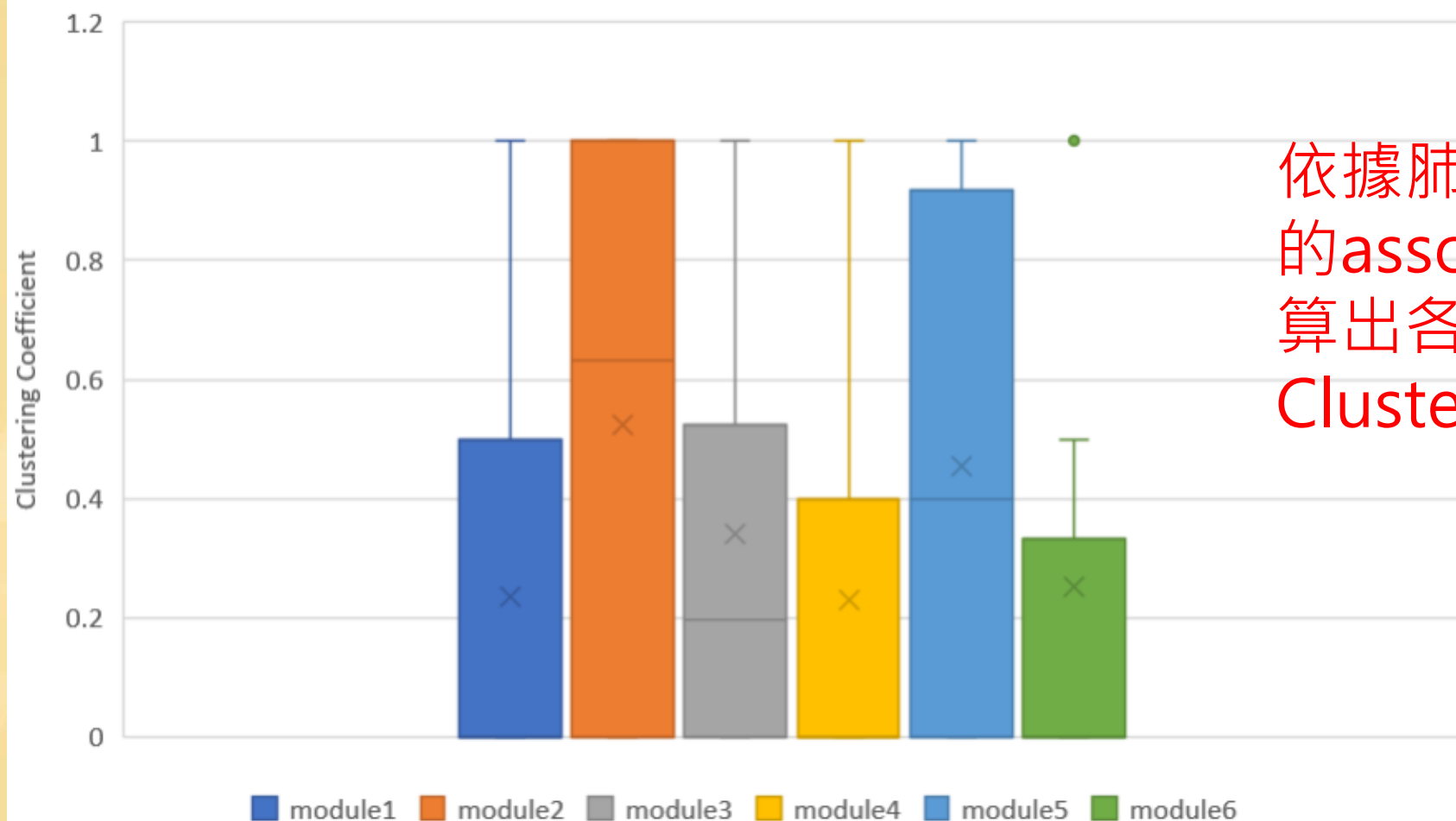
各module的Betweenness Centrality分佈圖



依據肺鱗狀細胞癌(LUSC)
的association network，
算出各個模組的
Betweenness Centrality

研究結果三:各基因網路模組 Clustering Coefficient分佈圖

各module的Clustering Coefficient分佈圖



依據肺鱗狀細胞癌(LUSC)
的association network，
算出各個模組的
Clustering Coefficient



結論

- 一. 參考肺腺癌(LUAD) 中長鏈非編碼核糖核酸對(lncRNA-lncRNA pair) 的皮爾森相關係數(PCC)的篩選比例729/6011778($PCC \geq 0.7$)，得到肺鱗狀細胞癌(LUSC)的 $PCC \geq 0.288401$ 。
- 二. 利用Cytoscape建立肺鱗狀細胞癌(LUSC)的association network，如結果圖所示。
- 三. 從肺鱗狀細胞癌(LUSC)的association network中挑選出六個 lncRNA較多的模組，並利用cytoscape中的分析工具，得到各基因網路模組每個節點的級數(degree)、中間性(betweenness centrality)和聚類係數(clustering coefficient)如結果圖所示。

未來展望

- 探討兩種癌症在相同比例下，PCC差距甚大的原因 (LUAD:0.7,LUSC:0.288401)
- 用Weighted Gene Correlation Network Analysis(WGCNA)(註8,9,10)找出每個基因模組中的中樞(hub)，最重要的長鏈非編碼核糖核酸(lncRNA)。



未來展望

- 利用本研究的肺鱗狀細胞癌(LUSC)基因網路模組，找出重要的長鏈非編碼核糖核酸(lncRNA)，作為判斷肺癌的標準或是治療的標靶。
- 可針對網路模組中的co-expressed mRNA 做 functional enrichment，如此能推測網路模組代表的意義
- 根據級數(degree)挑出幾個中樞(hub)做存活分析(使用cox proportional hazards model)(註11,12,13)



參考資料

Εύρηκα

1. Ono, H. et al. Long noncoding RNA HOTAIR is relevant to cellular proliferation, invasiveness, and clinical relapse in small-cell lung cancer. *Cancer Med.* 2014;3: 632–642.
2. Liu, X. H. et al. The long non-coding RNA HOTAIR indicates a poor prognosis and promotes metastasis in non-small cell lung cancer. *BMC Cancer.* 2013;13, 464.
3. Nakagawa, T. et al. Large noncoding RNA HOTAIR enhances aggressive biological behavior and is associated with short disease-free survival in human non-small cell lung cancer. *Biochem Biophys Res Commun.* 2013;436, 319–324.
4. Wu CH, Hsu CL, Lu PC, Lin WC, Juan HF, Huang HC. Identification of lncRNA functions in lung cancer based on associated protein-protein interaction modules. *Sci Rep.* 2016 Oct 27;6:35939.
5. Juan I Fuxman Bass, Alos Diallo, Justin Nelson, Juan M Soto, Chad L Myers & Albertha J M Walhout. “Using networks to measure similarity between genes: association index selection”. *Nature Methods.* 2013;10(12):1170-1171.

參考資料

Εύρηκα

6. Hsueh-Fen Juan & Hsuan-Cheng Huang. Gene Network Construction for Molecular Regulation. In Hsueh-Fen Juan & Hsuan-Cheng Huang (Eds), Systems biology : applications in cancer-related research. Singapore ; New Jersey : World Scientific. 2012(chap. 2, pp. 18).
7. Albert-László Brabási* & Zoltán N. Oltvai†. NETWORK BIOLOGY: UNDERSTANDING THE CELL'S FUNCTIONAL ORGANIZATION. NATURE REVIEWS : GENETICS . 2004; 5, 101-102.
8. Peter Langfelder & Steve Horvath. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008 ;Retrieved from <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-559>
9. Alexandra Maertens, Vy Tran, Andre Kleensang and Thomas Hartung. Weighted Gene Correlation Network Analysis (WGCNA) Reveals Novel Transcription Factors Associated With Bisphenol A Dose-Response. Frontiers. 2018;Retrieved from <https://www.frontiersin.org/articles/10.3389/fgene.2018.00508/full>.



參考資料

Εύρηκα

10. Peter Langfelder, Paul S. Mischel, and Steve Horvath. When Is Hub Gene Selection Better than Standard Meta-Analysis?. The National Center for Biotechnology Information. 2013; Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3629234/>.
11. Ru He & Shuguang Zuo . A Robust 8-Gene Prognostic Signature for Early-Stage Non-small Cell Lung Cancer. Frontiers. 2019;Retrieved from <https://doi.org/10.3389/fonc.2019.00693>.
12. Alboukadel Kassambara. Survival Analysis Basics. Retrieved from <http://www.sthda.com/english/wiki/survival-analysis-basics>
13. Alboukadel Kassambara. Cox Proportional-Hazards Model. Retrieved from <http://www.sthda.com/english/wiki/cox-proportional-hazards-model>



感謝

Eύρηκα

- 台大生科系 阮雪芬教授
- 李律 醫師
- 姚志鴻 高君陶 導師
- 生物 魏宏仁 老師
- 辛苦的父母親
- 兩班的各位同學

