

# 行政院國家科學委員會專題研究計畫 成果報告

## 從基因網路的次序結構變異探索癌症相關基因群 研究成果報告(精簡版)

計畫類別：個別型  
計畫編號：NSC 95-2221-E-468-006-  
執行期間：95年08月01日至96年07月31日  
執行單位：亞洲大學生物科技學系

計畫主持人：張培均

計畫參與人員：博士班研究生-兼任助理：石貴中  
碩士班研究生-兼任助理：陳彥州、李崇鴻

處理方式：本計畫可公開查詢

中華民國 96年10月31日

## 從基因網路的次序結構變異探索癌症相關基因群

計畫類別： 個別型計畫  整合型計畫

計畫編號：NSC95-2221-E-468-006-

執行期間：95年8月1日至96年7月31日

計畫主持人：張培均

共同主持人：

計畫參與人員：陳彥州、李崇鴻、石貴中

成果報告類型(依經費核定清單規定繳交)： 精簡報告  完整報告

本成果報告包括以下應繳交之附件：

- 赴國外出差或研習心得報告一份
- 赴大陸地區出差或研習心得報告一份
- 出席國際學術會議心得報告及發表之論文各一份
- 國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢

涉及專利或其他智慧財產權， 一年 二年後可公開查詢

執行單位：亞洲大學生物科技與生物資訊學系

中華民國 96 年 10 月 31 日

## 英文摘要

Keywords : cancer, microarray, gene ordering network, differentially expressed gene, fold-change, t-test, Mann-Whitney-Wilcoxon-Rank-Sum-test.

One of the main objectives in the microarray analysis is the identification of cancer-related genes. The earlier approaches such as simple fold-change, t-test, and Mann-Whitney-Wilcoxon-Rank-Sum-test are all based on the gene with differential expression in a tumor-versus-normal comparison. The fold-change method obvious disadvantage that it dose not provide a significance estimate for the observed changes and that the necessary cutoff values are essentially arbitrary. The t-test and the Mann-Whitney-Wilcoxon-Rank-Sum-test are more robust than the fold-change method, but it lack the biological meaning regarding the statistic significance. The evolution of cancer is the result of mutant accumulation that causes the variation of gene regulation network. The gene regulation network of one cell state corresponds to one gene ordering network. We have developed a novel method based on gene ordering network to discover the cancer-related genes that bring significance changes of gene network in a tumor-versus-normal comparison. With our method, the number of cancer-related genes is much more compares with the results from the other screening methods. However, these significance genes need to identify in the future.

## 中文摘要

關鍵詞：癌症，微陣列晶片，基因次序網路，差異表現基因，倍數法，t-檢定，Mann-Whitney-Wilcoxon 序數和檢定。

癌症相關基因的確認是微陣列分析的主題之一，早期的方法像簡單倍數法、t-檢定以及 Mann-Whitney-Wilcoxon 序數和檢定，均是針對癌症樣本對正常樣本實驗中相同基因的表現量差異作統計檢定，倍數法最明顯的缺點是，關於倍數臨界值的決定過於任意而缺少顯著性的估計，t-檢定以及 Mann-Whitney-Wilcoxon 序數和檢定雖然較穩健，然而，其統計上的顯著性卻缺少與之對應的生物意義。癌症的發生是基因突變累積的結果，以致於造成基因調控網路的變動，一個細胞狀態下的基因調控網路對應一個基因次序網路，根據基因在癌症樣本對正常樣本實驗中，是否會引起次序網路結構的重大改變，我們發展了一個新的癌症相關基因篩檢法。以此方法篩檢出的癌症相關基因數目明顯多餘其他方法所篩檢出的基因數目，無論如何，這些基因都有待進一步驗證。

## 研究目的

細胞的癌化是基因變異累積的結果，使得基因的表現或功能改變，進而細胞內整體基因間調控亦發生改變。癌細胞具有正常細胞所沒有的許多能力，例如；癌細胞的生長與細胞分裂是不受控制的，正常細胞在某些情況下會啟動自殺機制，癌細胞失去這種機制，癌細胞會刺激血管增生以獲得養分，癌細胞可以不斷做細胞分裂，癌細胞可以侵犯不同組織的其他器官，這些能力使癌症病人失去生命。

癌症相關基因是研究癌症的診斷、形成與治療的關鍵，近幾年來，基因微陣列技術被大量應用於癌症相關基因的研究，癌症相關基因的篩檢則是根據癌組織與正常組織間，基因的表現量是否有顯著差異，並以各種統計分析的方法來檢定，然而，統計上表現量差異不顯著的基因，卻不能斷定就一定與癌症無關。

基因表現量的次序結構，可以反應一個細胞的狀態，正常細胞以及不同類型的癌症細胞間，其基因表現量的次序結構皆不同。本計畫從基因表現次序結構變異的觀點，來發展一個全新的癌症相關基因篩檢方法。顯著的次序結構變異與顯著的表現量變異，我們將對這兩種篩檢癌症相關基因的觀點作分析與比較。

## 文獻探討

DNA 微陣列技術可以同時偵測數千甚至上萬個基因的表現，如此強而有力的工具被廣泛應用在癌症分類、癌症標記基因篩選、基因調控模型推論、功能基因體學等方面的研究(Winzler *et al.*, 1999; Chen *et al.*, 2001; Yu *et al.*, 2004; Morley *et al.*, 2004; Chen *et al.*, 2005)，其中一個最主要的應用，是去比較兩個不同細胞狀態的基因表現差異，例如；有特殊處理的細胞對未特殊處理的細胞，疾病組織對正常組織，突變細胞株對野生細胞株等等，這一類的研究，主要以統計學的方法去確認那些基因有明顯的表現差異，以下是常用的三種統計檢定基因有無表現差異的方法：

### 1. 倍數改變法 (Fold-Change)

這是早期 DNA 微陣列數據分析常用的方法(Schena *et al.*, 1995; DeRisi *et al.*, 1997)，其判斷是否有基因表現差異的依據，是看基因在實驗組與對照組的表現量增加或減少，是否有超過一定的倍數 ( $h$ )，並以下式來檢定：

$$\left| \log\left(\frac{\overline{x_2}}{\overline{x_1}}\right) \right| > \log(h)$$

其中  $\overline{x_1}$  與  $\overline{x_2}$  分別代表基因  $x$  在實驗組與對照組的平均表現量，在一定的顯著性下， $h$  的值與該基因的表現度有關(Chen *et al.*, 1997)。

## 2. t-檢定法 (t-test)

兩群或兩樣本 (實驗組與對照組) 的 t-檢定法其公式如下：

$$T_e = \frac{|\overline{x_1} - \overline{x_2}|}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$S_p = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

其中  $n_1$  與  $n_2$  為兩樣本的個別實驗次數， $\overline{x_1}$  與  $\overline{x_2}$  分別代表基因  $x$  在兩樣本實驗中的平均表現量，t-檢定通常要求實驗數據呈現常態分佈，因此，原始基因表現量數據須做對數轉換 ( $x \rightarrow \log(x)$ )，使接近常態分佈，若觀察值為  $T_{e,obs}$ ，則 p-value 為  $\text{Prob}(|T_e| > T_{e,obs})$ 。因此，可以定義臨界 p-value 作為有否基因表現差異的判斷。以此為基礎並進一步改進的方法則有微陣列數據顯著性分析法 (SAM) (Tusher *et al.*, 2001)。

## 3. Mann-Whitney-Wilcoxon Rank Sum 檢定

Mann-Whitney-Wilcoxon Rank Sum Test 用來檢定兩群體對於某一變數有否顯著差異。假設兩樣本分別做了  $n_1$  與  $n_2$  次的實驗，若要檢定基因  $x$  在這兩個樣本中的表現有無差異，首先，將兩群實驗關於  $x$  基因的表現量混合排序，並依序分配序數 1~N， $N = n_1 + n_2$ ，接著計算第一群的序數總和  $R$ ，並計算下式的值：

$$\frac{\left| R - \frac{n_1(N+1)}{2} \right|}{\sqrt{\frac{n_1 n_2 (N+1)}{12}}}$$

上式的值對  $R$  的分佈接近標準常態分佈，因此，可以用來估計 p-value，若觀察的  $R$  值為  $R_{obs}$ ，則 p-value 為  $\text{Prob}(|R| > R_{obs})$ ，因此，可以定義臨界 p-value

作為有否基因表現差異的判斷(Chen *et al.*, 1997; Chambers *et al.*, 1999)。

以上為主要用於篩檢癌症相關基因的統計檢定法，皆是針對個別基因表現量的變化。

我們在發展癌症診斷方法的過程中，提出了以基因網路的次序結構來做癌症分類的方法(Liu *et al.*, 2005)。當兩個基因*i*與*j*在DNA微陣列上的信號強度，如果具有顯著次序關係；即次序關係係數 $r_{ij}$ 大於某一顯著性門檻值，則建立這兩個基因的連線：

$$\gamma_{ij} = 1 - \sum_{s=1}^N \frac{[x_{sj} - x_{si}]^+}{N(x_{\max} - x_{\min})}$$
$$[x_{sj} - x_{si}]^+ = \begin{cases} x_{sj} - x_{si} & x_{sj} > x_{si} \\ 0 & x_{sj} \leq x_{si} \end{cases}$$

上式中 $x_{si}$ 是實驗樣本*s*中基因*i*的表現信號強度； $N$ 為樣本總數； $x_{\max}$ 與 $x_{\min}$ 則為實驗樣本中最大與最小基因表現信號強度。

次序結構網路是一個有向圖(directed graph)，連線 $i \rightarrow j$ 意味著基因*i*的信號強度比基因*j*來得低，我們定義*i*與*j*的連線為*i*流出而*j*流入，若*i*與*j*的連線具雙向性，我們定義為對等狀態。所以當某一個基因有很高的由內向外連線，表示這個基因有相對較低的表現信號強度；當某一個基因有很高的由外向內連線表示這個基因有相對較高的表現信號強度。因為我們不知道此次序關係係數 $r_{ij}$ 的分佈，因此，門檻值的決定採用隨機取樣測試的過程來模擬樣本空間的分佈：(a) 隨機自各個實驗樣本中選取一對基因構成一組基因對，重複此過程 5000 次；(b) 對於這些隨機選取各組基因對，分別計算次序關係係數 $\gamma_p$  ( $p = 1, 2, \dots, 5000$ )；(c) 由上述可以獲得一個次序關係係數的分佈，以這個次序關係係數分佈的顯著水平 5% ( $P < 0.05$ )做為建構次序結構網路的門檻值。

在研究的過程中，我們發現，不同類型癌症的次序結構網路，不僅具有保守性，其節點連線分佈皆遵循幕次定律(power law)，這不僅表示不同的癌症有不同的次序結構性，例如：局部自我相似性(Gisiger, 2001)，也意味著這種結構性的不同是廣泛的，而非少數基因引起的差異(Strogatz *et al.*, 2001; Albert and Barabasi 2005)。事實上，蛋白質交互作用網路與代謝路徑網路也具有相同的特性(Jeong *et al.*, 2001)。因此，我們利用不同細胞狀態具有不同的基因網路次序結構，來比較分析，探討造成癌症次序結構不同於正常細胞的基因，而這些基因在這二種細胞狀態下，不一定有明顯的基因表現差異。

## 研究方法

### (1) 微陣列資料庫來源：

我們使用 Oncomine database 作為微陣列基因表現的數據來源(Rhodes *et al.*, 2004)，其中包含 6,732 個 DNA 微陣列實驗，約 7,080,000 筆基因表現數據。由於 Affymetrix 的晶片有較好的數據品質，我們將只採用 Affymetrix 的晶片所做實驗數據。

### (2) 癌症相關基因的篩檢方法-比較癌組織與正常組織或癌細胞與正常細胞 基因網路次序結構差異的可能性估計：

我們將考慮每個基因的流入、流出與對等三種狀況，以流出為例；假設在癌症組織樣本裡，基因  $g$  流出之連線分別連接基因群  $S_t$  的基因集合，而在正常組織樣本裡，基因  $g$  流出之連線分別連接基因群  $S_n$  的基因集合，若基因群  $S_n$  與  $S_t$  的基因個數為  $x$  與  $y$ ，微陣列的基因總數為  $N$ ， $S_t$  與  $S_n$  的交集基因數為  $k$ ，利用超幾何分佈應用於估計兩個不同類別事件關連性的觀念(Bussemaeker *et al.*, 2001)，則基因  $g$  在兩種組織內的連線狀態關連性的 p-value 為：

$$p = \sum_{i=k}^{\min[x,y]} \frac{C_i^y C_{k-i}^{N-y}}{C_k^N}$$

令基因的流入、流出與對等三種狀況的 p-value 分別為  $p_i$ 、 $p_o$  以及  $p_e$ ，我們定義基因  $g$  在癌症組織樣本與在正常組織樣本裡，造成網路次序結構差異的可能性為：

$$P_s = p_i + p_o + p_e$$

$P_s$  的值越大表示差異性越大，對每個基因計算其  $P_s$ ，即可以估計每個基因對造成基因次序結構差異的貢獻程度。

## 結果與討論

本計畫篩檢出來的各種癌症相關基因如下表：



Cancer Type	Significance Gene
Pancreatic Cancer	AKR1C,AKR1C,ALDH1A,AMF,AMY2,AMY2,ANXA,ATP2B,BRWD,BZW,CCL,CD4,CEACAM,CLP,COL1A2,CPA1,CPA2,CPB1,CSDA,CSDE1,CTRB1,CTRB1,CXCL1,DICER1,DUOX2,EDN3,EGLN1,ELA2A,ELA3A,ELA3B,FBXO11,FBXO21,GAPDH,GATA4,GATM,GCG,GCLM,GNGT1,GP2,GPR20,H19,HIF1A,HMGA1,HNRPC,IAPP,IGF2,IGHA1,IGHG1,IL8,KLF6,KRT17,KRT18,KRT19,LAMC2,MALAT1,MAP4K4,MGAT4B,MRLC2,NFE2L2,NRBF2,NRIP1,NTAN1,PGLS,PLS3,PNLIP,PNLIPRP1,PNRC1,PRKAR2A,PRSS1,PRSS2,PSAP,RAB11A,RAB1A,RAE1,RANBP5,REG1A,REG1B,RHOQ,ROCK1,RPL22L1,RPL7A,RPS16,RPS2,RPS3,SAT1,SERPINA3,SGMS1,SLC3A2,SOD2,SQSTM1,STX16,SYPL1,TFPI2,TMED2,TMEM51,TNFRSF19,TUG1,TXN,USP12,USP3,USP34,USP6NL,VIM,VTA1,WWP1,YWHAB,YWHAQ,YWHAZ,ZNF266,ZNF395,ZNF451,ZNF580
Liver Cancer	PLEK,C11orf54,UQCRB,DUSP9,ZNF517,FLNA,CRLF1,SART1,OGDH,NCAPH2,GPT2,NBPF14,OBSL1,DAK,DDX54,MAP7D2,LARS2,PEO1,EIF4G1,TIRAP,TADA3L,IL1B,TMEM8,CHTF18,MYO18A,RBCK1,NINJ1,SETD2,NPLOC4,CREB3L2,FRAT2,ZNF259,SCAMP5,APEH,FZD2,TRPV2,NXN,GBF1,BMP4,TCEA3,TUBB3,NR2F6,NKD1,SNAPC2,ZNF426,RGNEF,SMTN,RANGAP1,USP4,NR5A1,IMPA2,CHPF,PCYOX1,DNAJC5,TBC1D17,MGC70857,EHMT2,SPG7,EP400,PPP2R5D,VIL1,EDC3,BCL2L1,AKT1,FASN,M6PRBP1,HDAC7A,CHST3,CHAF1A,FKBP8,NUDC,NAP1L4,HDAC10,ASPSR1,DHRS2,FAM129B,LOC93622,MYH9,MRPS26,GSK3A,RIC8A,TRRAP,MAEA,FLJ10769,CLK3,NFATC3,TNIP1,FGFR1,GDF15,TGFB1,RAGE,ITGA3,UBE2O,GORASP2,GSTT1,TRIOBP,IGF2,KIAA0323,PLAU,BAIAP2,BYSL,LSS,PHGDH,PHC2,LONP1,CSNK2A2,FSCN1,BFSP1,FLNC,PGAM1,ANXA11,EIF3S8,HDLBP,PPM1G,HYOU1,GRN,PLK2,CAPN1,VAT1,SAAL1,PTRF,GAMT,AP1M1,SHAHBP1,POLDIP3,COPA,MYO1C,GAPDH,GNB1,GLTSCR2,OAZ1,SREBF1,MYL9,PLAT,PRPF8,TRIM28,KEAP1,CAPN2,MAL2,HMG20B,DAP,C18orf24,TCF7,ALDOA,CALM3,GAPDH,MCM5,QPRT,SLC25A11,ENOSF1,PKM2,BRD4,PDXK,SNRPD3,GNB2,RNH1,PSMC3,SNRPB,FER1L3,USP11,RPS2,THBS1,CDT1,SNX5,CNN2,DHX30,KRT18,TPM1,TGM2,YIPF3,C20orf149,MSN,NT5DC2,COPE,STK25,IGF2BP1,CST3,RPL10A,CTNBL1,MAP4,KRT19,FARSA,ADAR,SND1,SCD,CPNE1,LMNA,PLK1,SLC2A1,CDC20,TAGLN,S100A16,KRT8,GDI2,MAGED1,COTL1,IGF2,SF3B2,EPB49,ARHGDI,CCNG1,TXNDC5,SERPINE1,TUFM,LASP1,PYCR1,ARF1,BAT3,CTNNA1,RPS9,ACTG1,DDX21,IGFBP3,BSG,ADFP,ACTN1,HLA-A,CCT7,AHCY,CAV1,ANXA2,SLC25A39,CLTC,IMPDH2,VIM,VCP,PRMT5,ENO1,SLC25A3,HMGA1,TPX2,HSP90AB1,PECI,HLA-B,TUBB2C,PHB,P4HB,SPARC,EEF1G,RPL4,TUBB,FN1,EIF4A1,RPS2,ANXA2P2,SLC25A6,CFL1,FTL,TF,RPL27A,RPL27,RPS24,RPL37A,RPS23,RPL19,APOA1,MYL6,CLU,SLC25A5,APOA2,PSMA7,RPS15,B2M,RPL21,TMSB10,VTN,RPL36,LOC729236,RPLP1,APOB,SERPINA1,LGALS1,RPL31,S100A6,RPS20,RPL32,H2AFZ,HMGB1,COX7C,SDHD,RPS11,RPL30,APOH,RPL37,RPL41,TTR,SNRPD2,ATP5O,ATF5,NDUFS5,COX6B1,SHFM1,RPL39,TDO2,RPL35A,ANG,SNRPG,DC2,HP,FGA,ATP5E,RPL35,HINT1,UQCRQ,NDUFA4,SERF2,RPS21,RPL34,RPL36A,COX7A2,TXN,CCDC72,RPS29,TMSB4X,SEPP1,RPL38,HSPE1,SERPINA3,RPS19,FGG,ALB,RPL41,HBA2,ACSL1,UCRC,AFP,COX17,HBG2,SERPINC1,NDUFA6,APCS,ITIH4,COX7B,CYP2E1,H19,CFHR1,C11orf10,C14orf2,NOLA3,MT1G,PLGLB2,H3F3A,ATP5I,SERPING1,CPS1,ALDOB,CRP,C14orf156,ORM1,HBA1,MEG3,C4B,SEC61G,RPS10,POMP,COX8A,SNRPE,

	TCEB2,UBL5,PAH,ATP5J,USMG5,FGL1,MALAT1,C9,ATP5J2,RPS27,HSPC152,AZGP1,SUB1,CFHR2,MGST1,PLG,CFB,FGB,HLA-C,CP,ATP5L,ANGPTL3,HRSP12,GC,HPX,COX6A1,RPS28,RPS26,SAA2,HBG1,ITIH3,FBXO10,FABP1,SAA1,COX6C,ALAS2,LOC646949,RPS27A,NDUFB2,UQCRB,EIF4G1,SPDEF
Prostate Cancer	IGFBP1,EDC3,ITGA3,CRISP3,EDEM2,EMILIN2,AYTL2,SHC1,G6PD,ALDH3B1,DAK,IDH2,STK25,QSOX1,LOC648232,NPLOC4,POMGNT1,BTBD2,C19orf6,OSGIN1,FAM38A,ZNF552,COG4,RXRA,EEF1A2,TPBG,FHOD1,ITGB2,WNT7B,TNFAIP2,ZNF629,LOC129607,SLCO4A1,MGAT4B,EFHD2,KRT14,PXN,GIPC1,AKT1,SEMA4B,EIF4G1,FLNA,GPR107,ZYX,MUC1,SLC16A3,PVRL1,SMG7,KLHL5,PSAP,C19orf48,AGRN,NUCB1,LMNA,PSMC3,LDLR,KRT5,DAP,SYNGR2,LOC728689,TGM2,PDIA4,GRN,TXNDC5,GRINA,CTSD,TRIM28,ATP1A1,WBP2,VCP,RPS2,S100A16,TPM3,SPINT2,GAPDH,YWHAQ,BSG,CALR,EEF2,PGAM1,P4HB,KRT8,ALDOA,SLC25A3,CTSB,C15orf15,UBE2V2,PKM2,FTL,RPL7,RPL19,PTMA,RPL27,RPLP1,LDHB,RPL35,TPI1,MIF,RPL21,ATP5L,SNRPD2,DBI,UBL5,S100A11,TCEB2,TMSB10,SHFM1,TMSB4X,UBB,B2M,HSPE1,SERF2,H2AFZ,ARHGDIB,COX7A2,ATP5E,RPS27L,CALM2,HMGB2,HLA-DRA,PSMA2,ERH,MRLC2,C14orf156,ATP5J2,SSBP1,SLC25A5,TXN,MRCL3,CCDC72,HINT1,HSBP1,FBXO10,ATP5J,RPS27,VDAC1,UCRC,ATP5C1,ATP5G3,MALAT1,UQCRQ,ORM1,SNRPG,TCEAL4,LOC51255,NDUFA12,COX17,NDUFA6,COX7A2L,APOA2,PIP,SEMG2,SEMG1,TGM4,NDUFB2,BIRC3,BYSL
Breast Cancer	ACPP,ACTG1,AKT1,ALDOA,APOA2,APOE,ARF3,ARHGDIA,ARHGDIB,ARPC1B,ARPC3,ATP5E,ATP5G1,ATP5G2,ATP5I,ATP5J,ATP5J2,ATP5L,AYTL2,B2M,CALM2,CALR,CCDC72,CD74,CDC37,CFL1,CKS2,COMMD6,COX17,COX6C,COX7A2,COX7B,COX7C,CTSD,CYB561,CYC1,DAD1,DBI,EEF1A1,EEF1A2,EEF1D,EIF3S9,EIF4A1,EIF4G1,FABP5,FASN,FBXO10,FKBP8,FLNA,FTL,G6PD,GAPDH,GAPDH,H2AFZ,H3F3A,HINT1,HLA-DRA,HMGB1,HMGB2,HSBP1,HSPC152,HSPE1,IDS,IFI6,IMPDH2,KLK3,KRT18,KRT19,KRT4,KRT8,LMNA,MAFF,MIF,MRPL41,MRPS18C,MYBL2,MYH11,MYL6,NDUFA1,NDUFA12,NDUFA3,NDUFA4,NDUFA6,NDUFB2,NOLA3,OAS3,OAZ1,OBSL1,PIP,PKM2,POMP,PRDX1,PREX1,PSMA2,PSMA7,PTMA,PYGB,RAB1B,RANGAP1,RPL19,RPL21,RPL23A,RPL27,RPL27A,RPL30,RPL31,RPL32,RPL34,RPL35,RPL35A,RPL36,RPL36A,RPL36AL,RPL38,RPL39,RPL41,RPL41,RPL41,RPL6,RPLP1,RPS11,RPS14,RPS19,RPS2,RPS2,RPS20,RPS23,RPS25,RPS26,RPS27,RPS27L,RPS28,RPS29,RPS7,RPS8,RPS9,S100A11,S100A2,S100A6,SAP18,SCD,SEC61G,SERF2,SF3B2,SHFM1,SKP1A,SLC9A3R1,SNRPD1,SNRPD2,SNRPE,SNRPG,SNRPG,SON,SPDEF,SRM,ST14,SUB1,TBC1D16,TCEB2,TMSB10,TMSB4X,TMSB4X,TPI1,TSKU,TUBB3,TXN,UBB,UBC,UBL5,UCRC,UQCRQ,USMG5,VAT1,VIM,WISP2,ZNF259
Lung Cancer	A2M,ACADVL,ACTA2,ACTG1,ADAM15,ADAR,AGER,AKR1B10,AKR1C2,AKR1C3,AKT1,ALB,ALDH3A1,ALDH3B1,ALDOA,ALOX5AP,ANG,ANXA3,AP1M1,AP3D1,APOA2,ARF1,ARF4,ARHGDIA,ARHGDIB,ARMC6,ARPC3,ASAHI,ASCL1,ASS1,ATF4,ATOX1,ATP5C1,ATP5E,ATP5G1,ATP5G2,ATP5I,ATP5J,ATP5J2,ATP5L,ATP6V1G1,B2M,BANF1,BAT2,BAT2D1,BCKDHA,BCL2L1,BIRC3,BLMH,BLOC1S1,BRP44,CALM2,CAPN1,CBX3,CCDC72,CCT3,CCT7,CD52,CD74,CDC25B,CDC37,CDK4,CDKN1A,CENPB,CKB,CKMT1B,CLDN18,CNN3,COMMD6,COX17,COX4I1,COX6A1,COX6B1,COX6C,COX7A2,COX7A2L,COX7B,COX7C,CPLX2,CREB3L1,CTGF,CTSD,CTSH,CXCL5,CYB561,CYC1,DAD1,DBI,DBNL,DC2,DCN,DDIT4,DDX11,DDX39,DDX5,

<p>DLC1,EDEM2,EEF1A2,EEF1G,EFHD2,EFTUD2,EGR1,EIF3S8,EIF4A1,EIF4G1,ENO1,EPHX1, EPS8L2,ERH,ESRRA,FABP5,FAM129B,FAM83A,FASN,FBXO10,FCGRT,FEN1,FKBP8,FLJ12949, FLNA, FN1, FOXA2, FSCN1, FTL, FUS, G6PD, GAPDH, GAS6, GDI1, GIT1, GJA5, GLTSCR2, GNB1, GNB2L1, GOT2, GPS1, GREM1, GUK1, H2AFZ, H3F3A, HBA2, HBB, HDGF, HDLBP, HES6, HIGD1A, HINT1, HLA-C, HLA-DMA, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DRA, HLA-DRB1, HLA-DRB1, HLA-E, HMGA1, HMGB1, HMGB2, HMGN3, HSBP1, HSPBP1, HSPC152, HSPE1, HYPK, IFITM1, IGFBP3, IGHA1, IGI, IL8, IMPDH2, INPPL1, IRAK1, ITGA3, ITM2B, ITPK1, ITB, KCTD12, KIF1A, KRT18, KRT19, KRT7, KRT81, LAMP3, LGALS1, LGALS3BP, LMNA, LMNB2, LSM3, LSM4, LUM, LYZ, MAFK, MALAT1, MAP2K2, MAZ, MBD3, MBIP, MCM7, MDH1, MGAT4B, MGC11257, MGP, MRCL3, MRLC2, MRPL13, MRPL18, MRPL38, MRPL41, MRPS18C, MT2A, MTCH1, MYBL2, MYL6, MYLK, NCAPH2, NDUFA1, NDUFA12, NDUFA2, NDUFA3, NDUFA4, NDUFA6, NDUFB2, NDUFB3, NDUFB5, NDUFC1, NDUFS5, NDUFV1, NEDD9, NOL4, NOLA3, NPNT, NUDC, OAZ1, OGDH, ORM1, PDCD10, PFDN5, PFKL, PFKM, PGAM1, PGAM1, PGD, PHGDH, PIH1D1, PIP, PKM2, PKN1, PLAU, PLEK, PLK1, PLP2, POLR2K, POLR2L, POMP, PPIF, PPP1CA, PSMA7, PSMC1, PSMC3, PSMD2, PSMD4, PSME2, PTGES2, PTMA, PTTG1, PXDN, PYCR1, PYGB, QARS, RAB11B, RAD23B, RAD51C, RALY, RANGAP1, RBCK1, RBX1, RGS5, RIC8A, RNASE1, RNASEH2A, RNF31, RPL17, RPL19, RPL21, RPL27, RPL27A, RPL30, RPL31, RPL32, RPL34, RPL35, RPL35A, RPL36, RPL36A, RPL36AL, RPL37, RPL37A, RPL39, RPL4, RPL41, RPL41, RPL41, RPL41, RPL7, RPL7A, RPLP1, RPS11, RPS15A, RPS19, RPS2, RPS2, RPS20, RPS21, RPS23, RPS24, RPS25, RPS26, RPS27, RPS27, RPS27A, RPS27A, RPS28, RPS29, RPS9, RPS9, RUVBL1, RUVBL2, S100A11, S100A2, S100A9, SAE1, SAP18, SAT1, SCGB1A1, SCP2, SDHD, SEPP1, SERF2, SERPINA1, SERPING1, SETD2, SF3B4, SFN, SFRS2IP, SFTPA2, SFTPA2, SFTPB, SFTPC, SGTA, SH3BGRL, SH3BGRL3, SHFM1, SHMT2, SIAHBP1, SIRT2, SKP1A, SLC16A3, SLC25A3, SLC25A39, SLC25A5, SLC25A6, SMARCA4, SNRPB, SNRPD1, SNRPD2, SNRPE, SNRPF, SNRPG, SNRPG, SON, SPARC, SPDEF, SPP1, SQSTM1, SRXN1, SSBP1, STK25, STOM, SUB1, TACC1, TAGLN, TBL3, TBRG4, TCEB2, TCOF1, TESC, TFF1, TIMM10, TIMP2, TIMP3, TMSB10, TMSB4X, TncRNA, TNFSF10, TRAP1, TRIB3, TRIM16, TRIM16L, TRIM22, TRIM28, TSPAN13, TUBA1B, TUBA4A, TUBB2A, TUBB2B, TUBB2C, TUBB3, TUFM, TXN, TXNDC13, TXNIP, UBB, UBE2R2, UBL5, UCHL1, UCRC, UQCRB, UQCRB, UQCRH, UQCRO, USMG5, USP11, VARS, VCP, VPS29, WDR18, WDR34, YKT6</p>
---

我們篩檢出的癌症相關基因數目比其他方法所篩檢出來的明顯的多，有些是未經實驗驗證的，因此有待進一步實驗驗證或與其他方法做比較。

## 參考文獻

1. Albert, R. and Barabasi, A.L. (2005) Statistical mechanics of complex networks. *Reviews of modern physics*, **74**, 47-97.
2. Bussemaker, H.J., Li, H., and Siggia, E.D. (2001) Regulatory element detection using correlation with expression. *Nature Genet.*, **27**, 167-171.
3. Chambers, A., Angulo, A., Amaratunga, D., Guo, H., Jiang, Y., Wan, J.S., Bittner, A., Frueh, K., Jackson, M.R., Peterson, P.A., Erlander, M.G., and Ghazal, P. (1999) DNA microarrays of the complex human cytomegalovirus genome: Profiling kinetic class with drug sensitive viral gene expression. *J. Virol.*, **73**, 5757-5766.
4. Chen, J.J., Lin, Y.C., Yao, P.L., Yuan, A., Chen, H.Y., Shun, C.T., Tsai, M.F., Chen, C.H. and Yang, P.C. (2005) Tumor-associated macrophages: the double-edged sword in cancer progression. *J. Clin. Oncol.* **23**, 953-964.
5. Chen, J.J.W., Peck, K., Hong, T.M., Yang, S.C., Sher, Y.P., Shih, J.Y., Wu, R., Wu, C.W. and Yang, P.C. (2001) Global analysis of gene expression in invasion by a lung cancer model. *Cancer Res.* **61**, 5223-5230.
6. Chen, Y., Dougherty, E.D., and Bittner, M.L. (1997) Ratio-based decision and the quantitative analysis of cDNA microarray images. *J. Biomed. Opt.*, **2**, 364-374
7. DeRisi, J.L., Iyer, V.R., and Brown, P.O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, **278**, 680-686.
8. Dorogovtsev, S. N. and Mendes, J. F. F. (2002) Evolution of networks. *Adv. Phys.*, **51**, 1079–1187.
9. Gisiger, T. (2001) Scale invariance in biology: coincidence or footprint of a universal mechanism? *Biol. Rev.*, **76**, 161-209.
10. Gordon, G.J., *et al.* (2002) Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma. *Cancer Research* **62**, 4963-4967.
11. Jeong, H., Mason, S.P., Barabasi, A.L. & Oltvai, and Z.N. (2001) Lethality and centrality in protein networks. *Nature*, **411**, 41–41.

12. Liu, C.C., Chen, W.S., Yang, P.C., Chang, P.C., and Chen, J.W. (2005) Topological-based cancer classification technologies using microarray data. *Formosan journal of Medicine.*, **9**. 650-666.
13. Rhodes, D.R. et al. (2004) OCOMINE: a cancer microarray database and integrated data mining platform. *Neoplasia*, **6**, 1-6.
14. Rhodes, D.R., Kalyana-Sundaram, S., Mahavisno, V., Barrette, T.R., Ghosh, D., and Chinnaiyan, A.M. (2005) Mining for regulatory programs in the cancer transcriptome. *Nature Genet.*, **37**, 579-583.
15. Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, **270**, 467-470.
16. Strogatz, S.H. (2001) Exploring complex networks. *Nature* **410**, 268–276.
17. Tusher, V.G., Tibshirani, R., and Chu, G. (2001) Significance analysis of microarrays applied to ionizing radiation response. *Proc. Nat. Acad. Sci.*, **98**, 5116-5121.
18. Winzeler, E.A., et al. (1999) Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901-6.
19. Yeoh, E.-J., et al. (2002) Classification, subtype discovery, and prediction of outcome in pediatric lymphoblastic leukemia by gene expression profiling. *Cancer Cell*, **1**, 133–143.

## 計畫成果自評

本計畫已經找出較多的癌症相關基因，目前仍有待進一步驗證。

本計畫並未如預期完成全部的研究，主要原因在於研究生之訓練與相關知識的學習進度不如預期，因此，拖延了計畫進度。