

行政院國家科學委員會專題研究計畫 成果報告

總計畫：智慧型系統在卵巢癌晶片之分子演化與控制(II)

計畫類別：整合型計畫

計畫編號：NSC94-2213-E-009-123-

執行期間：94年08月01日至95年07月31日

執行單位：國立交通大學電機與控制工程學系(所)

計畫主持人：李祖添

計畫參與人員：蘇宗輝，李；忠諭，謝承儒，傅建嘉，余宗晏

報告類型：精簡報告

處理方式：本計畫可公開查詢

中 華 民 國 95 年 9 月 28 日

行政院國家科學委員會補助專題研究計畫 成果報告
 期中進度報告

智慧型系統在卵巢癌晶片之分子演化與控制-總計畫

計畫類別： 個別型計畫 整合型計畫

計畫編號：NSC94-2213-E-009-123

執行期間：2005年8月1日至2006年7月31日

計畫主持人：李祖添 國家講座

計畫參與人員：蘇宗輝，李忠諭，謝承儒，傅建嘉，余宗晏

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：交通大學電機與控制學系

中華民國 95 年 7 月 31 日

智慧型系統在卵巢癌晶片之分子演化與控制-總計畫

Intelligent Systems in Ovarian Cancer with Microarray Data for Molecular Evolution and Control

計畫編號：NSC 94-2213-E-009-123

執行期限：94 年 8 月 1 日 至 95 年 7 月 31 日

主持人：李祖添

職稱：教授

電電子信箱：president@ntut.edu.tw

執行機構：交通大學電機與控制學系

一、中文摘要

卵巢癌是目前常致死的婦科疾病之一。而其如何從正常細胞如何轉換到惡性細胞仍有待瞭解。生物晶片是依可提供大規模疾病研究的重要方法。在生物晶片中的信息，經過智慧型系統處理後，可提供有關診斷、病程預測及治療的規劃，同時也可瞭解腫瘤形成的過程。而透過電腦輔助分析及模擬來瞭解生化演進過程可以可觀地增進對細胞進化的動態行為及結構的瞭解。所以，本計畫積極地尋求如何將先進的智慧型系統運用在卵巢癌晶片之分子演化與控制，以求能夠減少卵巢癌死亡的發生，本計畫將以生物資訊軟體系統開發(總計畫)與卵巢癌基因晶片資料前處理與分析(子計畫一)來連絡各相關子系統的智慧型系統研究，其中本計畫所考慮的子系統有：卵巢癌的預測與分類(子計畫二)、動態函數聚集、聚集區域回歸分析(子計畫三)、基因網路模型與穩定度(子計畫四)和多重序列排比(子計畫五)，用以建構一套智慧型生物資訊分析系統，並且利用基因路徑分析及視覺化研究(子計畫六)去模擬驗證此智慧型生物資訊系統。

關鍵詞：卵巢癌生物晶片、智慧型系統、預測與分類、基因網路模型、聚集區域回歸分析、基因路徑分析、多重序列排比。

英文摘要

Ovarian cancer is one of the most common causes of death from a gynecologic malignancy. The mechanism of transformation from normal cells to malignant cells remains to be clarified. Microarray provides a tool to study the diseases by a large-scaled approach. The information of microarray with intelligent system analysis can assist the physicians on the diagnosis, prediction

of prognosis, therapeutic planning and also shed light on figuring out the pathways of tumor formation. Computer assisted analysis and simulation of biochemical pathways can improve the understanding of the structure and the dynamics of cell processes considerably. In this research, we shall focus on the study of some key technologies of advanced intelligent system in the ovarian cancer with microarray data for molecular evolution and control. The group project consists of seven subprojects as:

1. Ovarian Cancer Microarray data preprocessing and analysis.
2. Classification and Prediction of Ovarian Cancer Using Microarray Data based on Intelligent Systems.
3. Computational Intelligence Approaches on Dynamic Functional Clustering Algorithm, Clusters of Interval Regression Analysis and Distance-Based Outliers Detection of Microarray Data for Ovarian Cancer
4. Modeling and Stability of Gene Network based on Computational Intelligence.
5. A Hybrid Algorithm to Multiple Sequence Alignment for Ovarian Cancer with Microarray Data.
6. Genome-Wide Pathway Analysis and Visualization Using Gene Expression Data.

In this three-years group project, the effectiveness of the developed subsystems shall be verified by the simulated and experimental results in the intelligent systems in ovarian cancer with Microarray data for molecular evolution and control.

Keywords: Ovarian Cancer with Microarray; Intelligent Systems; Prediction and Classification; Modeling of Gene Network; Clustering of interval Regression Analysis; Gene

二、緣由與目的

人類基因序列解碼定稿，學者認為，「後基因體時代」即將來臨，人類一層層揭開基因的神秘面紗，基因資訊不僅能顯示疾病的分子遺傳機轉，也是發展診斷及治療方法的重要標的。目前已知癌症、慢性病的致病機轉都與多重因子有關，若能找到癌病變最初時的基因變化，便有機會找到預防癌症的方法。生物資訊學目前研究發展方向即著重於以整合方式將基因序列轉成生物知識。進而了解每一個基因在生命正常過程或病變機制中所扮演的善惡角色。然資料得到後之解讀，目前解讀方法眾多，因此本整合型計畫提出一系列的智慧型系統方法去解讀生物晶片資料，同時開發一軟體系統進行方便之判讀工具。目前這方面之研究以生物晶片可最快述得到結果。有關生物晶片資料到基因表現量資料之解讀如圖一所示：

Ovarian cancer is the second most common gynecological cancer and is the leading cause of death from gynecological cancers, except breast cancer, with an incidence rate of about 15 per 100,000 women in western countries (Gatta, Lasota et al. 1998). Ovarian tumors are classified as epithelial tumors, which arise from the coelomic epithelium and constitute 80-90% of all cases, or as tumors of germ cell or sex cord/stromal cells (Chuaqui, Cole et al. 1998). The cause of epithelial ovarian carcinoma is unknown. The incidence of ovarian cancer is highest in industrialized countries and is lowest in developing countries. The low survival rate is related to the high proportion of women who have spread of disease beyond the ovaries at diagnosis. Furthermore, ovarian cancer tends to recur, even in patients who achieve a complete response to primary treatment.

There have been attempts to screen women without symptoms to achieve early detection of ovarian cancer; measurement of serum CA 125 antibody concentrations and abdominal or transvaginal ultrasonography have been done in postmenopausal women (Niloff, Klug et al. 1984). However, the low sensitivity and positive predictive value of such screening is a cause for concern. Thus, CA125 is thought to be robust only in following the response or progression of the disease, but not as a diagnostic or prognostic marker (Meyer and Rustin 2000). Thus, there is

an important need for additional diagnostic and prognostic markers for this disease.

Screening for molecular markers by specific antibodies is currently in progress. Ovarian carcinoma-associated antigen (OCA) is isolated from ovarian-cancer patients. NB/70K is also tested for useful marker. However, both OCA and NB/70K are not as sensitive and specific as CA125.

Macrophage colony-stimulating factor (MCS-F) and OVX-1 presented complimentary screening ability combined with CA125. Large scale screening was initiated for this possibility. Lysophosphatidic acid (LPA) was shown as a candidate in a preliminary report. Studies complimentary effects with CA125 are underway.

本計畫的主要目的是

為了解癌基因在整個細胞層次上之演化地位與表現趨勢，方便早期篩選可疑之癌化細胞與進行治療。

三、研究方法及成果

本計畫以人體細胞之癌生物晶片資料，提出相關新的研究方法(例如 Intelligent Systems、Neural Networks, Evolutionary Computation, Support Vector Machine/Regression, Machine Learning, Soft Computing 等)在生物資訊上之技術，從驗證、新基因發現到預測、基因網路建構、分群、分類、序列排比、基因路徑分析、視覺化與整合軟體開發之一系列的方法在癌基因分子演化與控制之研究。有關本整合型計畫之相關子計畫如下：

子計畫一：卵巢癌基因晶片資料前處理與分析。

子計畫二：基於智慧型系統使用生物晶片資料在卵巢癌的預測與分類。

子計畫三：智慧型計算之動態函數聚集、聚集區域回歸分析與距離基準之離異點偵測應用於卵巢癌的微陣列資料。

子計畫四：基於計算智慧之基因網路模型與穩定度之研究。

子計畫五：混合式演算法於多重序列排比在卵巢癌晶片之應用。

子計畫六：利用基因晶片做基因路徑分析及

視覺化研究。

子計畫七：卵巢癌基因晶片分子標記之確認與癌生物學之研究。

另外視覺化與分析軟體技術是以物件導向語言，Perl and Matlab進行開發。軟體整合開發：以Matlab整個子計畫軟體，另外一方面採用物件導向之UML分析及以物件導向軟體進行整合開發。

本總計畫負責督導、問題協調與整合各子計畫及系統軟體之建立。有關本整合計畫之基本分配架構圖如圖二所示。

除了整合工作之外，本總計畫要以物件導向軟體做整合及軟體開發物件導向UML分析，因此在智慧型系統在卵巢癌晶片之分子演化與控制之生物資訊分析軟體建立之研究如下：

生物資訊學之視覺化方法，以物件導向語言，Perl and Matlab進行開發，並依其特徵可分為：(1)靜態的描述與動態的呈現，(2)一維的排序、平面的相對關係、立體的結構以至於更高維度之交互作用，(3)維度縮減方式與非維度縮減方式，與(4)實體呈現與概念的表現。相關視覺化研究工作可包括基因圖譜之視覺化、蛋白質結構之視覺化、反應路徑之視覺化、基因陣列表現之視覺化等。改進現有視覺化工具到研發新的視覺化技術。以物件導向語言，Perl and Matlab進行生物資訊學之分析軟體開發。主要是以知識庫為基礎之軟體設計 (knowledge-based software)：

發展高效能的基因註解軟體，設計高功能的軟體工具來作精確的基因註解。

生物資訊採礦軟體 (biological data mining)：應用各式工具發掘基因體中有意義之生物學資訊。

各類重要註解 (annotations)、SNP的研究、intron-exon splicing sites的研究、重複序列計算 (repeat sequences)、基因家族 (gene family)、各種預測軟體 (prediction algorithms)。

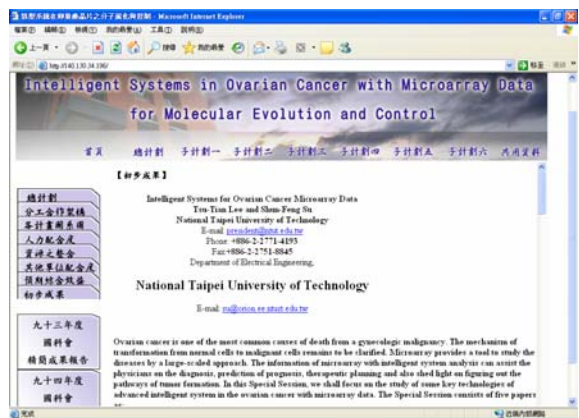
整合型實驗環境：

交通大學在電機與控制系將提供一間大型實驗室專供本計畫研發所需，本計畫雖為一跨校之整合型計畫，參與學校包括交通大學、

台灣科技大學、虎尾科技大學、宜蘭大學、華梵大學等校，但透過定定期開會及電子郵件商討，吾人將可確保本計畫可彼此配合順利進行。各子計畫所研製的各個系統將可整合到智慧型系統在卵巢癌晶片之分子演化與控制之軟體系統上，此軟體系統由伺服器完成相關畫面如下圖(a)到(e)所示：



(a) 伺服器系統畫面。



(b) 成果分享畫面。



(c) 以Perl建立之伺服器畫面。



(d) Web service建置。



(e) Perl生物計算建置。

並可在交通大學校區內實地測試各子計畫所開發之系統效能。當一新型系統設計完成後，為了要驗證其可行性與實用性，往往需要嘗試各種不同的測試與實驗，申請機構將提供一伺服器供各子計畫實地測試所將研發出來之系統效能，例如子計畫一、子計畫二、子計畫三、子計畫四、子計畫五、子計畫六、及子計畫七所開發出來之系統均將可安裝於伺服器實際驗證其效能。另外，如果要取得卵巢癌生物晶片資料，在申請機構將再提供另一伺服器存放由子計畫一所提供之資料，亦即建立一生物資訊資料庫網站，供各子計畫進行資料取得進行處理及上傳分析結果，提高各子計畫之交流，達到任何一子計畫一有結果，其他子計畫可透過此網站進行同步交流。另一些成果可參 T. T. Lee, J. T. Jeng and G. S. Huang, "Intelligent Systems in ovarian Cancer Microarray Data for Molecular Evolution and Control," Proceedings of 2006 CACS Automatic Control Conference, (Accepted), (2006)之成果發表論文。

四、結論

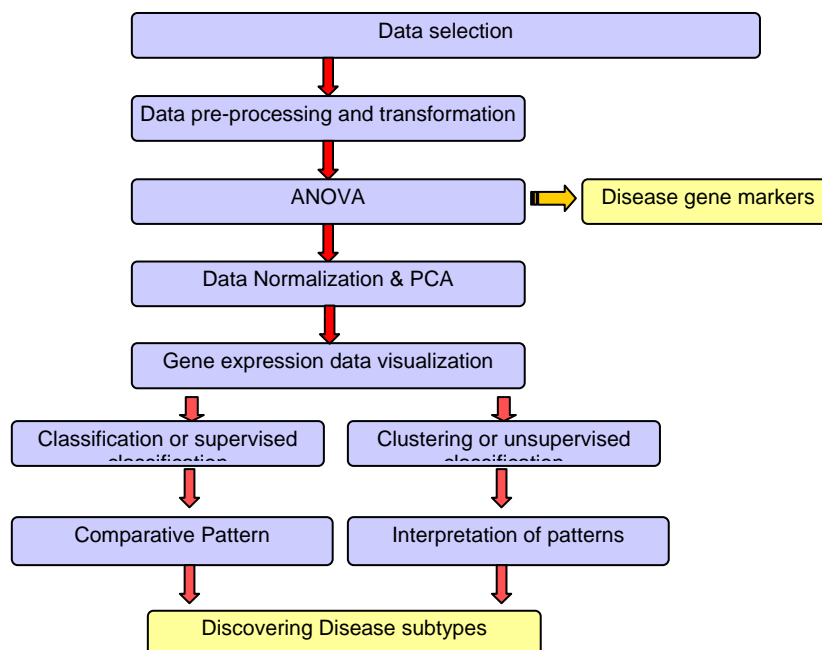
吾人已在交大成立「智慧型系統」研究中心，初期以研究符合人性化的前瞻性智慧技術為主。中長程重點則在研究智慧型生物資訊系統核心技術，並發展一個以交大及新竹科學園區為中心而建立的智慧生物資訊系統，提供各種技術、方法一個驗證與交流平台。這些結果，可為國內電子業、生物科技業及資訊業擬訂我國發展智慧型生物資訊系統策略的參考，以及相關單位制訂有關標準的主要參考。

生物資訊與生物資訊分析軟體是未來成長十分快速的產業。我國已建立完整的電子及資訊業的垂直整合及水平分工能力，以及優異的研發及製造能力，這些均是切入生物資訊與生物資訊分析軟體最佔優勢的地方。透過本計畫的實施，相信可為國內生物資訊、生物資訊分析軟體及資訊領域，建立自主性的技術，為國內相關業界走出一條路。此外，各分項計畫亦將有許多學術卓越的研究成果，預期這些成果均可在國際知名期刊發表。另外，相關技術，獲得突破後，也將申請專利。

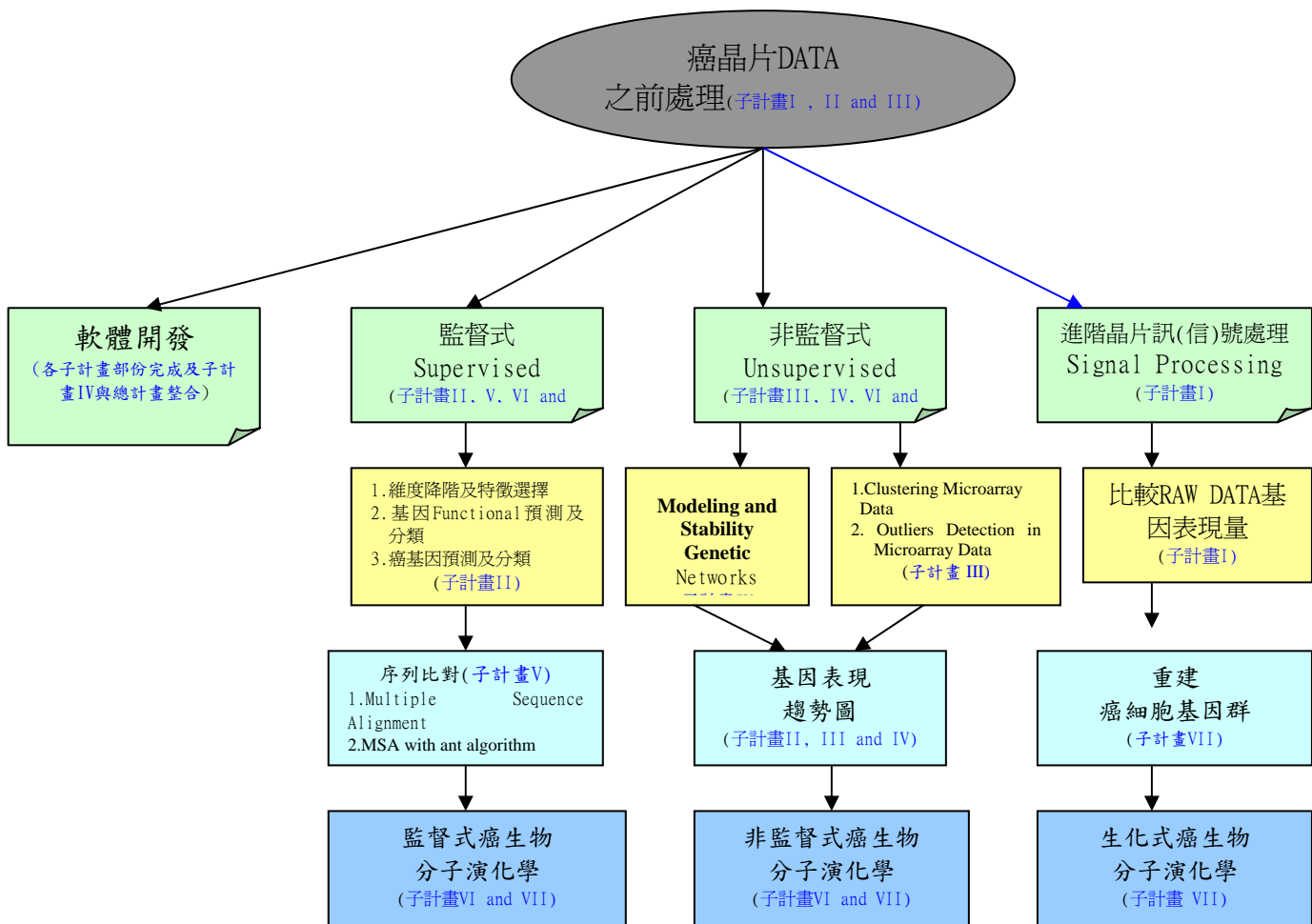
五、參考文獻

- Gatta, G., M. B. Lasota, et al. (1998). "Survival of European women with gynaecological tumours, during the period 1978-1989." *Eur J Cancer* **34**(14): 2218-2225.
- Chuaqui, R. F., K. A. Cole, et al. (1998). "Histopathology and molecular biology of ovarian epithelial tumors." *Ann Diagn Pathol* **2**(3): 195-207.
- Niloff, J. M., T. L. Klug, et al. (1984). "Elevation of serum CA125 in carcinomas of the fallopian tube, endometrium, and endocervix." *Am J Obstet Gynecol* **148**(8): 1057-8.
- Meyer, T. and G. J. Rustin (2000). "Role of tumour markers in monitoring epithelial ovarian cancer." *Br J Cancer* **82**(9): 1535-8.
- W.J. Ewens, G.R. Grant, *Statistical Methods in Bioinformatics: An Introduction*, Springer, 2001.
- 林仲彥等譯，生物資訊學—電腦技術，O'Reilly, 2002.
- G.B. Fogel and D.W.Corne, *Evolutionary Computation in Bioinformatics*, Morgan Kaufmann, 2003.
- Data mining in Bioinformatics 相關paper.

9. C.H. Wu, and J.W. McLarty, *Neural Networks and Genome Informatics*, Elsevier, 2000.
 10. A.D. Baxevanis and B.F. F. Ouellette, *Bioinformatics*, Wiley, 2001.
 11. R.C. Lee and W.M. Tepfenhart, *UML and C++ A Practical Guide to Object-Oriented Development*, Prentice-Hall, 2001.
 12. M. Nei and. Kumar, *Molecular Evolution and Phylogenetics*, Oxfard, 2000.
 13. G. Gibson and S. V. Muse, *A Primer of Genome Science*, Sinauer Associates, 2002.
 14. T. Speed, *Statistical Analysis of Gene Expression Microarray Data*, CRC Press, 2003.
 15. Amaratunga and J. Cabrera, *Exploration and Analysis of DNA Microarray and Protein Array Data*, Wiley, 2004.
- 本計畫之部分結果發表如下[16-22]：
16. Jin-Tsong Jeng, Yung-Cheng Lee, and Tsu-Tian Lee, "Classification of Ovarian Cancer based on Intelligent Systems with Microarray Data," 2005 IEEE International Conference on System, Man, Cybernetics, pp. 1053-1058, 2005.
 17. Chen-Chia Chuang, Jin-Tsong Jeng and Shun-Feng Su, "Dimension Reduction with Support Vector Regression for Ovarian Cancer Microarray Data," 2005 IEEE International Conference on System, Man, Cybernetics, pp. 1048-1052, 2005.
 18. G. S. Huang, Yao-Ching Hung, Ann Chen, Meng-Yen Hong, "Microarray analysis of ovarian cancer," *IEEE International Conference on Systems, Man, and Cybernetics*, pp. 1036 – 1041, 2005.
 19. Zne-Jung Lee, Chou-Yuan Lee, Kuan-Hung Liu, and Shun-Feng Su, "An Intelligent System for Multiple Sequences Alignment," *IEEE International Conference on Systems, Man and Cybernetics*, pp. 1042 – 1047, 2005.
 20. S. J. Wu, C. T. Wu, C. H. Chou and T. T. Lee, "Evolution-Based Gene Regulatory Network of Yeast Cell Cycle," *IEEE International Conference on Systems, Man and Cybernetics*, 2006. (Accepted)
 21. S. J. Wu, C. T. Wu and T. T. Lee, "Computation Intelligent for Eukaryotic Cell-Cycle Gene Network," *IEEE Engineering in Medicine and Biology Conference*, 2006. (Accepted)
 22. S. J. Wu, C. H. Chou, C. T. Wu and T. T. Lee, "Inference of Genetic Network of Xenopus Frog Egg: Improved genetic Algorithm," *IEEE Engineering in Medicine and Biology Conference*, 2006. (Accepted)
- 另外尚有一些成果正在審查中，未來會陸續發表於相關期刊及研討會上。



圖一 生物晶片資料分析流程圖



圖二 本計畫之基本分配架構圖。