Determination of Common Cancer-Related Genes Using Microarray Data
運用微陣列數據資料探討癌症之共同基因

Lillian Y. Chen, Yen-Chen Chen, Yuan-Chii G. Lee, Hung-Wen Chiu,
陳宇瑄、陳彥臣、李元綺、邱泓文

Graduate Institute of Medical Informatics, Taipei Medical University

臺北醫學大學醫學資訊研究所

Abstract

As cancer has drawn much of the attention worldwide these days, development of effective drugs using microarray data analytical results has become a popular method to search for drug targets. In order to determine common cancer-related genes for prostate adenocarcinomas, kidney carcinoma, hepatocellular carcinoma, ovarian adenocarcinomas, pancreatic adenocarcinomas and lung carcinomas, microarray datasets were analyzed via GeneSpring Version 6.1 based on their gene expressions. After having obtained 240 common cancer-related genes among six cancer types, a further analysis on the relationship of gene and biological pathway would be performed. MAPK signaling pathway and cytokine-cytokine receptor interaction have been found to have most genes involved in terms of the biological signaling process.

Keywords: *cancer, microarray, gene expression, MAPK signaling pathway, Cytokine-Cytokine receptor interaction*

Introduction

Cancer has currently become one of the deadly diseases affecting people's life. In Taiwan, cancer has ranked top cause of death since year 1982 as to that of second leading cause of death in the United States (3). As cancer has drawn great attention and focus in the medical field, many researchers around the world have dedicated their time in looking for a better cancer treatment. Until today, however, there have been no perfect medications being developed yet. Effective treatments on cancer patients basically rely on a better understanding of the tumour genes in relation to the specific cancer type. As a result, if researchers are able to define a group of cancer-related genes that are hidden in human bodies and waiting to be triggered, perhaps more helpful treatments can be developed based on those cancer gene targets will be available for all cancer patients. Previous understanding of gene expression levels in different cancer types by microarray hybridization have provided an idea that this is indeed a useful and eventually will be an essential method to identify possible biomarkers as well as drug targets (4, 5, 7, 11, 13). Many studies have completed their research in comparing gene expression profiles of cancer tissue to the corresponding normal tissue samples. Among all studies, Su *et al.* from the Genomic Institute of the Novartis Research Foundation (GNF) have published several well-known papers on the comparison of microarray gene expression for major cancer types as well as for normal tissue samples. For the purpose of this particular study, we picked six cancerous and normal microarray sample datasets obtained from Su's on prostate, kidney, liver, ovary, pancreas and lung carcinoma (9, 10). We re-analyzed and compared their gene expression in cancerous and normal tissue samples. We were looking for associated genes presented in six different cancer types. Moreover, we would like to find

the relationship between those associated genes and their biological pathways. The ultimate goal would be to define the most common pathway in which the majority of the associated genes have involved so that future drug development can be benefited from this finding.

Materials and Methods

Data Collection and Data Samples

Microarray data were obtained from public websites and/or given by the authors upon request. Tumour samples are composed of a total of 172 primary carcinoma sets, which include 31 prostate adenocarcinomas, 9 bladder/ureter carcinomas, 23 infiltration ductal breast adenocarcinomas, 19 colorectal adenocarcinomas, 13 gastroesophageal adenocarcinomas, 11 clear cell carcinomas of the kidney, 7 hepatocellular carcinomas, 25 serous papillary ovarian adenocarcinomas, 6 pancreatic adenocarcinomas, and 28 lung carcinomas. As to the normal microarray dataset, it contains 46 samples of human tissues, organs and cell-lines.

Data Analysis

Out of those available datasets, we have focused this study on six specific tumour types, which are the prostate adenocarcinomas, kidney carcinoma, hepatocellular carcinoma, ovarian adenocarcinomas, pancreatic adenocarcinomas and lung carcinomas. All datasets were presented in maximum hybridization intensity (AD) that has a very wide range of expression value patterns. The samples were first adjusted so that all values <20, including negative values, equal to a value of 20 by using excel VBA programming language (10). The adjusted datasets were then imported into a microarray analytical tool called GeneSpring Version 6.1 for further analysis. Each cancer specific dataset was first normalized to median among peer genes within its own tissue type, so was that for normal datasets. The normalized datasets were filtered to eliminate different expression level genes by having set the threshold to 1.5 fold. We would take answer of 1.5 fold from either the normal divided by cancer values, or vise versa. Moreover, the remaining genes would again be filtered by having compared with housekeeping genes mentioned in Warrington's paper (12). A list of 240 genes would be derived as a result of this filtration process. To classify the 240 genes in terms of their expression levels, we categorized each gene by doing hierarchical clustering (5). In addition, we also focused our attention on each gene's biological pathway by looking up on KEGG (http://www.genome.ad.jp/kegg-bin/mk_point_html). Figure 1 has shown an overall microarray data analytical process that we have followed to complete this study.

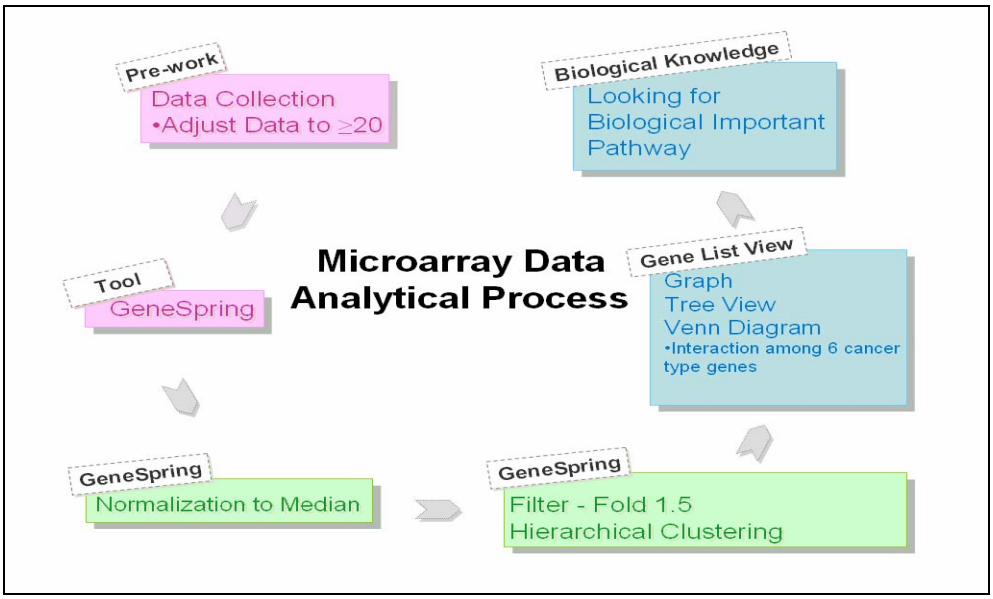

Figure 1: Flowchart for the process to analyze microarray datasets.

Result

After having filtered out genes that are not present in prostate adenocarcinomas, kidney carcinoma, hepatocellular carcinoma, ovarian adenocarcinomas, pancreatic adenocarcinomas and lung carcinomas, we illustrated gene-gene associations via Venn diagram (Figure 2). Figure 2a has shown that there are 903 common cancer-related genes among prostate, hepatocellular and lung carcinoma; Figure 2b has shown that there are 1941 common cancer-related genes among kidney, ovarian and pancreatic adenocarcinomas. As shown in Figure 2c, when having combined two groups of common genes altogether, only 240 genes have revealed equal importance in six cancer types. Since there are 240 common cancer-related genes among six cancer types, further analysis was required to categorize those genes based on their biological pathways. In total, more than fifty biological pathways have been confirmed to have one or more genes involved in their signaling process. However, there are two signaling pathways found to have more than 8 out of the 240 common cancer-related genes being involved, which are the MAPK signaling pathway and cytokine-cytokine receptor interaction. Table 1 and Table 2 have listed out genes that are involved in the two pathways individually along with a short description for each gene. In addition, Figure 3 has presented the MAPK signaling pathway obtained from KEGG website.

Table 1: Identified cancer-related genes involved in MAPK signaling pathway

Gene Name	Description
FGFR2	fibroblast growth factor receptor 2
PRKCBP1	protein kinase C binding protein 1
ARRB2	arrestin, beta 2
PDGFA	platelet-derived growth factor alpha polypeptide
PLA2G1B	phospholipase A2, group IB (pancreas)
PLA2G2A	phospholipase A2, group IIA
PRKCI	protein kinase C, iota
TP53	tumor protein p53 (Li-Fraumeni syndrome)
MAP2K3	mitogen-activated protein kinase kinase 3
CACNA1E	calcium channel, voltage-dependent, alpha 1E subunit
MAPKAPK5	mitogen-activated protein kinase-activated protein kinase 5
MAPKAPK2	mitogen-activated protein kinase-activated protein kinase 2

Table 2: Identified cancer-related genes involved in cytokine-cytokine receptor interaction

Gene Name	Description
CCR7	chemokine (C-C motif) receptor 7
CSF3R	colony stimulating factor 3 receptor (granulocyte)
EPOR	erythropoietin receptor
TNFRSF21	tumor necrosis factor receptor superfamily, member 21
CXCL1	chemokine (C-X-C motif) ligand 1
IL5	interleukin 5 (colony-stimulating factor, eosinophil)
IL13	interleukin 13
LIF	leukemia inhibitory factor (cholinergic differentiation factor)
PDGFA	platelet-derived growth factor alpha polypeptide
CCL15	chemokine (C-C motif) ligand 15
CCL25	chemokine (C-C motif) ligand 25
BLR1	Burkitt lymphoma receptor 1, GTP binding protein

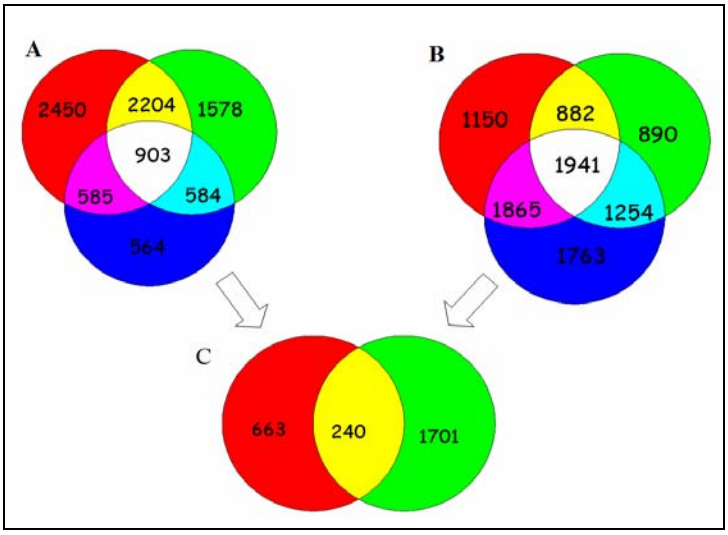

Figure 2: Interactions of genes among six cancer types. A: left circle = hepatocellular carcinoma gene numbers, right circle = lung carcinoma, bottom circle = prostate carcinoma gene numbers. B: left circle = kidney carcinoma gene numbers, right circle = ovarian adenocarcinomas gene numbers, bottom circle = pancreatic adenocarcinomas gene numbers. C: middle interaction ecliptic area = total cancer-related genes for six cancer types.

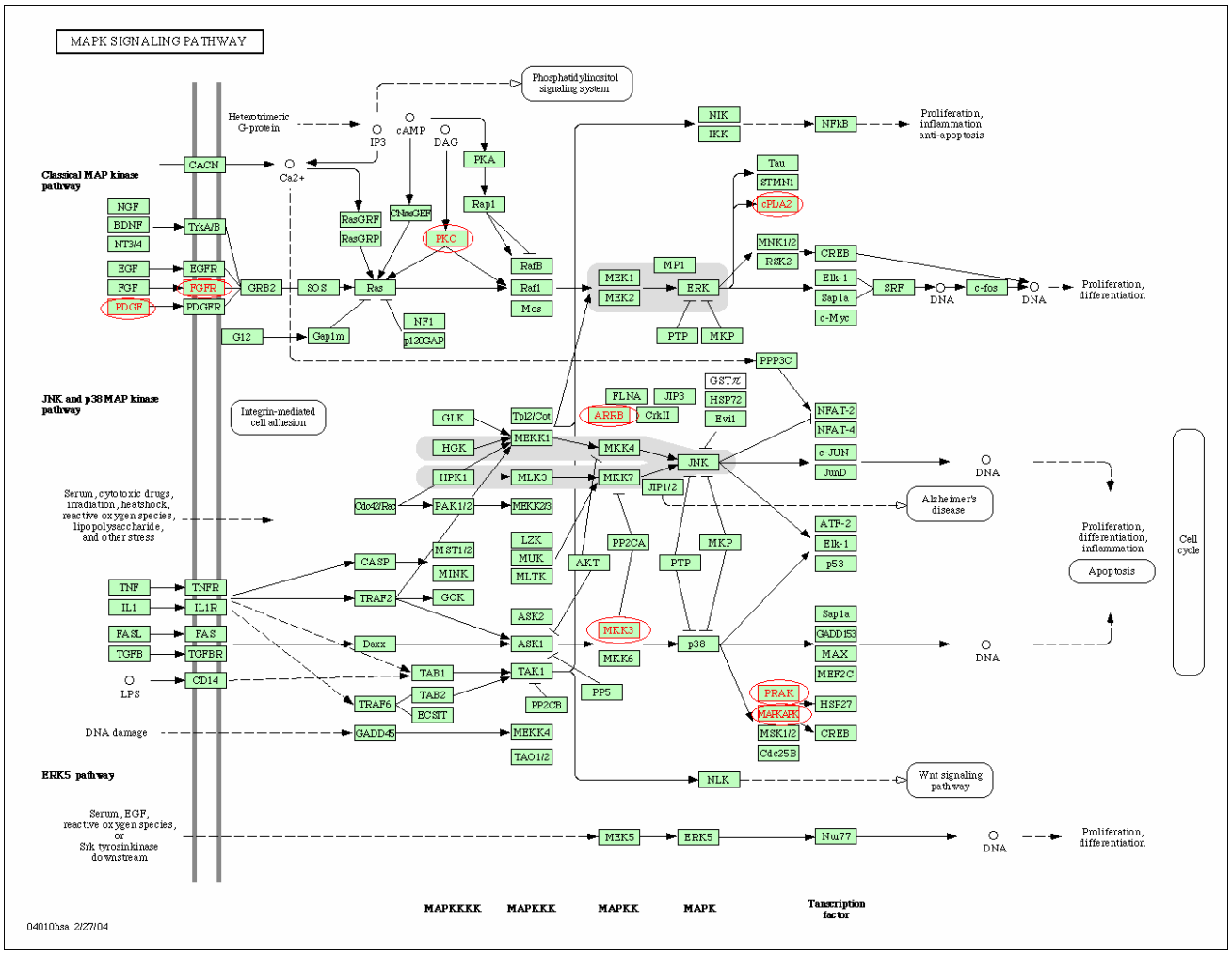

Figure 3: Pathway for MAPK signaling process. In total, eight cancer-related genes have been shown in circle. (*http://www.genome.ad.jp/kegg-bin/mark_pathway_www?23114/hsa04010.args/)*

Discussion

According to the search result of the biological pathways, it has shown that a great portion of the 240 genes is related to MAPK signaling pathway and cytokine-cytokine receptor interaction. MAPKs, or Mitogen-Activated-Protein-Kinases, are key signal transducting enzymes that are unique to eukaryotic organisms (1). MAPK signaling pathway is an important mechanism in a way that it not only transduces mitogenic signals from the cell membrane to the nucleus, but also controls gene expression, apoptosis, inflammation and cell proliferation. In addition, whether this pathway is functioning properly or not can also affect the pathology of various kinds of diseases, including chronic inflammation, heart disease, cancer etc (2). As to the cytokine-cytokine receptor interaction, it is a pathway containing many known cytokines that can either stimulate or suppress actions on immune responses since B cells, T calls macrophages, neutrophils etc. are participated in this process. Cytokines are peptide molecules that primarily regulate communication between immune system cells, and usually act locally by binding to the cell-specific cytokine receptor that is located in the cell membrane. The signal cascade starts in one cell, but eventually leads to obvious changes in target cell. Furthermore, another major function for cytokines is that they are also responsible for induce proper cell division in human (5). Having focused on the distinct functions of two signaling pathways, they are related either to the cell growth and death, or to the immune responses of the cells. As a result, accurate functionality of these two pathways has major effects on the human cell division (2, 5). It is very critical to have a complete and successful cell division in order to ensure the healthy production of human cells and to prevent possible defect cells entering the cell cycle; therefore, future drug development can focus on genes specifically in these two pathways, as they probably hold the promise to unveil the cancer causing mechanisms.

Conclusion

A total of 240 common cancer-related genes have been identified in six cancer types, which includes prostate adenocarcinomas, kidney carcinoma, hepatocellular carcinoma, ovarian adenocarcinomas, pancreatic adenocarcinomas and lung carcinomas. Among those genes, a great portion of genes is involved in MAPK signaling pathway and cytokine-cytokine receptor interaction.

Acknowledgement

I am greatly appreciated for Andrew I. Su from the Genomic Institute of the Novartis Research Foundation for providing the microarray datasets.

Reference

1. Chang, L. and Karin, M. (2001) Mammalian MAP kinase signalling cascades. *Nature*, 410(6824):37-40.
2. Gudermann, T., Grosse, R. and Schultz, G. (2000) Contribution of receptor/G protein signaling to cell growth and transformation. *Naunyn-Schmiedeberg's Archives of Pharmacology*, 361(4):345-62.
3. Ho, Y. (2004) Diabetes, pneumonia increases as causes of death. http://www.chinapost.com.tw /taiwan/detail.asp?ID=49759&GRP=B/
4. Liotta, L and Petricoin, E. (2000) Molecular profiling of human cancer. *Nature Reviews*, 1: 48-56.
5. Lu, Y. and Han, J. (2003) Cancer classification using gene expression data. Information Systems, 28: 243-268.
6. Nelson, D.L. and Cox, M.M. (2000) *Lehninger Principles of Biochemistry, Third Edition*. Worth Publishers: United States.
7. Rhodes, D. R., Yu, J, Shanker, K., Deshpande, N., Varambally, R. et al. (2004) Large-scale mata-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. *Proceeding of the National Academy of Sciences USA*, 101: 9309-9314.

8. Rhodes, D. R., Yu, J, Shanker, K., Deshpande, N., Varambally, R. et al. (2004) Oncomine: a cancer microarray database and data-mining platform. *Neoplasia*, 6(1): 1-6.
9. Su, A.I., Cooke, M.P., Ching, K.A., Hakak, Y., Walker, J.R. et al. (2002) Large-scale analysis of the human and mouse transcriptomes. *Proceeding of the National Academy of Sciences USA*, 99: 4465-4470.
10. Su, A.I., Welsh, J.B., Sapinoso, L.M., Kern, S.G., Dimitrov, P. et al. (2001) Molecular classification of human carcinomas by use of gene expression signature. *Cancer Research*, 61: 7388-7793.
11. Tsao, J.L., Tavare, S., Salovaara, R., Jass, J.R., Aaltonen, L.A. and Shibata, D. (1999) Colorectal adenoma and cancer divergence-evidence of multilineage progression. *American Journal of Pathology*, 154(6): 1815-1824.
12. Warrington, J.A., Nair, A., Mahadevappa, M., and Tsyganskaya, M. (2000) Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. *Physiological Genomics*, 2: 143-147.
13. Welsh, J.B., Sapinosa, L.M., Su, A.I., Kern, S.G., Wang-Rodriguez, J. et al. (2001) Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. *Cancer Research*, 61: 5974-5978.