行政院國家科學委員會專題研究計畫 成果報告

微型核糖核酸為致癌基因及其標靶物的蛋白質相互作用
研究成果報告(精簡版)

計畫類別: 個別型
計畫編號: NSC 97-2221-E-468-010-
執行期間: 97 年 08 月 01 日至 98 年 07 月 31 日
執行單位: 亞洲大學生物資訊學系

計畫主持人: 吳家樂
共同主持人: 李尚熾
計畫參與人員: 碩士班研究生-兼任助理人員: 王傑瑋

報告附件: 出席國際會議研究心得報告及發表論文

處理方式: 本計畫涉及專利或其他智慧財產權, 1 年後可公開查詢

中華民國 98 年 10 月 19 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

微型核糖核酸為致癌基因及其標靶物的蛋白質相互作用

計畫類別：□ 個別型計畫 □ 整合型計畫
計畫編號：NSC 97-2221-E-468-010-
執行期間：2008 年 08 月 01 日至 2009 年 07 月 31 日

計畫主持人：吳家樂
共同主持人：李尚熾
計畫參與人員：翁嘉偉、江長志、王傑瑋

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：生物資訊學系 亞洲大學
中華民國 98 年 10 月 19 日
ncRNAppl - A tool for identifying disease-related miRNA and siRNA targeting pathways
ABSTRACT

Summary: Currently, there are a number of databases which store microRNA (miRNA) information, and tools available which provide miRNA target prediction. In this paper we describe a novel web-based tool that integrate the miRNA-targeted mRNA data, protein-protein interactions (PPI) records, tissues, biochemical pathways, human disease and gene function information to establish a disease-related miRNA target pathway database. This database is unique in the sense that it links miRNA target genes with their PPI partners according to being tissue-specific, diseases-specific or both. The same approach is also applied to siRNA data. This database provides two types of searches; (i) tissue-specific, and (ii) disease-specific miRNA (or siRNA) targeting pathways. The search allows one to identify tissue-specific or disease-specific miRNA (or siRNA) target gene’s PPI partners two levels beyond.

The release version 1.0 is a freely accessible database available at http://ncRNAppi.cs.nthu.edu.tw, and http://ncRNAppi.bioinfo.asia.edu.tw/

INTRODUCTION

MicroRNAs (miRNAs) are a class of small non-coding RNAs (ncRNAs) that bind to mRNA and induce either translation repression or mRNA degradation. In the last few years, there have been reports, that miRNAs could cause cancers by acting as oncogenes and tumor suppressor genes. For example, a miRNA that targets the mRNA of a tumor suppressor would result in loss of that protective factor (Zhang et al. 2007). Small interfering RNAs (siRNAs) are small double-stranded RNA molecules that could inhibit gene expression by the RNAi mechanism.

In this study, we integrate the human miRNA-targeted (or siRNA-targeted) mRNA data, protein-protein interactions (PPI) records, tissues, pathways, and disease information to establish a disease-related miRNA (or siRNA) pathway database. We focused on cancer-related targets since it has been reported (He et al., 2005, Esquela-Kerscher & Slack 2006) that miRNAs could play the role of an oncogene (OCG) and tumor-suppressor-gene (TSG).

This database is set up to provide the following functionality, (i) for a given miRNA (or siRNA) ID, output the targeted miRNAs and its PPI partners, (ii) search for tissue-specific miRNA (or siRNA) and PPI partners, and (iii) identify disease-specific miRNAs (or siRNAs) and PPI pathways. Previously, two miRNA web services (Nam et al., 2008, Wang 2008) have been set up to supply functional annotation, but to the best of our knowledge, this database is the first one to address the relationships between miRNA (or siRNA), PPI, pathways, tissues and disease information. This tool could certainly assist users in identifying tissue- or disease-specific miRNA (or siRNA) targeting pathways.

The importance of this database can be understood in term of the regulation relationships between miRNA (or siRNA) and genes. For instance, if the upstream miRNA (or siRNA) is defective, its effect could be amplified downstream. As another illustration, given that a miRNA (or siRNA) targets gene A, which has two PPI partners, i.e. proteins B and C; and suppose that genes A and C are involved with the same disease, then it is highly probably that gene B is also related to the same disease.

DESCRIPTION OF INPUT DATASETS

The miRNAs and their target mRNAs information are retrieved from the TarBase v5 (Papadopoulos et al. 2009) database. PPI data is carried from BioIR (Liu et al. 2008), which integrate the major publicly available databases that contain literatures on PPI information for human proteins. BioIR is a PPI warehouse includes HPRD (Mishra et al. 2006), DIP (Xenarios et al. 2001, Salwinski et al. 2004), BIND (Alfarano et al. 2006), IntAct (Kerrien et al. 2007), MIPS (Mewes et al. 2008), MINT (Chatr-aryamontri et al. 2007) and BioGRID (Breitkreutz et al. 2008) databases.

Tissues, pathways disease and protein function information are carried from UniGene (Pontius et al., 2003), KEGG (Kanehisa et al. 2008), OMIM (Amberger et al. 2009) and Gene Ontology (Gene Ontology Consortium 2006) respectively, and they are parsed and linked to the PPI entries. The siRNA data are obtained from siRecords (Ren et al., 2006), which document siRNAs entries and their target mRNAs. Queries are retrieved by the siRecords IDs.

QUERY INTERFACE

To begin the query, user selects a miRNA (or siRNA) ID from the ID pull-down manual, and specifies the tissue type. The tool will return the miRNA (or siRNA) target genes and the nearest and the next nearest interacting protein neighbors. To trace the interaction partner of miRNA (or siRNA), the ncRNA node is treated as the initial node, then level one PPI is obtained by using BioGrid, these nodes are stored, and the next nearest PPI partners are obtained by performing PPI searches in BioGrid, hence, we obtained PPI nodes to the second level. Duplicate nodes are filtered out in this search. We do the search up to two levels only, more protein levels can be done if necessary.

User can also specify a cancer type (more than 30 types are available) associated with the miRNA (or siRNA) target gene, plus the nearest or the next nearest interacting protein neighbors. The platform will identify all the miRNA (or siRNA) target pathway.

SEARCH FEATURES AND ANALYSIS INTERFACE

The tool provides web-based data access and allows disease assignment for a specific node along miRNA targeting pathways. For example, a user selects an miRNA with the ID, let-7, checks the ‘OMIM Disease type for individual node’ box labeled with ‘Target’, and chooses the ‘OMIM Disease type for individual node’ box labeled with ‘Target’, and chooses the ‘OMIM Disease type for individual node’ box labeled with ‘Target’, then disease information recorded by OMIM will be retrieved by the platform. To narrow down the number of pathways, one can select ‘Yes’ under the “Common expression of target, Level-1 and level-2 nodes in KEGG” option, the results are shown in Figure 1.

Depending on the user’s choice, the pathways are ranked and displayed according to the Jaccrad index (p-value as well) for biological process or molecular function annotation.

User can further narrow down the output by checking the ‘Filter out receptor protein’ , ‘selection of transcription factor’ or ‘selection of cancerous protein’ box labeled with ‘Target’, ‘Level-1’ or ‘Level-2’. This operation allows the user to filter out receptor or transcription factor or cancerous protein node in the pathway. Cancerous genes information are obtained from the Tumor-Associated Gene database (Chan and Sun 2006). Similar searching method is applicable to siRNA.

The platform also provides disease-name keywords search in identifying disease-specific pathways, for instance, a user can type the word ‘breast’ in the ‘KEYWORDS’ box in order to do keyword search. We anticipate to regularly updating
ncRNAppi as long as a new version of TarBase or siRecords is released.

DISCUSSION

The tool ncRNAppi is set up by integrating human miRNA and siRNA target genes data, with the tissues, disease, pathways, cancerous genes and PPI information. By integrating those pieces of information, ncRNAppi provide a powerful tool for identifying cancer-related miRNAs or siRNAs. For instance, the tool allows the possibilities of predicting novel cancer genes through tissue or disease specific search. In summary, the database provides an easy means of investigating the regulatory role of miRNAs and siRNAs for cancer study.

REFERENCES


中文摘要

建置一個非編碼基因(微型核糖核酸及 siRNA)與其標靶物之蛋白質與蛋白質相互作用的資料庫，名為 ncRNAAppi。此資料庫整合了人類組織、生物網路、疾病及功能等訊息，從而建立一個查詢與癌症相關的非編碼基因資料庫。資料庫可作兩種功能：依(1)人類組織，及(2)疾病，查詢非編碼基因其標靶物之蛋白質與蛋白質相互作用。

Self-assessment

We have completed the major aims of the proposal, that is deriving the miRNA target genes’ protein-protein interactions partner up to the second level. A web-based service was set up which provide the PPI and pathway ranking services.

Our results are presented, either oral or poster presentations, in international conferences and local conferences.

Publications

期刊論文


國際性研討會論文或壁報


| 報告人姓名 | 吳家樂 |
| 时间 地點 | 13-16 Jan. 2009 Peking, PRC |
| 會議名稱 | (中文) 2009 年亞太生物資訊研討會 (英文) The Seventh Asia Pacific Bioinformatics Conference 2009 |
| 發表論文題目 | (中文) 利用蛋白質功能與蛋白質功能相關性預測蛋白質功能 (英文) Protein Function-Function Correlation Approach for Protein Function Prediction |

服務機構及職稱：亞洲大學生物資訊系教授
本會核定補助文號：NSC97-2221-E-468-010
報告內容應包括下列各項:
一、參加會議經過

Jan. 13
- 下列是我在會議中參加的談話:
  - Tutorial 1: Olga G. Troyanskaya: Systems biology based on integrated analysis of functional genomics data
  - Keynote Speech
    David Lipman: Molecular evolution and epidemiology of seasonal influenza
  - Session 2: Gene expression and microarrays

Jan. 14
- Session 3: Protein structure, location and function
- Poster presentation

Jan. 15
- Invited Session 2
  - Martin Vingron: Transcriptional regulation: computational methods, statistics, and coregulation
  - Michael Eisen: Understanding and exploiting the evolution of Drosophila regulatory sequences
  - Session 9: Gene expression, microarray, disease
  - Session 11: Networks and systems biology
  - Session 13: microRNAs and RNAi

Jan. 16
- Invited Session 3
  - John Mattick: A new understanding of the human genome
  - Session 17: gene expression, microarray data analysis

二、與會心得

The Asia-Pacific Bioinformatics Conference 2009 is an annual forum for exploring research, development and novel applications of Bioinformatics which is held at the International Convention Center, Tsinghua Science Park, from Jan. 13 to 16.

The scientific program of APBC 2007 included 3 keynote talks, 3 tutorials, 35 oral presentations, 112 poster presentations and a HP industrial sessions as well.

The symposium has received 204 papers and each submitted paper was reviewed by at least two program committee members. All accepted papers had at least 2 positive recommendations. The program committee accepts approximately 37% of papers, that is a total of 75 papers. A variety of papers was presented at this conference and the topics include DNA sequence analysis, gene regulation and regulation analysis, RNA structure and function, protein structures study, proteomics, biological pathways, disease and medical application, and evolution study.

I had a poster presentation on Jan. 14, 1:00 p.m. Title of my poster presentation is “Protein Function-Function Correlation Approach for Protein Function Prediction”. On the other hand, I had attended most of the talks during the four days conferences.

In my personal opinion, bioinformatics researches are growing very rapidly. During APBC05 there were 35 papers presented but the number is doubled to 75 papers this year. More and more researchers are working in the areas such as data integration, gene micro-array analysis, proteomics and system biology. There are several good tutorials, talks and posters presented in this conference. Attended talks with their titles are listed in below.
Jan. 13, 2009

Tutorial:
Systems biology based on integrated analysis of functional genomics data
Professor Olga G. Troyanskaya

The tutorial presented an overview of recently developed methods for integrated analysis of functional genomic data and outlined current challenges in the field. The focus was on the development and use of such methods for gene function prediction, understanding of protein regulation, and modeling of biological networks.

Keynote:
Molecular evolution and epidemiology of seasonal influenza
Prof. David Lipman

Session 2: Gene expression and microarrays

B1-04 Network-based support vector machine for classification of microarray samples.
By Yanni Zhu, Xiaotong Shen and Wei Pan.


Jan. 14, 2009

Session 3. Protein structure, location and function

D1-03 Predicting disordered regions in proteins using the profiles of amino acid indices. Pengfei Han, Xiuzhen Zhang and Zhi-Ping Feng.

D1-04 A method to improve protein subcellular localization prediction by integrating various biological data sources. Thai Quang Tung, Doheon Lee.

Poster session

PB2-01 zhenyu xu, Wu Wei, Julien Gagneur and Lars Steinmetz. Prevalent use of bidirectional promoters generates pervasive transcription in yeast


PC1-02 Characteristic analysis of miRNA precursors in metazoan species Xiaobai Zhang, Xiaofeng Song, Huinan Wang and Huanping Zhang.

PC1-05 Rapid Evolution of Mammalian X-linked Testis microRNAs Xuejiang Guo, Bin Su, Zuomin Zhou and Jiahao Sha.

PD1-22 Evolution of protein-protein interaction networks for plant bZIP transcription factors
Ying he, Grigoris Amoutzias and Yves Van de Peer.

PD2-24 CPSARST & CPDB - an efficient search tool and a comprehensive database of circular permutation in proteins
Wei-Cheng Lo, Chi-Ching Lee, Che-Yu Lee, and Ping-Chiang Lyu.

PE2-04 Identification of disease genes via network alignment of human interactome and
phenome
Xuebing Wu and Rui Jiang.

**PF1-04** An Approach to the Recovery of Associations between Protein Domains and Complex Diseases
Wenhui Wang, Rui Jiang and Yihui Luan.

**Jan. 15, 2009**
**Invited Session 2**
Martin Vingron: Transcriptional regulation: computational methods, statistics, and coregulation

Michael B. Eisen: Understanding and exploiting the evolution of Drosophila regulatory sequences

**Session 9. Gene expression, microarray data analysis and disease classification**

**B1-06** A statistical framework for integrating two microarray data sets in differential expression analysis. Yinglei Lai, Sarah E. Eckenrode and Jin-Xiong She.

**B1-07** Comparison of Affymetrix data normalization methods using 6,926 experiments across five array generations. Reija Autio, Sami Kilpinen, Matti Saarela, Olli Kallioniemi, Samps Hautaniemi, Jaakko Astola.

**B1-08** Integrative disease classification based on cross-platform microarray data. Chun-Chi Liu, Jianjun Hu, Mrinal Kalakrishnan, Haiyan Huang, Xianghong Jasmine Zhou.

**B1-09** Principal component tests: applied to temporal gene expression data. Wensheng Zhang, Hong-bin Fang, Jiuzhou Song.

**Session 11. Networks and systems biology**

**E2-02** Finding motif pairs in the interactions between heterogeneous proteins via bootstrapping and boosting. Jisu Kim, Kyungsook Han.

**E2-03** A new graph-based method for pairwise global network alignment. Gunnar W. Klau.


**Session 13. microRNAs and RNAi**

**C1-01** Predicting microRNA targets in time-series microarray experiments via functional data analysis. Brian J Parker, Jiayu Wen.

**C1-02** A structural interpretation of the effect of GC-content on efficiency of RNA interference. Chi Yu Chan, C. Steven Carmack, Dang D. Long, Anil Maliyekkel, Yu Shao, Igor B. Roninson, and Ye Ding.

**C1-03** Computational identification of condition-specific miRNA targets based on gene expression profiles and sequence information. Je-Gun Joung, Zhangjun Fei.


**Jan. 16, 2009**

**Invited Session 3**

John Mattick: A new understanding of the human genome

**Session 15. Association study and genomic variation**

**F2-09** Partial correlation analysis indicates causal relationships between GC-content, exon density and recombination rate variation in the human genome. Jan Freudenberg, Mingyi Wang, Yaning Yang and Wentian Li.

**F2-10** Minimizing recombinations in consensus networks for phylogeographic studies. Laxmi Parida, Asif Javed, Marta Melé, Francesc Calafell and Jaume Bertranpetit and Genographic Consortium.

**Session 17: gene expression, miicroarray data analysis**

**B1-10** Biclustering of gene expression data using reactive greedy randomized adaptive search procedure. Smitha Dharan, Achuthsankar S Nair.

三、考察參觀活動(無是項活動者省略)

無

四、建議

In summary, the symposium had a lot of discussions. The level and quality of the talks are very good. The organizer had done a very good job in organizing the conference.

五、攜回資料名稱及內容

資料名稱:
(1) Proceedings of the Seventh Asia-Pacific Bioinformatics Conference, and
(2) Conference Program booklet - The Seventh Asia-Pacific Bioinformatics Conference.

六、其他

無