MEDINFO 2007
HISA Ltd Australia
© 2007 The authors. All rights reserved.

Development of a Colon Cancer Extension for the AMANDA Microarray Management Database

Guillermo López-Campos[a], Oscar García-Hernández[a], Juan Pedro Sánchez Merino[a], Alejandro Romera-López[b], Beatriz Pérez-Villamil[b], Fernando Martín-Sánchez[a]

[a] *Medical Bioinformatics Dept. Institute of Health" Carlos III", Majadahonda. Spain*
[b] *Medical Oncology Dept. Hosp. Clínico San Carlos, Madrid. Spain*

Abstract and objective

The management of information is a key aspect during microarray experiments analysis and results interpretation. The common use of this technique with clinical samples requires an adequate clinical annotation of those samples and the inclusion of that annotation into microarray information management systems. We present a module for clinical annotation of colon cancer samples for microarray experiments. This module has been developed to be integrated with a previously developed microarrays database (AMANDA). This new module stores information related with the clinical annotation of samples and patient information used during the experiments such as patient evolution, chemotherapy or surgical procedures. The module also includes information about the histological analysis of the samples used for gene expression analysis. The integration of this information into the microarray information system facilitates the processes of quality control and assessment of microarray experiments as well as further clinical interpretation of the results.

Keywords:

colonic neoplasm, microarray analysis, databases

Introduction

The recent advances in molecular biology and its techniques are being translated into medical applications and a better understanding of physio-pathological mechanisms underlying the disease states. One of those relevant techniques is microarray technologies, these techniques and their application for the study of pathological processes is becoming more and more frequent. Application of microarrays in medicine is mainly focused on the study of gene expression levels, and more specially on the correlation of those expression levels with relevant pathological states. Most of these works done with microarrays have been focused on the analysis of gene expression during cancer onset and development. Some examples of the aims pursued with these studies are the identification of relevant gene signatures for the prognosis of disease(1;2), differentiation and reclassification of disease(3) and others(1;4;5). This kind of experiments generate huge amounts of data that must be analyzed and correlated with relevant clinical data in order to facilitate their interpretation and extract the most of them. Therefore it's necessary to develop systems that can manage simultaneously both experimental data and clinical data for this purpose.

Methods

The design of the new module for the clinical annotation of samples used in microarray experiments was developed on AMANDA microarray database. AMANDA is a MIAME(6) compliant microarray database previously presented in Medinfo 2004.

The extension module has been developed in MS Access due to the familiarity of the user with this environment, continuing the philosophy of AMANDA database. This new module is composed of tables and forms in which the clinicians fill in all the selected fields.

The module was developed in collaboration with clinical oncologists who gave feed back of what information is relevant for their studies and will be required to correlate gene expression with the clinical outcomes of the disease. In the development of the extension module controlled vocabularies and terminologies as well as codes were used with the aim of standardizing the database as well as minimizing the mistyping errors during database filling processes.

The new module includes tables and forms related with:

- Patient data (AJCC defined stage group, ECOG scale level, whether the patient undergo surgery, what type of surgery, the therapies used and finally also a record for patient follow-up regarding disease state).
- Tumor data (aspects regarding tumor location in colon, Dukes' scale, metastasis degree, whether there are affected nodes and their number, and also a tumor extension scale).
- Sample Data: the data collected includes the reference to those histological preparations used for sample selection including aspects such as type of histology, % tumoral cells, % of necrosis or inflammation degree in that piece of tissue.

The forms for data input as well as some reports have been designed in order to present users with some of the most relevant data available in the database. For this purpose several queries have been especially done and formatted.

Results

A new customized version of AMANDA database has been generated by the inclusion of this new module. The use of a modular design had the benefit of using an already designed system that is compliant with microarray standards. With this new module it is possible to achieve a higher level of accuracy in the clinical annotation of samples, in a microarray environment.

During database design non standardized terminology was avoided as much as possible in order to facilitate further integration of the data with other systems. Where it was possible the use of standard terms and controlled vocabularies and codes was used. This extension module is focused on colon cancer and therefore some of the terminologies uploaded in the database are focused only on this particular disease (v.g. anatomical location), while other fields can be directly applied on any other type of cancer (v.g. metastasis degree).

The system has been developed using an user friendly interface with several forms to facilitate the user the uploading of data into the system in a environment that is well known (MS Access) for both clinicians and researchers. In this interface it's possible to find forms designed for data uploading. There are also customized reports based on queries presenting the data in the most demanded fields by the physicians as well as by the molecular biology researchers.

Conclusions

The annotation and management of all the processes undergone by the samples ensures the quality control of all these processes and it's very important in order to minimize the presence of experimental artifacts that may affect the final results of these experiments.

A remarkable aspect is that the interface must be easy to use and intuitive to ensure the participation of clinicians in filling the required fields. For this reason an MS access environment was selected, clinicians are already familiarized with this environment and they can use it immediately with few if any explanations about the system.

The use of a unified system for storage of both clinical and experimental data and information related with the microarray based analysis of clinical colon cancer samples facilitates the interpretation of these complex experiments.

References

[1] 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002; 415(6871):530-536.

[2] Vasselli JR, Shih JH, Iyengar SR, Maranchie J, Riss J, Worrell R et al. Predicting survival in patients with metastatic kidney cancer by gene-expression profiling in the primary tumor. Proc Natl Acad Sci U S A 2003; 100(12):6958-6963.

[3] Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999; 286(5439):531-537.

[4] Mariadason JM, Arango D, Augenlicht LH. Customizing chemotherapy for colon cancer: the potential of gene expression profiling. Drug Resist Updat 2004; 7(3):209-218.

[5] Mischel PS, Cloughesy TF, Nelson SF. DNA-microarray analysis of brain cancer: molecular classification for therapy. Nat Rev Neurosci 2004; 5(10):782-792.

[6] Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet 2001; 29(4):365-371.