

Rosetta Uses MathWorks Tools to Predict the Clinical Outcome of Breast Cancer Patients

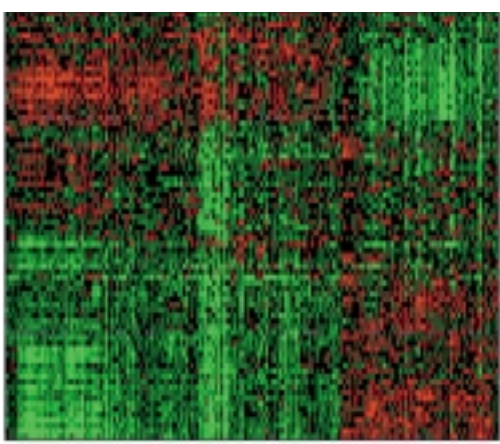
Rosetta Inpharmatics, a wholly owned subsidiary of Merck & Company, recently collaborated with the Netherlands Cancer Institute (NKI) to develop a tool that enables clinicians to determine a breast cancer patient’s prognosis based on the gene expression profile of the primary tumor. This project is an example of how programmers, when equipped with the right software, can respond promptly to researchers’ requests for custom analysis tools.

“MathWorks tools are integral to this type of custom analysis work,” says Dr. Hongyue Dai, Director of Custom Analysis and MATLAB Tools at Rosetta Inpharmatics/Merck Research Laboratories. “MATLAB frees us to focus on data analysis instead of programming. It greatly speeds up our coding process because we don’t have to write lower-level routines that are already in the MATLAB library.”

capabilities with the ability to handle large datasets rapidly. The software had to be flexible enough to allow for trial and error when selecting features and constructing classifiers.

“One of the key challenges in microarray experiments is image analysis,” explains Dr. Dai. His team needed an effective means of extracting signal intensities from TIFF images of microarray slides to determine how much of a gene was present in a particular cell. Because the TIFF images are too large and complex to be processed by hand, programmers would need to preprocess the images and design a batch process to extract the relevant data.

* Early Breast Cancer Trialists’ Collaborative Group



Two-dimensional clustering of DNA microarray gene expression data from breast cancer samples. Red indicates over-expression; green indicates under-expression. Patients with the expression pattern similar to the top of the plot generally have poor outcome.

THE CHALLENGE

To accurately predict the clinical outcome for breast cancer patients

THE SOLUTION

Use MathWorks products to develop a tool that lets clinicians make a prognosis based on the gene expression profile of the patient’s primary tumor

THE RESULTS

- Accurate prediction of disease outcome
- Fast, effective response to scientists’ needs
- Flexibility to adjust algorithms whenever necessary

THE CHALLENGE

It is difficult to determine the best course of treatment for a patient with breast cancer. Patients at the same stage of the disease and receiving the same treatment can have markedly different outcomes. Chemotherapy and hormone therapy reduce the risk of distant metastases by about a third, but studies show that 70% – 80% of patients receiving this treatment would have survived without it.*

Dr. Dai and his colleagues were asked to develop a tool that would enable cancer researchers to determine which genes in breast cancer patients were strong predictors of future metastases. To do this, they would need software that coupled powerful statistics

“*MathWorks tools are integral to the custom analysis work that we perform. MATLAB frees us to focus on data analysis instead of programming. It greatly speeds up our coding process.*”

Dr. Hongyue Dai,

Rosetta Inpharmatics/Merck & Company

THE SOLUTION

NKI researchers collaborating with the Rosetta team followed the progress of a group of 117 patients for more than five years. They examined the original DNA samples of patients who had had a poor outcome in order to identify genes whose expression level was associated with that result. Using that data, Dr. Dai's programmers used MATLAB to perform DNA microarray analysis, which identified genes that were strong predictors of distant metastases. "Since MATLAB and the Image Processing Toolbox are fully integrated and the MATLAB platform is very good for matrix calculation, we did not have to spend time writing the low-level image processing and the basic data analysis routines like vector and matrix calculations," notes Dr. Dai.

The programmers then developed an unsupervised, hierarchical clustering algorithm in MATLAB that enabled them to group the patients' tumors based on the dominant expression features. They then developed a classifier based on the genes that carry the prognosis information. They discovered that 70 genes correlated tightly with the patients' outcome, indicating that a prognosis could be determined based on the gene expression profile of the primary tumor.

The Informatics group also used MATLAB to prototype algorithms and code for their commercial product, the Rosetta Resolver Gene Expression Data Analysis System. Based on the same premise as the breast cancer prediction tool, Rosetta Resolver includes tools for high-powered analysis, visualization, and storage of gene expression data. Dr. Dai notes that MATLAB considerably accelerated the prototyping process on this product.

APPLICATION AREAS

- Algorithm development
- Biomedical engineering
- Image processing

PRODUCTS USED

- MATLAB®
- Image Processing Toolbox

THE RESULTS

■ **Accurate prediction of disease**

outcome. The gene profiling method enabled the scientists to accurately predict the outcome of disease. Compared to the clinical method currently in use, the microarray-based classifier can reduce unnecessary toxic chemotherapy from 90% of the breast tumor patients to nearly 40%.

■ **Fast, effective response to scientists'**

needs. "Our research scientists are happy with the quick feedback," Dr. Dai says. "Using MathWorks tools, we can respond to their requests very fast, and it's easy for the scientists to use these tools. Using the GUIs that we develop in MATLAB, they can access functions without having to remember the underlying code."

■ **Flexibility to adjust algorithms**

whenever necessary. "If we had done the work in C, we would have had to write a function, compile it, link to some library, and repeat that process every time we made even a little change," notes Dr. Dai. "In MATLAB, we can do this work much more easily using the command-line interface."

To learn more about Rosetta Inpharmatics, visit www.rosettatabio.com.

www.mathworks.com