Mining high-throughput biological data: methods, algorithms and application

Eytan Domany

Dept of Physics of Complex Systems and
Center for Systems Biology
Weizmann Insitute of Science
Rehovot, Israel

DNA chips are novel experimental tools that have revolutionized research in molecular biology and generated considerable excitement. A single chip allows simultaneous measurement of the level at which thousands of genes are expressed. A typical experiment uses a few tens of such chips, each devoted to one sample - such as material extracted from a tumor. Hence the results of such an experiment consist of a table, of several thousand rows (one for each gene) and 50 - 100 columns (one for each sample). Extracting relevant information from such a large, complex and noisy data set requires development of novel methods of analysis.

I will briefly demonstrate how we combine standard statistical analysis with novel unsupervised methods (clustering[1], bi-clustering[2] and sorting[3]) to mine expression data obtained from leukemia[4], cervical[5] and colon cancer patients.

If time permits, I will describe a novel experimental tool - antigen chip – and the manner in which it can be used to predict whether an individual (mouse) will or will not become diabetic[6].

1. Blatt et al, *Physical Review Letters* **76**, 3251 (1996)
2. Getz et al, *PNAS* **97**, 12079 (2000).
3. Tsafrir et al, *Bioinformatics* **21,** 2301 (2005)
4. Rozovskaia et al, *PNAS* **100**, 7853 (2003)
5. Rosty et al, *Oncogene* **24,** 6367 (2005)
6. Quintana et al, *PNAS* **101**, 14615 (2004)