

New Insights in Clinical Impact of Molecular Genetic Data by Knowledge-driven Data Mining

Daniel Berrar*

Intelligent Bioinformatics Systems
German Cancer Research Institute
Im Neuenheimer Feld 280
69120 Heidelberg, Germany.
d.berrar@dkfz.de

Martin Granzow*

Intelligent Bioinformatics Systems
German Cancer Research Institute
Im Neuenheimer Feld 280
69120 Heidelberg, Germany.
m.granzow@dkfz.de

Werner Dubitzky

Intelligent Bioinformatics Systems
German Cancer Research Institute
Im Neuenheimer Feld 280
69120 Heidelberg, Germany.
w.dubitzky@dkfz.de

Roland Eils

Intelligent Bioinformatics Systems
German Cancer Research Institute
Im Neuenheimer Feld 280
69120 Heidelberg, Germany.
r.eils@dkfz.de

* both authors contributed equally to this paper.

ABSTRACT

Traditionally, classification of complex genetic diseases such as cancer has been performed on the basis of nonmolecular criteria such as tumor tissue type, pathological features, and clinical stage. It has been generally accepted that some patients grouped into a given category will have a certain survival prognosis and response to a particular therapy. While several studies have recently reported on the application of microarray gene expression analysis for molecular classification of cancer, attempts to integrate complex and heterogeneous molecular genetic data with clinical parameters are just at the beginning. Based on heterogeneous microarray gene expression and interphase cytogenetics data in combination with clinical patient data we have applied machine learning approaches for the classification of B-Cell chronic lymphocytic leukemia (B-CLL) patients into genetic risk groups. We believe this knowledge-driven approach is an important step forward for our capabilities to study complex functional relationships between molecular genetic and clinical data. It is bound to become a wide spread tool for future studies in clinical genomics.

ADDITIONAL AUTHORS

Stephan Stilgenbauer (Department of Internal Medicine III, University Hospital Ulm, Robert-Koch-Str. 3, 89081 Ulm, Germany. Email: stephan.stilgenbauer@medizin.uni-ulm.de). Klaus Wilgenbus (Department of Exploratory Research, Boehringer Ingelheim, Austria, Dr. Boehringer Gasse 5-7, 1121 Vienna, Austria. Email: klaus.wilgenbus@vie.boehringer-ingelheim.com). Hartmut Döhner (Department of Internal Medicine III University Hospital Ulm, Robert-Koch-Str. 3, 89081 Ulm, Germany. Email: hartmut.doehner@medizin.uni-ulm.de). Peter Lichter (Organization of Complex Genomes, German Cancer Research Institute, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. Email: p.lichter@dkfz.de).

REFERENCES

- [1] Granzow M., Berrar D., Dubitzky W., Schuster A., Azuaje F.J., Eils, R. Tumor Classification by Gene Expression Profiling: Comparison and Validation fo Five Clustering Methods. ACM SIGBIO Newsletter, vol. 21, no. 1: 16-22, April 2001.
- [2] Zwiebel J.A, Cheson B.D. Chronic lymphocytic leukemia: staging and prognostic factors. Semin. Oncol. 25, 42-59 (1998).
- [3] Julius G., Merup M. Cytogenetics in chronic lymphocyte leukemia. Semin. Oncol. 25, 19-26 (1998).
- [4] Stratowa C., Löffler G., Lichter P., Stilgenbauer S., Haberl P., Schweifer N., Döhner H., Wilgenbus, K.K. cDNA Microarray gene expression analysis of B-cell chronic lymphocytic leukemia proposes potential new prognostic markers involved in lymphocyte trafficking. J Cancer 91: 474-480, 2001.
- [5] Döhner H., Stilgenbauer S., Benner A., Leupolt E., Krober A., Bullinger L., Döhner K., Bentz M., Lichter P. Genomic aberrations and survival in chronic lymphocytic leukemia. N Engl J Med 2000 Dec 28;343(26):1910-6.
- [6] SPSS Clementine. <http://www.spss.com/clementine>.
- [7] Witten I.H., Frank E. Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann Pub., San Francisco, 1999.
- [8] Berrar D., Dubitzky W., Solinas-Toldo S., Bulashevska S., Granzow M., Conrad C, Kalla K., Lichter P., Eils R. A Database for Comparative Genomic Hybridization Analysis. IEEE EMB, July 2001, to appear.