E4. Gene expression profiles: What the clinicians need to know

Juliane Hannemann, Marc J. van de Vijver

Dept of Pathology, Netherlands Cancer Institute, Amsterdam, The Netherlands

Introduction

Prognostic and predictive factors play an important role in the treatment of breast cancer. While many statistically significant associations have been identified, unfortunately not many of these associations are sufficiently strong to be also clinically useful [1].

Genome-wide monitoring of gene expression using DNA microarrays makes it possible to study thousands of genes in a tumour sample in a single experiment. By looking for an association between the gene expression pattern and tumour behaviour, it should be possible to identify new prognostic and predictive factors.

Microarray-based expression profiling has shown promise with the preliminary demonstration that clustering techniques can predict clinical outcome in various malignancies, including malignant lymphoma [2,3], leukaemia [4], prostate cancer [5,6], malignant melanoma [7], lung cancer [8,9] and breast cancer [10]. Data in breast cancer have demonstrated the ability of microarray-based expression profiling to predict disease-free survival and overall survival from profiles in breast cancer surgical specimens [11–13].

What is a DNA microarray?

A microarray is an orderly arrangement of known or unknown DNA samples attached to a solid support. One array may contain many thousands of spots and can be obtained by a number of different methods. The probes attached to the solid support can be cDNAs, oligonucleotides of varying length or genomic sequences [14,15]. The target sequence hybridised to probes on the array may be radioactively or fluorescently labelled.

Technically, it is possible to analyse all estimated 30,000–40,000 genes in the human genome for their expression. It is also to be expected that the function of the proteins encoded by many of these genes will be elucidated in the coming years. A major advantage of microarray analysis is that specific properties of cells can be recognised by the expression level of a large set of genes. This has been operationally defined as "expression signatures" [16].

Data analysis: unsupervised and supervised classification

The main methods to identify categories of tumours based on gene expression profiles are unsupervised and supervised classifications. With supervised methods, clinical or pathological information is used to find correlations with gene expression patterns. With unsupervised methods, the tumours are grouped on the basis of gene expression pattern; the main unsupervised method used is two-dimensional hierarchical cluster analysis.

A mathematical algorithm termed hierarchical clustering is used to order the gene expression data. Both the tumours and the genes are ordered according to similarities in regulation of the genes. Thus, gene expression analysis can be used to subclassify tumours on the basis of hierarchical cluster analysis in specific subgroups. Several "signatures" can be analysed that reveal specific properties of the tumour cells; and the contribution of non-tumour cells to the tumour mass. It is to be expected that the hierarchical cluster analysis will become more sophisticated; and that an increasing number of "signatures" can be recognised, making microarray data on tumour samples in relation to clinical behaviour an increasingly powerful approach.

To find gene expression patterns that can predict the clinical behaviour of tumours, it is more appropriate to use a supervised method that makes distinctions among the specimens on the basis of predefined clinical and pathological information. Several methods for supervised classification have been developed; for a brief overview of such techniques: see Ref. [17]. Fundamental to each of these techniques is that a combination of genes is identified, whose expression can predict tumour behaviour (for example: risk of developing distant metastases; responsiveness to specific forms of chemotherapy). After the identification of a predictive gene expression pattern by supervised classification, validation of the signature in a sufficiently large patient series is essential.

Correlation studies of gene expression in clinical breast cancer samples

For clinical use of gene expression profiling, the main aim is to identify gene expression profiles associated with specific clinical endpoints. When such gene expression

1359-6349/$ – see front matter © 2004 Elsevier Ltd. All rights reserved.

profiles are identified, the next step will be to implement gene expression profiling in the diagnostic process for breast cancer patients.

The first large series of clinical breast cancer samples was reported by Perou and colleagues who obtained microarray portraits of a set of 65 surgical specimens of human breast tumours from 42 individuals [18]. A difference in gene expression pattern between oestrogen receptor (ER)-positive and ER-negative tumours has been found in several studies [19–21].

To date, only few studies on larger series of patients have been performed, limiting the use of gene expression profiling to the research setting. As a result of the large number of datapoints for each tumour, results of small studies need to be interpreted with caution, because statistical flaws can easily result in the identification of gene expression patterns that cannot be reproduced in independent patient series. With these caveats in mind, in the coming years new diagnostic tests for the prediction of breast cancer behaviour will undoubtedly emerge. The treatment of breast cancer is becoming complex, requiring a growing number of choices and decisions. In the future, these choices and decisions in individual patients can be tailored based on clinical, pathological and gene expression parameters.

References

[1] Isaacs C, Stearns V, Hayes DF. New prognostic factors for breast cancer recurrence. Semin Oncol 2001; 28(1): 53–67.

[2] Alizadeh AA, Ross DT, Perou CM, van de Rijn M. Towards a novel classification of human malignancies based on gene expression patterns. J Pathol 2001; 195(1): 41–52.

[3] Rosenwald A, Wright G, Chan WC et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. N Engl J Med 2002; 346(25): 1937–1947.

[4] Staudt LM. Molecular diagnosis of the hematologic cancers. N Engl J Med 2003; 348(18): 1777–1785.

[5] Dhanasekaran SM, Barrette TR, Ghosh D et al. Delineation of prognostic biomarkers in prostate cancer. Nature 2001; 412(6849): 822–826.

[6] Varambally S, Dhanasekaran SM, Zhou M et al. The polycomb group protein EZH2 is involved in progression of prostate cancer. Nature 2002; 419(6907): 624–629.

[7] Bittner M, Meltzer P, Chen Y et al. Molecular classification of cutaneous malignant melanoma by gene expression profiling. Nature 2000; 406(6795): 536–540.

[8] Garber ME, Troyanskaya OG, Schluens K et al. Diversity of gene expression in adenocarcinoma of the lung. Proc Natl Acad Sci USA 2001; 98(24): 13784–13789.

[9] Beer DG, Kardia SL, Huang CC et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat Med 2002; 8(8): 816–824.

[10] Chang JC, Wooten EC, Tsimelzon A et al. Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. Lancet 2003; 362(9381): 362–369.

[11] Sorlie T, Perou CM, Tibshirani R et al. Gene expression patterns of breast carcinomas distinguish tumour subclasses with clinical implications. Proc Natl Acad Sci USA 2001; 98(19): 10869–10874.

[12] van de Vijver MJ, He YD, van't Veer LJ et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002; 347(25): 1999–2009.

[13] Huang E, Cheng SH, Dressman H et al. Gene expression predictors of breast cancer outcomes. Lancet 2003; 361(9369): 1590–1596.

[14] Hughes TR, Mao M, Jones AR et al. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nat Biotechnol 2001; 19(4): 342–347.

[15] Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ. High density synthetic oligonucleotide arrays. Nat Genet 1999; 21(1 Suppl): 20–24.

[16] Alizadeh AA, Eisen MB, Davis RE et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000; 403(6769): 503–511.

[17] Knudsen S. A Biologist's Guide to Analysis of DNA Microarray Data. 1st ed. John Wiley & Son, Inc., 2002.

[18] Perou CM, Sorlie T, Eisen MB et al. Molecular portraits of human breast tumours. Nature 2000; 406(6797): 747–752.

[19] 't Veer LJ, Dai H, van de Vijver MJ et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002; 415(6871): 530–536.

[20] Gruvberger S, Ringner M, Chen Y et al. Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. Cancer Res 2001; 61(16): 5979–5984.

[21] West M, Blanchette C, Dressman H et al. Predicting the clinical status of human breast cancer by using gene expression profiles. Proc Natl Acad Sci USA 2001; 98(20): 11462–11467.