COMMENTARY

DNA microarrays in breast cancer: the promise of personalised medicine

See page 1590

Breast cancer is clinically heterogeneous, with varying natural history and response to treatment. Despite much effort to identify clinical measures of risk, methods to accurately predict an individual's clinical course are lacking. Whilst lymph-node status at diagnosis is the most important measure for future recurrence and overall survival, it is a surrogate that is imperfect at best. About a third of patients with no detectable lymph-node involvement, for example, will develop recurrent disease within 10 years.[1]

The clinical heterogeneity of breast cancer is probably due to the genetic complexity of individual tumours, which have multiple somatic mutations and epigenetic changes that influence the expression of many genes that drive tumour growth, invasion, and metastasis. Different breast tumours, moreover, may arise from distinct cell-types.[2] This complexity has been difficult to study with traditional methods which are best suited to studying one gene at a time. The advent of DNA microarray technology, however, has recently enabled the quantitative measurement of complex multigene expression-patterns in human cancer.[3]

Gene-expression profiling by DNA microarrays uses nucleic acid polymers, immobilised on a solid surface, as probes for gene sequences. DNA microarrays are relatively easy to use, yield gene-expression measurements for thousands of genes simultaneously, and can be used in large numbers of samples in parallel.[4] The results can be used to accurately diagnose and molecularly classify tumours,[5–7] assess their propensity to metastasise,[8] and predict response to combination chemotherapy.[9] Thus there is keen interest in defining the gene-expression profiles of all human tumours to create a new generation of clinically useful cancer diagnostics. There is also great hope that genetic information from these studies will lead to a deeper mechanistic understanding of the molecular pathways that cause cancer.

Breast cancer has been particularly fertile ground for exploring the diagnostic usefulness of microarrays. Recent studies suggest that gene-expression patterns of primary tumours are better than available clinicopathological methods for determining the prognosis of individual patients.[6,10,11] In this issue of *The Lancet*, Erich Huang and colleagues extend these observations by using microarray-derived gene-expression profiles to classify individual breast tumours by their likelihood of having associated lymph-node metastases at diagnosis and by 3-year recurrence risk. These investigators first used unsupervised learning to cluster about 12 000 genes into groups based on similarity of gene expression across breast cancer samples. They then used singular-value decomposition to determine a "metagene" for each cluster; this metagene is not an actual gene, but rather a feature that encompasses much of the discriminatory information in a given cluster of genes. They then fed these metagenes into a decision-tree algorithm that "learns" to distinguish lymph-node negative from lymph-node positive tumours with these metagenes and then "classifies" unknown samples based on this training.

The predictive models that are described by Huang and colleagues use multiple abstracted features (metagenes) for classification. Implicit in this approach is the idea that highly accurate molecular classification for difficult clinical problems is possible only when the information from many genes is combined. Previous studies have consistently proven this point. It is not clear, however, why Huang and colleagues found it necessary to use highly abstracted features. Others have used relatively simple methods to identify individual genes, which in combination can be used for accurate classification and prognostication in breast cancer.[6,10] The use of abstracted metagenes poses two problems. First, this derivative information cannot be easily studied by conventional approaches. Second, gleaning biological or mechanistic understanding from abstracted metagenes is a formidable challenge. Further studies will be required to fully determine whether there are appreciable advantages to this complex analytic approach.

Huang's study also raises the important question of what it means to "validate" an observation based on DNA microarrays. The purest way to validate a classifier is by training it on one sample set and then testing it, without modifications, on a second independent test set. Huang and colleagues use this approach to validate prediction accuracy for lymph-node status. Without a test set, another approach is to use leave-one-out cross-validation within a single dataset. This approach involves holding out a sample from a set, training a classifier on the remaining samples, testing the classifier on the left-out sample, recording the classification result, and subsequently repeating this procedure for each sample in the set. Due to limitations in sample number, Huang and colleagues use cross-validation to determine the accuracy of their 3-year recurrence predictor. Since cross-validation generally overestimates classification accuracy, further study will be needed to determine the true accuracy of their classification method for breast cancer recurrence.

The use of microarrays to predict an imperfect surrogate measure such as lymph-node status is unlikely to markedly change patients' care, since there are presently surgical methods for examining nodal status, including sentinel-node mapping and lymph-node dissection. The ability to accurately predict long-term recurrence with microarrays, however, might prove very important if subsets of patients who will not relapse can be

spared the toxicity of adjuvant chemotherapy. Huang's study and data from van de Vijver et al[11] offer hope that this predictive ability might indeed be possible in the not too distant future.

Perhaps equally important is what Huang's study might be saying about the biology of cancer. Traditional models of tumorigenesis hold that most primary tumour cells have low metastatic potential, but rare cells within large primary tumours acquire metastatic capacity through additional somatic mutation. These models thus predict that genetic markers of metastatic behaviour should not be detectable with microarrays in primary tumours, since microarrays can only detect expression patterns arising from significant portions of a tumour. Microarray studies, however, have recently shown that a cancer's natural history—including metastasis and response to treatment—is indeed encoded in a large proportion of primary tumour cells.[8,9,12,13] Interestingly, the primary tumour genes that predict the presence of lymph-node metastases at surgery in this study are largely different from the genes that predict the development of distant metastasis (ie, recurrence), suggesting that local and distant metastases may be governed by distinct molecular programmes. Huang and colleagues thus provide further evidence for the emerging view that primary tumour metastasis may be determined by initial mechanisms of transformation rather than metastasis-enabling mutations themselves.[8,14]

*Sridhar Ramaswamy, Charles M Perou

Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA 02115, USA (SR); and Whitehead/MIT Center for Genome Research, Cambridge, MA (SR); and Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA (CMP) (e-mail: sridhar@genome.wi.mit.edu)

1 Cole BF, Gelber RD, Gelber S, Coates AS, Goldhirsch A. Polychemotherapy for early breast cancer: an overview of the randomised clinical trials with quality-adjusted survival analysis. *Lancet* 2001; **358:** 277–86.
2 Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature* 2000; **406:** 747–52.
3 Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999; **286:** 531–37.
4 Ramaswamy S, Golub TR. DNA microarrays in clinical oncology. *J Clin Oncol* 2002; **20:** 1932–41.
5 Ramaswamy S, Tamayo P, Rifkin R, et al. Multiclass cancer diagnosis using tumor gene expression signatures. *Proc Natl Acad Sci USA* 2001; **98:** 15149–54.
6 Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 2001; **98:** 10869–74.
7 Yeoh EJ, Ross ME, Shurtleff SA, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* 2002; **1:** 133–43.
8 Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. *Nat Genet* 2003; **33:** 49–54.
9 Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000; **403:** 503–11.
10 van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002; **415:** 530–36.
11 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:** 1999–2009.
12 Pomeroy SL, Tamayo P, Gaasenbeek M, et al. Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature* 2002; **415:** 436–42.
13 Ye QH, Qin LX, Forgues M, et al. Predicting hepatitis B virus-positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning. *Nat Med* 2003; **9:** 416–23.
14 Bernards R, Weinberg RA. Metastasis genes: a progression puzzle. *Nature* 2002; **418:** 823.

Enfuvirtide, a new drug for HIV infection

HIV fusion and entry into target cells occur through a series of interactions between viral envelope glycoproteins and host receptors. The envelope gp41 subunit undergoes a conformational change that facilitates fusion of viral and cellular membranes; inhibition of this change should prevent virus-cell fusion and infection of target cells.[1] The fusion process represents a pharmacological target that is unique among presently available antiretroviral agents, which act after HIV has infected the target cell. Enfuvirtide (T-20, Fuzeon) is a synthetic 36-aminoacid peptidomimetic that binds to a region of gp41 and prevents the conformational change necessary for fusion of HIV to the CD4+ cell.[2] This agent is the first of a class of binding-fusion-entry inhibitors to receive regulatory approval; its development demonstrates that fusion can be selectively inhibited, which results in significant reductions in plasma HIV RNA.

The recommended adult dose of enfuvirtide is 90 mg subcutaneously, twice daily. The elimination half-life averages 3·8 hours, which supports this dosing interval.[3] Maximum and minimum plasma concentrations of enfuvirtide average 5 (SD 1·7) μg/mL and 3·3 (1·6) μg/mL, respectively.[4] In phase 1/2 studies, plasma HIV RNA concentrations were lower when plasma concentrations of the drug were above 1 μg/mL throughout the 12-h dosing interval.[2,5] The rate of removal of enfuvirtide from plasma is positively correlated with bodyweight, and, interestingly, is 20% lower in women than in men after correction for weight. No adjustments for bodyweight or sex are recommended in the approved labelling; however, data are not provided to show that none are necessary.[4] Clinically significant interactions between enfuvirtide and hepatically metabolised agents are not expected. Indeed, co-administration with ritonavir, or saquinavir plus ritonavir, increased trough concentrations of enfuvirtide by 14% and 26%, respectively; co-administration with rifampin decreased enfuvirtide concentrations by 15%. The elimination pathways of enfuvirtide remain to be elucidated.

Enfuvirtide's safety and efficacy have been established in two pivotal studies of the same design. 501 patients in the USA, Canada, Mexico, and Brazil were enrolled into TORO 1, and 504 patients in Europe and Australia were enrolled into TORO 2.[6,7] Both trials were randomised open-label studies of a new optimised background antiretroviral regimen with or without enfuvirtide. The optimised background regimen consisted of three to five antiretroviral agents selected for viral genotype and phenotype. Eligibility criteria included: 6 months of treatment with at least one nucleoside and one non-nucleoside reverse-transcriptase inhibitor and two protease inhibitors, documented resistance to these three classes of drugs, or both; and plasma HIV RNA over 5000 copies per mL. Patients who enrolled had a mean of 7 years' previous therapy and exposure to a mean of twelve antiretrovirals.[7] The primary efficacy endpoint was the mean RNA change from baseline to week 24, which was $-1{\cdot}696$ $\log_{10}$ in the enfuvirtide recipients compared with $-0{\cdot}764$ $\log_{10}$ ($p<0{\cdot}001$) in those who received just the optimised background regimen.[7] This difference in viral load of $-0{\cdot}93$ $\log_{10}$ in TORO 1 was similar to the $-0{\cdot}78$ $\log_{10}$ difference in TORO 2.[6] Enfuvirtide was well tolerated. Reactions at the injection site were the most common adverse event and nearly all patients reported at least one reaction. Only 2·8% of enfuvirtide recipients in TORO 1 discontinued therapy because of injection-site reactions and adherence was high, which indicates that