Gene expression profiling for individualised breast cancer chemotherapy: success or not?

John Ioannidis

Results of a recent study indicate that gene expression profiling seems to improve prediction of chemotherapy effect in breast cancer, but methodological caveats remain worrisome.

A study by Paik et al. (see opposite) has shown that a well-characterised recurrence score (RS) using information on the expression of 21 genes can successfully separate women who benefit from breast cancer chemotherapy from those who do not. Their study has several strengths: the RS has been developed with careful attention to both laboratory and statistical procedures and has been standardised to become commercially available, RS has already been found to predict recurrence and survival in a validation dataset (NSABP B-14),[1] and a meticulous training phase was carried out using three independent databases to maximise generalisation. Moreover, the treatment–RS interaction is demonstrated in 651 node-negative, oestrogen-receptor-positive women, a sample size much larger than used for most previous research on gene expression profiles.[2]

The interaction term between treatment and RS identified by Paik et al. has borderline statistical significance (P=0.038). Interestingly, even though the study population is derived from a randomised trial (NSABP B-20), the distribution of RS levels differs significantly between the tamoxifen and the tamoxifen plus chemotherapy arms (P=0.036). This result illustrates that biases and chance alone may yield similar P values to those found for the interaction term; however, an interaction is not necessarily required for a predictive score to be useful in therapeutic decisions. In the low-risk group, absolute risk is so low that chemotherapy is not recommended in any case.

The question is whether RS provides treatment guidance in addition to that available using routine information (e.g. age, tumour grade and receptor levels). RS correlates with and might be more informative than these classic predictors; however, even age, while clearly seen to have an interaction with the treatment effect in the full NSABP B-20 database, did not reach nominal significance in the 651 patients analysed in this study, owing to limited power. We need very large studies to discern the exact incremental benefit of RS interactions over classic predictors. Such predictors, which are routinely available, should be included in prognostic models.

The greatest concern regarding Paik et al.'s study is that tamoxifen-treated patients from the NSABP B-20 study were used in the original development of the RS, and data from these patients were important in the selection of the 21-gene signature[1]. RS is thus expected to (and does) differentiate the risk within the tamoxifen arm, since it has been trained purposely on these data. Conversely, RS does not appropriately differentiate recurrence risk in the NSABP B-20 chemotherapy arm. This contrast of good predictive performance in the tamoxifen arm and poor performance in the chemotherapy arm is what causes the significant treat-

John Ioannidis is Chair of the Department of Hygiene and Epidemiology at the University of Ioannina School of Medicine, Ioannina, Greece, with a joint appointment at Tufts University School of Medicine, Boston, Massachusetts, USA. This article was first published in *Nature Clinical Practice Oncology* 2006 vol. 3 no. 10, and is reproduced with permission. www.nature.com/clinicalpractice, doi:10.1038/ncponc0631,

ment–RS interaction effect. Given that the tamoxifen arm was a training dataset, the correct interpretation of the data is not necessarily that RS is a superb predictor of treatment response. An alternative interpretation is that RS, while previously validated in the independent NSABP B-14 dataset,[2] now fails to be validated in the independent data of the chemotherapy arm of NSABP B-20.

As gene expression profiling moves from exploratory research into clinical practice, rigorous testing with fully independent validations should continue.[3] Useful molecular signatures need to be trained and tested on several thousands of patients.[4] The validation work to date is retrospective and thus provides only preliminary evidence. The TAILORx trial, a large prospective trial of 8,000 patients, will try to validate this 21-gene signature in the clinical setting. Similarly, the MINDACT trial will try to prospectively validate a different 70-gene prognostic signature. As we move into large-scale evidence, making sense of gene expression profiling remains a fascinating challenge.

References

1. S Paik et al. (2004) A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 351:2817–2826
2. EE Ntzani, JP Ioannidis (2003) Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. *Lancet* 362:1439–1444
3. JP Ioannidis (2005) Microarrays and molecular research: noise discovery? *Lancet* 365:454–455
4. L Ein-Dor et al. (2006) Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proc Natl Acad Sci USA* 103:5923–5928

Synopsis

S Paik, G Tang, S Shak, et al. (2006) **Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer.** *J Clin Oncol 24:3726–3734*

Background. A 21-gene recurrence score (RS) assay has been developed and validated to quantify the probability of distant recurrence in women with node-negative, oestrogen-receptor-positive breast cancer.

Objective. To determine whether the RS can also predict the magnitude of chemotherapy benefit.

Design and intervention. Tumour tissue samples were obtained from the National Surgical Adjuvant Breast and Bowel Project (NSABP) B-20 trial, which investigated the value of adding chemotherapy (methotrexate and fluorouracil with or without cyclophosphamide) to 5 years of tamoxifen therapy in 2,363 patients with node-negative, oestrogen-receptor-positive breast cancer. Patients were enrolled between 17 October 1988 and 5 March 1993. Gene expression was measured using the Oncotype DX assay (Genomic Health Inc., Redwood City, CA). Each RS was determined by measuring the expression of 16 cancer-related genes and 5 reference genes, and was calculated on a scale of 0–100. Prespecified cutoff points for low-risk, intermediate-risk and high-risk disease were RS<18, RS in the range 18–30, and RS ≥ 31, respectively.

Outcome measures. The primary endpoint was freedom from distant recurrence. Cox proportional hazards models were used to study the interaction between chemotherapy treatment and RS as a continuous variable. Analysis was also performed using the predefined RS risk categories.

Results. Gene expression results were successfully obtained for 227 patients treated with tamoxifen alone and 424 patients treated with tamoxifen plus chemotherapy. Patients did not benefit equally from chemotherapy; those with a high risk of recurrence had a greater magnitude of benefit from chemotherapy than those with an intermediate or low risk of recurrence. Adding chemotherapy to tamoxifen improved the 10-year Kaplan–Meier estimate for freedom from distant recurrence from 60% to 88% in the high-risk group. The high-risk category benefited from chemotherapy, with a large reduction in distant recurrence at 10 years (relative risk [RR] 0.26, 95% CI 0.13–0.53; decrease in absolute risk 27.6%). This benefit was less clear for patients in the intermediate-risk group (RR 0.61, 95% CI 0.24–1.59; increase in absolute risk 1.8%), but a clinically important benefit from chemotherapy could not be excluded. No reduction in distant recurrence at 10 years was demonstrated for patients in the low-risk category (RR 1.31, 95% CI 0.46–3.78; increase in absolute risk 1.1%). In a multivariate analysis, the interaction between chemotherapy treatment and RS was significant (P=0.038); however, no clear cutoff point for RS could be defined.

Conclusion. The RS can predict the magnitude of benefit from chemotherapy for patients with node-negative, oestrogen-receptor-positive breast cancer, as well as the likelihood of recurrence of breast cancer, and could be used to select patients who would respond well to chemotherapy.

Acknowledgement: The synopsis was written by Petra Roberts, Associate Editor, *Nature Clinical Practice*.