Methods for Gene Expression Profiling in Clinical Trials of Adjuvant Breast Cancer Therapy

Soomyung Paik

Abstract

Although endocrine therapy is highly effective in the treatment of endocrine receptor–positive breast cancer, chemotherapy has been shown to provide clinical benefit when added to tamoxifen. However, baseline risk after tamoxifen treatment is low, especially in patients who are axillary node negative, that significant overtreatment will result if chemotherapy is given to every patient. Robust prognostic and predictive markers need to be developed to identify those at high risk of treatment failure. Although comprehensive gene expression profiling methods do offer promise, they require fresh or frozen tumor samples. To take advantage of existing archived tissue blocks with clinical follow-up collected from finished clinical trials, such as National Surgical Adjuvant Breast and Bowel Project trials B-20 and B-14, technologies that allow interrogation of archived blocks for gene expression profiling need to be realized. Recent developments in gene expression profiling technologies are discussed with their implications in clinical management of endocrine receptor–positive breast cancer.

Whereas endocrine therapy has resulted in significant improvements in clinical outcomes of hormone receptor–positive patients, the addition of chemotherapy has been shown to provide additional benefit even in node-negative patients. However, due to excellent baseline risk after 5 years of tamoxifen in the node-negative setting, almost 85% of node-negative patients would be overtreated if chemotherapy were given to every patient. Unfortunately, there has been no predictive marker for benefit from chemotherapy available, and the current clinical practice guidelines are based on the assumption that patients in all risk groups derive a similar degree of benefit from chemotherapy. Thus, current practice guidelines from the National Comprehensive Cancer Network and the St. Gallen meeting, which use classic clinicopathologic features to assess baseline risk, assign only ~10% of node-negative, hormone receptor–positive patients into a low-risk group that does not require chemotherapy. This has resulted in significant overtreatment of this group of patients. Therefore, it has become imperative to develop more robust prognosticators and predictive markers of response to chemotherapy.

Recent advances in genomics have provided tools that allow us to interrogate expression levels of entire genes in tumor cells. Studies using such tools have shown that there is tremendous heterogeneity in the molecular composition of breast cancer such that each tumor is unique, suggesting that treating every patient with same therapy probably is not the most ideal approach. However, these gene expression tools usually require very high quality RNA as a starting material. The main hurdle in identifying predictive markers for treatment benefit in the adjuvant setting, aside from the important issue of statistical power, is the lack of fresh frozen tumor tissue from large phase III adjuvant trials. Therefore, methods that allow gene expression profiling of formalin-fixed, paraffin-embedded tumor specimens, which are often fairly old, are in need. Several promising methods for gene expression profiling have been developed recently. Some of these methods will be reviewed briefly, with a focus on their potential in clinical applications.

The initial step in any gene expression profiling method is synthesis of cDNA from mRNA species extracted from the tumor tissue (Fig. 1). In the usual method of gene expression profiling, reverse transcription of the mRNA using an oligo-dT primer, designed to bind the polyadenylic acid tail of the molecule, is used to generate cDNA. In formalin-fixed tissue, the RNA has been modified by the addition of mono-methylol groups, especially to adenine, so that the oligo-dT primer cannot bind with high efficiency (Fig. 2; ref. 1). Furthermore, for unknown reasons, the RNA becomes heavily fragmented over time during storage of the blocks (2). These two problems significantly limit cDNA synthesis using RNA extracted from paraffin blocks. Several methods have been devised to overcome these limitations.

Gene Expression Profiling with Microarrays

Whereas chemical modification of RNA by formalin restricts the binding of oligo-dT primers to the polyadenylic acid tail and impedes the efficiency of reverse transcription, heating in high-pH Tris buffer can partially reverse the modification and...
allow the reverse transcription to proceed. Therefore, for relatively fresh paraffin blocks with high molecular weight RNA preserved in the specimen, the usual method of cDNA synthesis can be applied. This seems to be the basis of Paradise System from Arcturus (Mountain View, CA), which is also said to have an RNA extraction protocol that allows optimized microarray performance when used together with the Affymetrix GeneChip X3P (Santa Clara, CA) or the Agilent array (Palo Alto, CA). However, due to heavy fragmentation of RNA extractable from paraffin blocks that are more than a few years old, this method does not perform very well with such materials. Therefore, to take advantage of Paradise System, it is recommended that the RNA be extracted from the blocks as soon as they are collected and stored as RNA or cDNA instead.

TransPlex whole transcript amplification kit (Rubicon Genomics, Inc., Ann Arbor, MI) bypasses the need for an intact polyadenylic acid tail by using random primers for cDNA synthesis. Adaptor-based PCR is used to amplify the cDNA. We have tested the performance of TransPlex kit by comparing gene expression in 14 cases of breast cancer with known estrogen receptor (ER) status for which the blocks were 10 years old. In microarray studies, percent present call is a term used to describe how many probes give out signal above background. In a typical experiment with fresh tumor tissue, one would expect a percent present call of ~50%. Leave-one-out cross-validation is another validation method in which one case from the cohort is set aside and the rest of the cases are used to build a predictive model—in this case for ER status—and to see if the model predicts the status of the hold-out case correctly; this process is then repeated for all cases in the cohort. Percent present call in Affymetrix GeneChip X3P of 25% to 35% was achieved. The top most differentially expressed gene between ER+ and ER− groups detected by Affymetrix GeneChip was the ER gene. In the leave-one-out cross-validation, ER status (defined by immunohistochemistry) of all but one case was correctly classified. These data suggest that the TransPlex kit can be used for gene expression profiling of older paraffin blocks with high confidence. Because microarray assay allows interrogation of essentially all human genes at once, this method can be used for initial candidate gene selection.

### Real-time Reverse Transcription-PCR

Unlike microarrays, which allow interrogation of all expressed genes but at the expense of low dynamic range, the real-time reverse transcription-PCR (RT-PCR) assay allows interrogation of the expression level of one gene at a time but with great accuracy and a wide dynamic range. Fragmented RNA extracted from paraffin blocks can be a reasonable substrate for real-time RT-PCR assay (3). However, gene-specific priming is required for cDNA synthesis for each gene target because cDNA synthesis for the entire species of mRNA using oligo-dT primed reverse transcription is not possible for the fragmented and chemically modified RNA. This means that the assay for each gene has to be done in a separate reaction tube from the point of cDNA synthesis. Highly accurate pipetting using robotics is therefore required for reproducible assay results. For this reason, when more than a handful of genes are to be assayed, fairly sophisticated laboratory facilities are required. In addition, Cronin et al. (2) have shown that absolute signal decreases significantly if blocks have been stored for long time, resulting in ~100-fold reduction in signal if the block is 10 years old compared with freshly made block. However, careful normalization based on genes with minimal variation of expression level among different tumor samples can largely compensate for these differences in absolute signal (2).

In collaboration with Genomic Health, Inc. (Redwood City, CA), we have developed a multigene prognostic index for node-negative, ER-positive breast cancer treated with tamoxifen called OncotypeDx, based on 21 genes assayed by this method (3). To develop this assay, candidate genes (n = 250) were selected from literature and microarray data for breast cancer. These genes were tested in three independent cohort studies, including cases from the National Surgical Adjuvant Breast and Bowel Project (NSABP) trial B-20 (4). Univariate analysis showed that >40 genes correlated with clinical outcome in B-20 cohort. By selecting reproducible prognostic genes among three
independent cohorts with robust PCR performances, a multivariate prognostic model called *recurrence score* was developed that included 16 cancer-related genes and 5 reference genes. Whereas the majority of genes, composed of 16 genes, are ER (ER, PGR, BCL2, and SCUBE2) and proliferation related (Ki67, STK15, Survivin, CCNB1, and MYBL2), there are other genes (HER2, GRB7, MMP11, CTSL2, GSTM1, CD68, and BACG1). The unscaled recurrence score (RSu) was calculated with the use of coefficient that is defined on the basis of regression analysis of gene expression. Recurrence score (RS) was rescaled from the unscaled recurrence score as follows: RS = 0 if Rsu < 0; RS = 20 × (Rsu-6.7) if 0 ≤ Rsu ≤ 100; and RS = 100 if Rsu > 100. Final validation of the recurrence score was achieved by examination of its performance in a completely independent cohort from NSABP trial B-14, which was not used in the model building process (4). In the validation study, the assay was found to provide better and more reproducible indication of prognosis for ER-positive tumors in node-negative patients than age, tumor size, or histologic grade (3). Compared with National Comprehensive Cancer Network or St. Gallen criteria, recurrence score was able to categorize more patients into a low-risk group that had similar 10-year distant disease-free survival rates as low-risk groups identified by these criteria (Table 1).

One of the most interesting aspects of the study is the demonstration of a direct relationship between the level of ER mRNA expression and the degree of benefit from tamoxifen in NSABP trial B-14 comparing tamoxifen versus placebo. This is a logical finding because ER is the target for tamoxifen but one that has been very difficult to show using other means of ER expression measurement, including ligand binding assay. These data raise the possibility that for targeted therapies, accurate quantitative measures of the biological target will provide accurate predictors of response, an approach that we strongly believe needs to be tested for trastuzumab and bevacizumab. Because the recurrence score was developed for tamoxifen-treated patients, its performance in patients treated with aromatase inhibitors is unknown. Studies are being planned to analyze materials from Arimidex or Tamoxifen Alone or in Combination and MA.17 trials. The U.S. Breast Cancer Inter-group is prepared to launch a very large trial (Trial Assigning Individualized Options for Treatment) to stratify patients into different risk groups and randomize the intermediate risk group to hormonal therapy (including aromatase inhibitor) or hormonal therapy with chemotherapy. Whereas this trial will take a very long time to finish, it is hoped that the tumor blocks procured from this trial can be used to further optimize recurrence score or to help develop other more robust assays.

---

**Table 1. Comparison of OncotypeDx recurrence score with other clinical guidelines (data from tamoxifen arm of NSABP B-14)**

<table>
<thead>
<tr>
<th>Risk category</th>
<th>NCCN % of patients</th>
<th>DRFS10</th>
<th>St. Gallen % of patients</th>
<th>DRFS10</th>
<th>OncotypeDx % of patients</th>
<th>DRFS10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>7.9</td>
<td>0.93</td>
<td>7.9</td>
<td>0.95</td>
<td>50.6</td>
<td>0.93</td>
</tr>
<tr>
<td>Intermediate</td>
<td>NA</td>
<td>NA</td>
<td>33.2</td>
<td>0.91</td>
<td>22.3</td>
<td>0.86</td>
</tr>
<tr>
<td>High</td>
<td>92.1</td>
<td>0.85</td>
<td>58.8</td>
<td>0.81</td>
<td>27.1</td>
<td>0.69</td>
</tr>
</tbody>
</table>

Abbreviations: NCCN, National Comprehensive Cancer Network; DRFS10, 10-year distant recurrence-free survival; NA, not applicable.
cDNA-Mediated Annealing, Selection, Extension, and Ligation

One promising method is cDNA-mediated Annealing, Selection, Extension, and Ligation (DASL) developed by Illumina (San Diego, CA; ref. 5). Based on its bead array platform, the DASL assay can monitor expression of up to 1,536 sequence targets (512 genes at 3 probes per gene) in 500 ng of total RNAs derived from formalin-fixed, paraffin-embedded tissue samples that have been stored for up to 12 years, according to the description of the company. The DASL assay monitors gene expression by targeting sequences in cDNAs with sets of query oligos composed of multiple parts. In addition to gene-specific sequences, the query oligos contain primer landing sites for PCR amplification and an address sequence for hybridization to the universal bead array. Because randomers are used in the cDNA synthesis and because the query oligos target cDNA sequences spanning only ~50 bases, partially degraded RNAs can be used in the assay. In its design, the DASL assay resembles RT-PCR with highly multiplexed templates but with only three PCR primers. Because the oligos all share the same primers, and the amplicons are of a uniform size, the amplification step is expected to maintain an unbiased representation of transcript abundance.

We have assessed the performance of DASL using 24 cases of 10-year-old archived paraffin blocks with known ER status. ER, progesterone receptor, and insulin-like growth factor receptor were found to be significantly differentially expressed between ER- and ER+ tumors. On leave-one-out cross-validation using DASL results, 85% prediction accuracy was achieved for predicting ER status determined by immunohistochemistry. Given its low cost and the high capacity for multiplexing of the assay, DASL seems to be a very promising method that needs to be further evaluated.

Conclusion

Several options are available for gene expression profiling using RNA extracted from old formalin-fixed, paraffin-embedded tumor blocks. Whereas GeneChip analysis using TransPlex whole transcriptome amplification and DASL provide relatively inexpensive capability to profile >500 genes at once using a small amount of starting material (<200 ng RNA, usually requiring one to two 5-μm sections) and great discovery tools, real-time RT-PCR method is the preferred clinical assay platform based on excellent dynamic range of measurement and reproducibility.

Open Discussion

Dr. Myles Brown: Have you tried the Genomic Health gene set using DASL probes and tested whether that works as well as your assay?

Dr. Paik: One of the strengths of the Genomic Health assay is the dynamic range that it can provide, which is special to that active PCR because actually it is hybridized to the RNA. That is where it gets the signal. I don’t think it is going to evolve as a clinical assay, but it is going to be a great initial screening tool. We think that the eventual commercialization of the assay will have to be RT-PCR, just to be reliable.

Dr. James Ingle: The relationship of HER2 and PR to outcome is a little disappointing in your data. Do you want to comment on that? From the molecular markers we have available, the only one that seems to be ready for prime time is ER. From the clinical data, you would expect some signal from the HER2 and the PR, so to see nothing—is anybody else surprised?

Dr. Paik: The main reason that HER2 and GRB7 are in the OncotypeDx assay panel is because in the model-building set using the tamoxifen treated arm from the B-20 trial, HER2 and GRB7 were among the top contenders. They were very significant prognosticators. In UB 410, it was a complete failure. It could be just a selection bias.

Dr. Mitch Dowsett: The recurrence score is much more prognostic based than it is tamoxifen response based. Had the NSABP and Genomic Health got together and said, what we want to do is to find a predictor of benefit from tamoxifen, certainly the resulting gene panel would come out differently weighted and with possibly many different genes. I think the proliferation genes are really dominant here in determining the prognostic aspect.

Dr. Paik: Yes, it is entirely possible that if you went that route, looking for tamoxifen-response genes from the beginning, you might find genes that did not have prognostic significance in the tamoxifen arm because that population all had a response to tamoxifen.

Dr. Stephen Johnston: Because it is prognostic rather than predictive, what is the bottom line here on how this will be used to help make clinical decisions? I had a patient from the States who was put on tamoxifen and had an OncotypeDx done. It was obviously being used to decide whether or not she was going to go onto chemotherapy. What is the guidance here about whether this assay is of use in making clinical decisions? What does it add over and above PR and HER2 status?

Dr. Paik: We didn’t develop this as a predictive test for tamoxifen or endocrine therapy; we developed it as a prognostic test for tamoxifen-treated patients so that we can assess the baseline risk, which might aid in decision-making for chemotherapy. Luckily, it turned out to interact with the chemotherapy benefit, with patients with higher recurrence scores deriving more benefit from chemotherapy. So, for that decision making it might be useful. But for tamoxifen benefit it is simply an exploratory analysis. Because of the large confidence interval and low event rate, I don’t think we can draw a line to say which patients should not get tamoxifen.

Dr. Ingle: Could we have a point-and-counterpoint discussion about the two different studies, the NASBP B-14 and the M.D. Anderson study [Cin Cancer Res 2005;11:3315–9], which did not corroborate the value of the 21-gene panel?

Dr. Paik: To me, there was no real contradiction between the two studies. If you look at the M.D. Anderson study, the assay performed as expected. ER correlated with PR, ER correlated with the IGFR, and HER2 correlated perfectly with GRB7. So it is not that the assay did not work; rather, in that clinical cohort the recurrence score did not predict recurrence. The same was true for tumor grading in that cohort, so it might be a patient selection issue.

Dr. Aman Buzdar: Yes, it was a small study in node-negative, receptor-positive patients who did not receive tamoxifen. That is one of the differences between that
subgroup and the NSABP patients. The question is whether that is a self-selected subpopulation—because these were patients coming to M.D. Anderson for treatment who did not receive any systemic therapy—or is the recurrence score only predictive in the presence of tamoxifen therapy? That question cannot be answered clearly without running another cohort of patients with similar characteristics, patients who are ER positive and also node negative but did not receive any systemic therapy.

**Dr. Dowsett:** That is an extreme example of studies using these untreated populations as a control group, and yet they are untreated for special reasons. This sort of investigation needs to be done in the context of a randomized trial, or if you don't do it in the context of a randomized trial, you need to be very careful about the conclusions that you take from the data.

**Dr. Paik:** One has to realize that the Oncotype DX is not a perfect assay; it is definitely influenced by other factors. The ROC curve, the actual sensitivity and specificity of the assay, is not over 85%. So this is a method in evolution. I still regard it as a feasibility demonstration and nothing more.

**Dr. Eric Winer:** I was going to address Dr. Johnston's question a bit more by saying that on this side of the Atlantic, I don't know many academic breast cancer doctors who have embraced this assay wholeheartedly. On the other hand, I know many who have ordered it six to ten times over the past 6 months, and I would put myself in that group. The data from the prognostic standpoint are pretty solid. For a woman taking tamoxifen, this gives you information about her risk of distant recurrence at 5 and 10 years. The data are less solid in predicting the benefit of chemotherapy and in using it as a predictor of tamoxifen benefit.

**Dr. Johnston:** In terms of deciding who will get chemotherapy, other simple factors like age, tumor size, vascular invasion, quantitative level of ER, and so on, are already there for helping make that decision.

**Dr. Winer:** I agree with you. So how much better is this than a really good pathologist sitting next to you and giving you highly accurate tumor grading, quantitative ER, and good HER2? The problem is that level of pathology consult is not always available in the community. So what is potentially very helpful about this assay is the standardization.

**Dr. Ingle:** It ought to be added that we need to study this prospectively; such a study has been in the works for 2 years and will hopefully be starting up.

### References


Methods for Gene Expression Profiling in Clinical Trials of Adjuvant Breast Cancer Therapy

Soonmyung Paik


Updated version
Access the most recent version of this article at:
http://clincancerres.aacrjournals.org/content/12/3/1019s

Cited articles
This article cites 5 articles, 1 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/12/3/1019s.full.html#ref-list-1

Citing articles
This article has been cited by 2 HighWire-hosted articles. Access the articles at:
/content/12/3/1019s.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.