

NCI 2007 PRE-OP THERAPY IN BREAST CANCER  
35\_SESSION 6\_4 - PUSZTAI

DR. KATHY PRITCHARD: Thank you very much, Marco. Thank you for that talk. And next I'd like to introduce Dr. Lajos Pusztai from the MD Anderson hospital. He's going to speak to us about molecular profiling.

DR. LAJOS PUSZTAI: Thank you. I would like to thank the organizing committee for inviting me to this meeting.

I would like to discuss the value of gene expression profiling that it may contribute to neoadjuvant studies. [Brings up the slides] So, I will briefly review the potential uses of gene expression profiling in the neoadjuvant treatment setting and the limitations of the technology. I was also asked to comment on the relative merits of single baseline biopsy-based discovery strategies compared to serial biopsies during treatment. I'll also address some issues around sample size calculations in genomic studies.

So, preoperative therapy offers a unique opportunity to correlate baseline molecular markers with subsequent response to treatment. The most common justification for gene expression profiling in neoadjuvant studies is to test the hypothesis that new, mRNA-expression-based predictors of response could be discovered. The predictor could be a single gene or a combination of many genes. And one could also test a-priori-defined predictors of response.

A second important, but under-appreciated, reason to perform molecular profiling of cancers included in clinical trials is that this will create a molecular database. This database represents a comprehensive inventory of the molecular machinery of cancer at the mRNA level. It is important to realize the scientific value of creating such databases.

**PRE-OP THERAPY IN BREAST CANCER**  
**35\_SESSION 6\_4 - PUSZTAI**

So, a molecular database can have at least the same, if not greater, research value than a clinical database that we routinely create for all studies. This, of course, assumes that we accept gene expression profiling with DNA microarrays as a reliable measure of mRNA expression. I do not have time to present data to support this assumption; however, this data exists.

A comprehensive quality assurance... quality assessment of this technology has been performed, including extensive comparisons with RT-PCR measurements. And the results were published in a series of articles in *Nature Biotechnology* late last year -- and this is on the slide, for your reference.

So, I would like to illustrate the value of publicly available gene expression databases through two examples. The biomarker literature is full of markers that are associated with better or worse prognosis in various clinical trials. However, in most clinical trials, patients receive, quite appropriately, multiple different therapies, including surgery, chemotherapy, and endocrine therapy.

The critical, and often unanswered, question regarding the proposed marker is whether the marker predicts prognosis or predicts sensitivity to a particular therapy. With the advent of prospectively generated, public gene expression databases, these questions can now be addressed for mRNA-based measurements of markers.

This slide contains references to three such databases. The first includes 286 node-negative breast cancers that were treated with surgery alone. The second includes 267 ER-positive patients treated with five years of tamoxifen. The third one includes 133 patients who received neoadjuvant T/FAC -- Taxol, or paclitaxel,/FAC chemotherapy.

**PRE-OP THERAPY IN BREAST CANCER**  
**35\_SESSION 6\_4 - PUSZTAI**

All used the same gene expression profiling platform and the relevant clinical outcome data is available for each case.

You can test any of your favorite gene, gene signature, or molecular pathway in these datasets, to examine separately its prognostic, or chemo- or endocrine-response predictive values. We tested the predictive value of microtubule-binding protein Tau [MTP-Tau], our favorite gene, in

ER-positive breast cancers. The results are shown in this slide.

In the prognostic dataset, which is the first panel, there was no difference in Tau expression between those who relapsed and those who did not at five years. Therefore, Tau has no significant prognostic value. In the tamoxifen-treated dataset, those who relapsed had significantly lower Tau expression compared to those who did not. This suggests that patients with low Tau expression do not benefit from tamoxifen as much as those with higher Tau expression.

Kaplan-Meier survival curves are also shown by tertiles of Tau expression on the lower half of the slide. In the neoadjuvant paclitaxel-FAC dataset, which is the third panel, we examined if Tau expression correlated with pathological complete response to this chemotherapy, and we observed that patients with pathological complete response indeed had a significantly lower Tau expression compared to those with residual disease.

In summary, through the combined analysis of these three publicly available datasets we could establish that this particular single marker does not predict prognosis in the absence of systemic therapy, but it may identify a subset of ER-positive patients who are at risk of recurrence with tamoxifen alone and, at the same time, are also highly sensitive to chemotherapy.

**PRE-OP THERAPY IN BREAST CANCER  
35\_SESSION 6\_4 - PUSZTAI**

In the second example, investigators examined the value of an a-priori-defined predictive signature that were developed from cell-line models. Predictors of response to paclitaxel, 5-FU, doxorubicin, and cyclophosphamide in vitro were applied to human gene expression data from patients who underwent preoperative chemotherapy. The same neoadjuvant dataset as described on the previous slide was used.

The investigators reported that patients who were predicted to respond using the cell-line-based predictor indeed had higher rates of pathological CR. Regardless of whether this predictor will hold up in future validation studies, the important message is that this research approach is feasible and it is made feasible by publicly available gene expression data.

I would like to return to the main use of transcriptional profiling in the context of neoadjuvant studies. This is, to discover new predictors of response to therapy. I believe this is possible. Several such predictors have been published and some even tested in small, independent validation studies. We recently reviewed accomplishments in this field in a review article that is provided as a reference on the slide. However, the limitations of this technology are also becoming clearer and must be considered.

So, challenges. Some of the challenges in this field are well known. Low response rates seen in small, discovery studies pose a serious statistical challenge. Also, frequently, over 20,000 individual gene expression measurements are compared in the process of searching for differentially expressed genes that are associated with response. The risk of false discovery is, therefore, high. However, there are several statistical methods to adjust for multiple comparisons.

**PRE-OP THERAPY IN BREAST CANCER**  
**35\_SESSION 6\_4 - PUSZTAI**

What I would like to bring your attention to is another, lesser-known confounder. This is, the association between clinical pathological variables and large-scale gene expression patterns. The expression... the estrogen-receptor expression and histological grade are associated with large-scale gene expression patterns. ER-positive cancers differ from ER-negative tumors in the expression of well over 2,000 genes.

High-grade tumors differ from low-grade tumors almost at the same extent. Both ER-negative status and high grade are also associated with greater sensitivity to many different chemotherapies. Therefore, unadjusted comparison, pharmacogenetic comparison, of responders with non-responders will yield gene lists that are dominated by ER and grade-associated genes.

There are several statistical methods to adjust for these co-variables. Small sample size often makes these adjustments difficult. However, when such adjustments are made, the number of informative genes drops dramatically. This slide shows that when we perform unadjusted comparison to identify differentially expressed genes between those who have complete response and those who have residual disease, greater than 1,000 genes can be identified, with the predicted false discovery rate of five percent.

This means that only about five percent of these genes may not be truly differentially expressed. However, when we match cases by ER and grade, it becomes statistically difficult to identify genes, with confidence, that are differentially expressed. In this case-control study, shown on the left panel, no gene could be detected with a false discovery rate of less than five percent when ER and grade were matched.

I would like to emphasize that this does not mean that there are no truly differentially expressed genes between responders and non-responders other than the genes regulated

**PRE-OP THERAPY IN BREAST CANCER**  
**35\_SESSION 6\_4 - PUSZTAI**

by ER status and histological grade. These results only show that this particular discovery strategy can be high-risk for false-positive results.

To further illustrate this, this slide shows the top ten differentially expressed genes identified in the previous case-control study. The p-values for all these genes are very low. But the predicted false discovery rate associated with this list is 50 percent. This means that half of these genes may represent real discovery, the other half, not.

Importantly, two of the top 10 probes, including topo-isomerase II, which was over-expressed in patients who had pathologic CR to the FAC neoadjuvant chemotherapy. This is consistent with prior knowledge that implicates TOPO II as the currently best single gene predictor of increased sensitivity to anthracyclines. However, which of the remaining eight genes represent true markers of response versus spurious discovery? We will only know with further experiments.

I would like to emphasize that strong associations between ER, grade, and large-scale gene expression patterns bias pharmacogenomic discovery towards finding “general chemotherapy sensitivity” signatures. In fact, these predictors may be, to a large extent, the molecular equivalents of a combined ER-and-grade score.

Nevertheless, these signatures still could be useful. Several publications suggested that gene signatures that are predictive of prognosis or response to therapy track clinical-variable-based predictions closely, but not perfectly. Also, they appear to outperform, or at least complement, clinical-variable-based prediction models.

In our experience, a genomic predictor of pathological CR showed higher sensitivity and greater value to rule out who will not achieve this favorable response than clinical-

**PRE-OP THERAPY IN BREAST CANCER**  
**35\_SESSION 6\_4 - PUSZTAI**

predictor-based... than a clinical predictor based on a combination of ER, grade, and age. It is also very important to keep in mind that, by identifying the molecular differences between low-grade and high-grade tumors, and between ER-positive and ER-negative cancers, we increase our understanding of cancer biology. We hope that this will eventually translate not only to better treatment selection, but to better drugs for therapy.

I was specifically asked to at least touch on the complex issues of designing pharmacogenomic predictive studies. There is a vast literature on sample size calculations for genomic discovery studies. As all sample size calculations, these are based on certain assumptions about the study population, the predictor to be discovered, and the response rate.

There are two main approaches to estimate sample size. One is based on powering the study to identify individual, differentially expressed genes. The other uses existing and accumulating data to assess predictor performance as the sample size increases and estimates projected performance for a larger sample size.

Examples are shown on this slide. The upper table shows how discovery sample size increases as the mean difference decreases relative to the standard deviation, described as the standard effect size in statistical terms, for any particular gene. The estimated sample size that is needed for discovery also increases as the response rate decreases.

The lower panel shows learning curves for genomic predictors in different datasets. These curves are generated by developing predictors from increasingly bigger and bigger datasets, and extrapolating from these observed results how the predictor would work... how the predictor would work if developed from larger datasets. Please note that the

**PRE-OP THERAPY IN BREAST CANCER**  
**35\_SESSION 6\_4 - PUSZTAI**

steepness of these learning curves varies from dataset to dataset, which means that, for a different prediction problems, different discovery sample sizes may be appropriate.

Validation studies are easier to design because more is known about the performance of the predictor. However, different investigators may have different definitions of validation. For example, one could define validation as proving that marker-positive patients have definitely higher pathological CR rate than unselected or marker-negative patients. This would require showing that the lower bound of the 95-percent confidence interval of the positive predictive value of the test is greater than the upper bound of the 95-percent confidence interval of the pathological CR rate in unselected patients.

Others may argue that it may be more important to know the precision of the test with great certainty. This would require to power the study so that the standard deviation around the point estimate is acceptably low. I would like to point out that even a technically validated and accurate test may not necessarily improve clinical outcome of the tested population. Therefore, the ultimate and most important validation of any diagnostic assay is to show that it improves patient outcome.

I was also charged with discussing the potential research uses of gene expression profiling on serial biopsies collected before and during therapy. The rationale for this approach is that transcriptional changes in response to therapy will be more predictive of outcome than a baseline gene expression signature.

This makes sense. But this approach is also fraught with additional challenges. In addition to all of the caveats that apply to developing predictors from single, pre-treatment biopsies, there are a number of other problems that make successful discovery

**PRE-OP THERAPY IN BREAST CANCER**  
**35\_SESSION 6\_4 - PUSZTAI**

from serial biopsies difficult. The optimal timing of taking the follow-up biopsy is usually unknown. There is substantial time variation due to non-compliance.

There is missing data. After all, patients exercise their right that these are optional procedures. However, what I would like to draw your attention to is a challenge that relates to gene expression profiling technology itself. Important changes in the expression of a few dozen genes in response to therapy may be blurred by technical noise.

This slide shows a scatter-plot of replicate experiments. The same RNA was profiled on two separate occasions, and the results of 24,000 individual measurements are plotted against each other. A perfect correlation and a 100-percent concordance of results would be a straight line at a 45-degree angle. What you see here is pretty close to it. The concordance of these replicate measurements was 98 percent.

However, this very high reproducibility still allows for approximately 1.3 percent of all genes to have greater or equal than two-fold difference. This translates into over 300 genes. The implications of this are very important. Modest changes in the expression of a few dozen genes may not be readily discernible in the background of this technical noise in a small study.

This slide is a scatter-plot of gene expression data of the same tumor before and 48 hours after doxorubicin-docetaxel preoperative chemotherapy. You cannot avoid to notice the similarity to the previous slide. The transcriptional changes in response to chemotherapy are very small compared to the stability of the vast majority of gene expression measurements. In fact, in a small pilot study, they could not identify any single gene that showed consistent change in its expression in response to the given chemotherapy.

**PRE-OP THERAPY IN BREAST CANCER**  
**35\_SESSION 6\_4 - PUSZTAI**

I have no doubt that such genes exist, but a substantially larger sample size than most phase 2 neoadjuvant studies will be needed to identify them with any degree of statistical certainty.

So, I would like to return to the various genomic predictors that have been published or publicly presented and, I believe, are ready to be tested for their value in clinical trials. There are at least two multi-gene prognostic models that underwent several rounds of validation. Also, many groups have shown that mRNA-based ER and HER2 measurements correspond closely to immunohistochemistry and FISH results, and may even be more accurate.

There are several multi-gene predictors that could define a subset of ER-positive patients who are highly sensitive to endocrine therapy. Some of these, such as OncotypeDX, have been studied more extensively than others. Nevertheless, there are more than one endocrine response predictors that are ready to be tested in the clinic.

Similarly, there are numerous gene signatures that appear to identify individuals with substantially higher-than-average chemotherapy sensitivity. Most importantly, what you see on this slide -- that is, multiple genomic prediction reports from a single biopsy -- is technically feasible today.

All of the previously shown genomic tests are imperfect. The predictive results often come with considerably large confidence intervals, indicating uncertainty in the true predictive values. I believe an important challenge for us is to find out how one can integrate a series of imperfect tests into a clinically useful diagnostic strategy. We have done this, with some success, in the field of diagnostic imaging.

**PRE-OP THERAPY IN BREAST CANCER**  
**35\_SESSION 6\_4 - PUSZTAI**

This is our strategy to integrate our genomic predictors into a neoadjuvant clinical trial. Patients who are willing to participate in this molecular triaging program will undergo a single, needle biopsy for gene expression profiling. We will examine ER and HER2 status by mRNA-based methods as well as with routine diagnostic methods.

For the ER-positive patients, we will also estimate endocrine sensitivity by applying the 200-gene endocrine sensitivity index. We will also make a prediction regarding sensitivity to paclitaxel-FAC chemotherapy by using the 30-gene prediction of pathologic CR. Those who over-express HER2 will receive trastuzumab-containing preoperative chemotherapy. Those who are predicted to accomplish pathologic CR to neoadjuvant paclitaxel-FAC will receive that chemotherapy. And those who are predicted to be highly sensitive to endocrine therapy, but not sensitive to chemotherapy, will receive six months of preoperative endocrine treatment. Each of these treatment arms will serve as separate validation studies for the particular predictor that was used for the triaging.

An important fourth group of patients include those who are predicted not to accomplish pathologic CR and are also predicted not to be sensitive to endocrine therapies. These patients are in need for more effective new treatments. Therefore, we will offer for these individuals participation in investigational therapeutic studies on a separate protocol.

Please note that a simple way to measure the clinical value of this triaging approach is to calculate the pathological CR rate for the entire triaged group. If this approach indeed selects patients for the most appropriate treatment modality, then the overall pathological CR rate for the group must be significantly higher than the well-established historical pathological CR rate of 25 percent that is achieved if we treat all patients with our current best chemotherapy.

**PRE-OP THERAPY IN BREAST CANCER**  
**35\_SESSION 6\_4 - PUSZTAI**

You may be interested to see how chemo- and endocrine-sensitivity predictions relate to each other. We applied the 200-gene endocrine sensitivity index and the 30-gene pathologic CR predictor to the same 126 cases. Approximately half of those with low sensitivity to endocrine therapy were predicted to accomplish pathological CR. On the other hand, almost no patients with intermediate or high endocrine sensitivity were predicted to achieve this favorable response to chemotherapy. These observations are consistent with clinical experience and with the reports of other pharmacogenomic research groups.

An important question remains: how to treat those who are predicted not to do well with existing therapies? One simple approach is to treat these individuals with the best current standard therapy plus a promising new drug, for example, bevacizumab. However, it is also possible to design more molecularly-tailored studies that also test the predictive value of a particular gene signature.

This slide shows an example. Dasatinib is a multi-targeted kinase inhibitor. It binds at least 19 different kinases, including the src family. Bristol-Myers Squibb also developed and published a genomic predictor of dasatinib response. Using gene expression data, it is possible to simultaneously calculate several different potential predictors for dasatinib response.

For example, one can measure the expression of all the targets of dasatinib, and calculate the combined target index, to identify those who have a very high target expression. One can also examine the activity of a src pathway in each tumor, by calculating the pathway index using all of the published literature. And one could also apply the Bristol-Myers Squibb predictor.

**PRE-OP THERAPY IN BREAST CANCER**  
**35\_SESSION 6\_4 - PUSZTAI**

We have performed all of these calculations and superimposed the results on the chemotherapy response prediction. This slide shows a two-way scatter-plot of the dasatinib target index on the x axis, and the BMS prediction scores on the y axis. Each dot represents one tumor. Cancers predicted not to achieve pathologic CR to paclitaxel-FAC are marked with the small black squares around them.

These individuals, if they also have a high score on at least one of the three distinct dasatinib predictors, could be considered eligible for dasatinib in combination with standard chemotherapy. The goal of such a study would be to increase pathological CR rate to significantly above the rate that is predicted with standard chemotherapy alone, which is about five percent.

Clearly, this design did not address the activity of dasatinib or any other new drug tested in this manner in patients with other molecular characteristics or in unselected patients. However, this can easily be done with some modifications to this design, in a separate study.

So, in conclusion, I would like to stress that genes are not independent variables. There are large-scale, coordinated expression patterns that are associated with clinical variables. This helps to discover “general chemotherapy sensitivity” signatures, but, at the same time, it hinders the discovery of regimen-specific markers.

Microarrays give a reliable and reproducible snapshot of global gene expression status of breast cancer. This allows building a large molecular database that will be an invaluable resource for discovery and hypothesis testing. Raw data must be made public, and a uniform platform is desirable if we want to accomplish this. I believe the time is right to

**PRE-OP THERAPY IN BREAST CANCER**  
**35\_SESSION 6\_4 - PUSZTAI**

start to test the clinical value of response predictors, be they are single genes or multi-gene predictors, prospectively.

Oncotype DX, our [MD Anderson] 30-gene predictor, topo-isomerase II, luminal A molecular class, and several others are available to be tested in clinical trials. I think individually moderately accurate predictors may be assembled in the future into a clinically relevant diagnostic strategy. Thank you.