Presentation by Soonmyung Paik, MD

DR PAIK: I think the bottom line when you are considering the predictors of response to targeted therapy is the premise that if you have more target, you should have more benefit. So any data that depart from this premise should be looked at with suspicion.

Let me give you kind of a proof of principle of that concept.

In B-14, an NSABP trial evaluating placebo versus tamoxifen, if you view the tamoxifen arm of the patients, if you look at the amount of the target with the right assay, you should expect this kind of diagram, or curve, that shows poor prognosis with low ER and good prognosis with higher ER because of a linear response to tamoxifen.

But if you consider the data generated by immunohistochemistry performed centrally — maybe I did it wrong, but essentially what it showed is the completely nonlinear correlation between the amount of ER by protein measurement and the degree of benefit.

If you view the same data measured with real-time PCR, as part of the Onco*type* assay, there was an absolutely linear correlation with a very strong interaction *p*-value when you looked at placebo versus tamoxifen. So we do have proof-of-principle data that show that if you measure correctly, the amount of the target should be predictive.

Now, when you consider the B-31 data, the trastuzumab data, in the adjuvant setting, they are totally "heretic" in the sense that there was absolutely no interaction, no matter how you look at the data, by DNA copy number or amount of the protein. So the suspicion here is that maybe the assay is bad.

There were two presentations at the San Antonio meeting, one by NCCTG and the other one from the HERA trial, essentially evaluating the same question.

And the data are very striking in the sense that the trend for benefit seems to be there regardless of how you dissect the data — by immunohistochemistry or FISH or the presence of the polysome or any "ploid" in chromosome 17. So even for the immunohistochemistry-negative patient, there seems to be a trend for benefit.

And if you consider the copy number, or DNA, the FISH, by ratio, indicated no trend for any interaction showing smaller DNA amount, less benefit.

The other question was polysome 17 because of the CALGB data presented by Dr Kaufman showing that in the advanced disease setting, maybe the polysome was the determinant for response. But in the adjuvant setting in the NCCTG data, there was no interaction.

NSABP data regarding polysome 17, we have exactly the same finding — no correlation with degree of benefit. So regardless of whether you have polysome 17 or not, there is a benefit.

In the HERA trial, they also evaluate the central assay results. In this case, everybody had to be central assay-positive to be eligible for the trial, either by IHC or FISH. So you cannot really look at the real negative cases. But regardless, you could look at the different categories defined by immunohistochemistry or FISH. And you can see that there was no interaction between any assay result and the degree of benefit.

And here is another look at the data, similar to the NCCTG data. There seems to be no interaction between amount of the DNA copy for HER2 and degree of benefit. In fact, there seems to be a trend going the opposite direction — high copy number, less benefit. Obviously, there is an interaction *p*-value, but it is not significant.

So in the B-31 study, if you dissect the data in any way that you want, they are significant. They show no evidence for the interaction.

So the test, the scientific question there is, is it a testing issue, that maybe FISH and immunohistochemistry are just bad assays?

Or, similar to what we saw in B-14 for estrogen receptor staining, is it just because of the artifact of subset analysis, that imbalance of certain clinical factors or other covariables that we don't understand? Or is there a completely different mechanism of action between the trastuzumab used in the adjuvant setting versus the metastatic setting?

So we first examined whether we can prove that our central assay was wrong. We performed a microarray analysis on 400 cases randomly selected from B-31 and evaluated the differentially expressed genes between central assay, negative versus positive disease.

If you match the data to each chromosome — so 44,000 "proops" to each chromosome — and now plotting the distance between the centers of the chromosomes for the degree of difference, then the color we presented in blue is an extremely significant difference. And you can see that the only place where there is a coordinate underexpression of genes is in chromosome 17 at the HER2 locus in these tumors that we call central assay-negative. They do not have HER2 amplification, biologically. So they are really negative cases.

And if you look at individual RNA level, for example, HER2, compared between central assay-negative versus positive, the *p*-value is zero, even after adjustment for 44,000 comparisons. This is a true difference. There is an underexpression of the HER2 gene, as well as if you view the adjacent gene in the same amplicon: Again, very significant underexpression. So there is direct biological evidence that the assay was not wrong in doing what it is supposed to do. These are FISH-negative cases in reality.

On our assay we performed what we call a branched DNA capture of the RNA. Essentially, when you consider the amount of the RNA for HER2 in HER2-negative by FISH and IHC or negative — positive by both, or either of them positive — there is a linear correlation. But when you evaluate other trials, such as B-30, and randomly select the cases and do the FISH, you can see that the distribution of the HER2 signal in what's called HER2-negative in the trial is identical to the negative cases in B-31. So, there is no selection difference, although they were originally called positive by some labs. Indeed, they have no difference. And this, the HER2 RNA level-wise from any HER2-negative cases in any other trial.

But you can see that there is a high overlap of the HER2 FISH-negative cases and the FISH-positive cases. So there is a possibility that if we measure RNA level in the correct way, maybe we can find a better cutoff than what FISH shows.

We did a microarray analysis, so we looked for other genes that are predictive for trastuzumab response. We found about 300 or more genes that interact at a very significant level with trastuzumab benefit. Here's one example. In the C-MYC locus, it's not C-MYC itself. That has a linear prediction or a correlation with a trastuzumab response. The upper one third level shows a high degree of benefit, the lower one third almost no benefit. But there are about 300 of them, and they intersect with HER2 RNA at a right angle, so there's no correlation between the two. And we don't know which one will be actually more predictive when measured correctly.

This all might be because of the way we are measuring these assays. If you measure with PCR, HER2 might be more predictive, for example. So a lot of research has to be done in this way.

One promising technology that was presented at San Antonio was the Veritag Assay, or proximity assay. Essentially it works by this principle: Say you have two different antibodies targeting a molecular at about 200 nanometers apart in physical space, and the antibodies are designed in structure so that one has what they call a molecular seizer that can be activated by a red light, and the other one has a reporter that can be measured on the mass spec. Essentially, when they are very close to each other, the seizer will cut the reporter from this antibody. So you can demonstrate the proximity of two different moleculars, or a single molecular, and then you can quantify the amount of the proximal proteins. It can be used for the total amount of HER2: If you have two antibodies binding close to each other on the same molecular and you have a quantitative assay, or you can target two different moleculars, let's say HER2 and HER3, or HER2 and HER2, and when they are only making the heterodimer or homodimer, then they will cut out the reporter, so you can measure them also.

On this poster, essentially, they showed that they can measure that in the paraffin block and show the correlation not only in the cell line but also in the clinical samples. And it seems to be a highly reproducible assay on its own.

And if you compare it to immunohistochemistry in the paraffin block or FISH, there is a linear correlation, but the correlation is not perfect. We see the same situation for the RNA. The scatter is fairly wide. And when you compare the RNA level with this protein assay, there is a linear correlation, but there is a scatter. So which one is more linear, I think, will be the final determinant of which assay is the best.

So what they show in this study actually is that when they examined HER2 positivity by these assays, the amount of the total HER2 protein itself seemed to be somewhat prognostic, although it was a highly exploratory analysis, hunting for the cutoff point.

This needs to be validated, so we are conducting a study in B-31, evaluating approximately 800 cases, and we are currently analyzing the data. What we know is that there is a good correlation with the RNA level, and RNA level was not highly predictive of the response. So we expect that it's not going to be predictive of response, either, but we'll have to see.