A New Approach to Functional Genomics

Bruce Blumberg

University of California, Irvine

Thank you very much, Dr. Kanno. I would like to express my appreciation to the organizers for the invitation to this meeting, and particularly Dr. Kanno and Dr. Inoue.

What I would like to talk about today is what I think is a relatively new method that we have been working on in my laboratory. One goal of our research is to bridge the gap between genomic studies and high throughput screening. In order to do that, we need to identify what are the new target genes through which chemicals act, and which we can use to better understand the way in which chemicals affect the living organism.

The fundamental problem is that there are lots of genes coming from genome sequencing, from EST sequencing, and we would like to know what is the function of these genes.

Much of what is currently practiced to understand gene function relates to when and where the genes are expressed. This primarily works when you already know what the genes are and you are able to construct chips with known genes such as the Affymetrix GeneChips and various flavors of microarray chips.

There is some ability to study novel genes by this method, but it is not nearly as well developed. One way that we think is very potentially fruitful to understand gene function is to identify what genes interact with. So if you have a target of interest, what are the other gene products that interact with that target?

There are a few ways you can do that. The classical biochemical method is an old standby and that works well, but it is not very high throughput. There are a variety of yeast-based screens which can be used for this: phage display, expression cloning, and the whole new field of proteomics has application to this.

The fundamental problem is that inside the cell, most genes interact in large complexes, so it is very rarely the case that you have one-protein acting in isolation to accomplish a function. This slide uses the Wnt signaling pathway to illustrate this.

As you can see, you start with a signal, which is an extra cellular molecule, it acts through a receptor which is already a complex of a large number of molecules, you have an intracellular signal transaction pathway that gets even more complicated with multiple interactions. It is a very, very large macromolecular complex now, and you eventually wind your way to the nucleus, where there is a transcription factor that directly modulates gene expression.

If we only knew what a few members of this pathway were, how could we rapidly figure out what are the others? Because these other members of the complex may turn out to be important new targets for pharmacological and toxicological studies and for potential activity as endocrine disrupting compounds.

It would be desirable that we can do this quickly on the whole genome basis because it took more or less 20 years for dozens of laboratories working around the world to generate this picture.

So the kinds of methods that typically get used are phage display, where you express a fusion protein on the surface of a phage particle and ask the question, “Does this interact with some target of your interest?” This is a very valuable method and gives you good information. It is a binary system. There is one protein interacting with one other protein.

A good feature is that you can manipulate the biochemical stringency, which you cannot with all types of methods. It does take lots of attempts to get this optimized so that it works, and it is not usable for

all targets, and the likeliest result is that it takes a long time to get through the whole process to end up with your high affinity targets.

You are probably all familiar with the yeast two-hybrid system, where you are getting to use the fusion protein of a bait and a DNA binding domain, and you ask the question, "If I fuse some other fragment of a protein to an activation domain, which ones interact with the bait to reconstitute an active transcription factor and lead to the activity of a reporter gene.

Although this is very widely used, it has a number of what I think are difficult problems to overcome. It is also a binary system. You cannot manipulate the binding parameters; you are stuck with whatever is the intercellular nature of the yeast cell. You cannot detect interactions between proteins and complexes. Again, it takes a long time and lots of false positives.

Expression cloning is a method that is perhaps a little closer to the type that would be ideal. In expression cloning, you typically use large pools of cDNAs and test these pools for the presence of a function that you are interested in, and there are lots of ways that that is done. When you find a positive pool, then you separate and retest until you end up with a single component that has the activity that you want.

This is somewhat better than the other approaches in that it is completely functional, you can do it *in vivo*, and it works with secreted proteins and receptors, which all methods do not. Of course, the disadvantage is that it is not very fast. There is lots of screening and re-screening and you see the same molecules over and over again. There are sensitivity issues due to the size of the pool, and there is extensive retesting.

We have developed a new method that we call molecular interaction screening. The principle of this method is that instead of using large pools, we use relatively small pools of cDNAs, and use these to produce pools of protein, which are then assayed for function. There are lots of functional assays that can be used.

Our base material is a collection of cDNAs. It can be a cDNA library, or better yet a collection of individual genes, which are sometimes called UniGenes. These are in 384-well plates. We, then, use laboratory automation to pool those source plates into daughter plates. The size of the pools can be optimized — we settled on 96 as a very convenient size.

In an automated way, we make the bacteria, prepare DNA pools and use those to prepare protein. This protein is then used in the functional assay. So once you get a positive in this screen, the next step is to take that single positive and unpool the components and test it again. These positive pools, remember, since you have made them from a known source material, have a defined composition. So it is only a two-step process to get from the beginning of the screen to the end.

Here is a cartoon depicting how the pool looks. We have individual plates here and these are pooled into a daughter plate. We have 96 pools of 96. The current way that we do this assay is a radioactive scintillation proximity assay. We make radioactive protein *in vitro*, then ask the question "Can this protein interact with a target that we have immobilized on a special kind of plate?" When there is an interaction between something in the pool and the target, you get a scintillation event that you can detect.

This assay has a number of very good points. One of the important ones is that this target can be anything that you can produce. It does not have to be a fragment of a protein. It can be of any complexity. The second is that this is what we call an equilibrium assay so there is no washing or removal of unbound components. So it is insensitive to proteins that have a high off rate.

As I said, the advantage of this is that we can have an arbitrary size and complexity of the product. You can imagine that if you can form a macromolecular complex, that you can use that. This pool of cDNAs is relatively normalized: it has 96 components each present in approximately the same amount. That gives you a better chance to see molecules that would be ordinarily in low abundance.

Various end point assays are possible. Our favorite assay currently is a radioactive assay, but you could just as easily use a fluorescent or a luminescent or a Biacore type assay. An important point is that you can do saturation screening of the genome. Our system can do 150,000 cDNAs in a screen. The most complex genome, the human genome has, more or less, 45,000 genes. So we easily have the capability to saturate that.

The screening process has only two steps from start to finish and takes around two weeks. The downside is that you need the equipment and the consumables are not inexpensive. But at the end, the cost of the screen is really very low compared to the cost of labor by doing it in the traditional way.

So what kinds of screens are possible? It is certainly possible to do the standard protein-protein interaction assays in the same way that you do them with yeast or with phage display or with proteomic assays.

One of the areas where we have an advantage over other methods is that we can detect an interaction between a protein and a complex of proteins. My lab works on nuclear hormone receptors; you can easily use for instance a receptor dimer as the target, or you can go even more complex than that: you can make a protein and macromolecular complex. For instance, a nuclear receptor heterodimer bound to its target DNA sequence in the presence or absence of the ligand.

You can also do nucleic acid protein interactions to identify enhancer binding proteins or RNA binding proteins, and to detect interactions between the cDNA encoded proteins and small molecules that you can immobilize on the plate. Such interactions are not easy to detect by other methods unless you have the ability to make high specific activity radioactive compounds.

The assay that we are working on in the laboratory concerns the retinoic acid receptor. As you may know this is a nuclear hormone receptor and exists in three types, α, β and γ.

An interesting feature is that all three of these types of receptors can occur in the same cell at the same time. They all bind to the same natural ligand, retinoic acid, and they all bind to the same high-affinity DNA target elements. But there is a lot of evidence that the target genes that they activate differs from one receptor to another. If the receptors each bind the same DNA sequence and the same hormone, where does that specificity come from?

It is commonly thought that that must derive from interactions with cofactors, but which cofactors? None of the known receptor coactivators or corepressors even discriminate between different nuclear receptors, let alone between closely related receptors like RAR α, β or γ. So it is our great interest to do *in vitro* identification of cDNAs encoding proteins that can interact with the whole heterodimer in the presence or absence of the hormone.

Here is a cartoon showing how such an assay looks. The DNA element with the pre-bound RXR-RAR complex is attached to the plate. There is a mixture of radioactive proteins, and only when you can get something that specifically interacts with this complex will you detect a signal.

Here is how the data from such an experiment looks in a different way. This is 96 pools of 96 proteins in each pool. This pool here has a known interacting protein in it, and the goal of this assay was to detect which of the other pools had another known interacting protein in it. We did this blindly, and you can easily see there is one that stands out above the background.

In vitro assays are very nice, but we would also like to be able to do *in vivo* cell-based assays. The system is also amenable to doing those kinds of assays. Instead of making protein *in vitro* from these pooled DNAs, we can put those into cells and then test the function.

For instance, if you put these pools into a cell, is a reporter gene activated? Is the growth of the cells promoted or inhibited? Do the cells now have a ligand or a receptor that they did not previously have? This is important because cellular pathways have the property that they interact with each other and they sometimes do it in unexpected ways.

Here is the Wnt signaling pathway that I used before. In simplified fashion, you can see that pathway here, and this shows that there are several other pathways that can interact with the same ligand receptor complex to get entirely different results.

So we would like to be able to identify what are the other cellular pathways that can interact with this one to either promote or repress its activity. You do that in a very similar way. You make an engineered reporter cell that has the ability to report when you have affected the terminal event in the pathway, and you detect that by typical reporter gene assay; we like luciferase. You put your pools of DNA in the cell and ask the question; "Do we now see a response?"

So what kind of screens can you do using this methodology? You can do functional detection of enhancer interacting proteins, which is a relatively difficult method to do by other techniques. You can identify new components of a cellular signaling pathway. For instance, the FGF signaling pathway, although very important does not have well-defined components.

Also, screening for drugs that modulate particular components of a pathway. One of the ongoing projects at UCI is work on Wnt signaling in colon cancer. Because my background is in hormone receptors, we can also do assays that will identify small molecules, drugs, natural products, and endocrine receptors that affect the activity of these receptors.

The assay that we are doing to validate the cell-based assay uses a nuclear receptor that we have identified called the steroid and xenobiotic receptor, SXR. This receptor has target genes that are cytochrome P450s, in particular the CYP3 family.

Those of you who have worked with P450s know that they have the unhappy property that the promoters are not active in cell lines. You can only use primary hepatocytes to measure the activity of these promoters. They are troublesome, expensive, and someone has to die in order for us to work with primary hepatocytes.

So, it is fairly clear that there are one or more factors missing in the cell that would confer the ability of this promoter to be active. Our approach is to, then, transfect these cDNA pools from a liver library into a hepatocyte line — and we are using HepG2 — containing the P450 promoter linked to luciferase, and ask the question "Which of these now will enable this promoter to be active?"

Here is how the system looks. The system is based on a CRS robot arm. We have a variety of liquid handlers and automated carousels, and it has the capacity to do 400 plates in an unattended manner; you just load it up and go. It is designed to do *in vitro,* biochemical, and cell-based assays. Also, a necessary thing that we need to be able to do is to convert bacteria into a plated form amenable to these types of experiments. Our patent was just issued and we are working very hard to make this technology a reality.

In summary, this robotic molecular interaction screening system that we have developed is, I would argue, an ideal system to identify gene function. It is truly functional genomics in that we are identifying genes strictly by their function, by their ability to interact with other genes of interest.

It can be broadly applied to biological problems. We are using it primarily to work with nuclear receptors, but it is certainly possible to apply it to other transcription factors and even to extracellular molecules.

One of the things that we like about it most is that it is capable of detecting interactions with a target of arbitrary complexity. That means it can be one protein, two proteins, five proteins bound to DNA. It does not matter. As long as you can produce the complex, you can detect proteins that interact with it. You can do both *in vitro* and *in vivo* screens, and it is very rapid — it takes only two to three weeks from start to finish.

The availability of full-length cDNA sequences, either from cDNA libraries or collections such as have been developed at Riken makes this system much more useful in the future, rather than much less useful.

My time is just about up, so I am going to skip to the end and acknowledge my collaborators. There are three very talented students in the lab, Amee Patel, Bridget Riggs, and Gaurav Sharma are the ones that have been developing this assay.

I would like to stop there and answer any questions that you might have.

Q&A

Kanno : Thank you. Any questions, any comments please?

Q : Bruce, what happens if more than one protein is required to restore activity, let's say for 3A4 induction in Hep2G cells?

Blumberg : You mean if one molecule is insufficient…

Q : Required but insufficient, yes.

Blumberg : Because we are able to screen pools, we have the possibility to combine those pools in different ways and to constitute perhaps more than one factor that you would hypothesize is needed for activity. It is my bias that it is a single protein, but it is always possible that it is more than one.

Q : This may be a naive question, but how does the phosphorelation status of a protein factor into the interaction?

Blumberg : It is not a naive question; it is actually a very good question. Certainly whether a protein is phosphorelated or not can affect its interaction with other proteins. If you wanted to test that, you would modulate the phosphorelation state of the target. You could produce it in a phosphorelated or unphosphorelated state, and could then look for proteins that interact specifically with one state versus the other.

Q : So that would add another layer of complexity then…

Blumberg : Yes.

Kanno : Thank you very much.