AFFYMETRIX
MICROARRAY
BULLETIN

JULY ISSUE: ADVANCES IN CANCER RESEARCH

EXPRESSION PROFILING ▪ COPY NUMBER ▪ SEQUENCE ANALYSIS

July 2005

Microarray Sequencing of Mitochondrial Genome Uncovers New Clues for Cancer Detection *by Wes Conard*

Johns Hopkins' Dr. Anirban Maitra and Duke University's Amanda Baumann weigh in on emerging methodologies and technologies for mitochondrial studies of cancer.

Scientists led by Dr. Anirban Maitra, assistant professor of pathology, genetic medicine and oncology at Johns Hopkins Medical Center are using a new microarray-based method to rapidly sequence the genome of human mitochondria to find mutations linked to cancers, including prostate and pancreatic cancer. The microarray, called the MitoChip, sequences more than 12,000 bases of the mitochondrial genome, and takes less than a third the time of conventional sequencing methodologies.

"Right now we're using the MitoChip to look at a larger number of pancreatic tumor and juice samples to see if we are able to detect tumor-specific mutationsin pancreatic juice samples," said Maitra. "While we developed the MitoChip to study cancer biomarkers, it certainly has uses in non-cancer applications as well, everything from evolutionary studies to metabolic diseases to aging."

Maitra recently spoke with Amanda Baumann, a pre-doctoral fellow in Duke University's department of Pharmacology and Cancer Biology regarding the new challenges facing mitochondrial research now that the genome sequence can be readily accessed. The two discussed:

- The various types of mitochondrial mutations linked to cancer and models available for validation
- The "mass advantage" of mitochondrial genetics and the associated difficulties of distinguishing heteroplasmies
- The sensitivity, sample preparation, and genome coverage of microarray-based resequencing technology

Formerly at Emory University, Baumann worked with Dr. John Petros to prove that mitochondrial DNA mutations increase tumorigenicity in prostate cancer. The group published their work in the January 2005 Proceedings of the National Academy of Sciences, USA.

MITOCHONDRIAL GENOME FACT BOX

Size:	16,500 bases
Coding sequence:	Over 12,000 bases
Mitochondrial mutations:	Present in 70 to 90 percent of cancers
Cancer histologies:	Include breast, lung, prostate, pancreatic, head, neck, and bladder
Nuclear mutations:	p53 is mutated in only about 50 percent of human cancers

Mutation discovery and in vivo models

Baumann: Did you use the MitoMap website and the Cambridge Reference Sequence to identify the mutations that you published in your 2004 Genome Research publication? Were they all amino acid altering or did you just distinguish them as polymorphic versus non-polymorphic base changes?

Maitra: Actually, it was a mixture of synonymous and non-synonymous changes. We were primarily interested in finding differences between normal and cancer samples. In this initial study it was not that important to us if the mutation directly led to an amino acid change, but rather if we could identify nucleotide variations between matched normal and cancer samples. From a biomarker perspective, finding such base pair differences between normal and tumor is the most important issue.

Having said that, if you are actually studying, for example, genotype-phenotype correlations or doing a bit more of a functional type of analysis, then that information becomes very important. And we did find many of those examples that had been referenced in MitoMap, or in previous studies in the literature. But at the same time, we found a bunch of other changes as well. I'm not sure what the significance of them is, except that they were not present in the matched normal tissue.

Baumann: We were absolutely amazed at the number of amino acid altering mutations we found when we started sequencing whole mitochondrial genomes. How can we find more of an in vivo model for some of these mutations? Obviously, it's not currently possible to really make a transgenic mouse where you can alter specific bases in the mitochondrial genome.


Amanda Baumann

Maitra: One of the biggest challenges has always been attributing exactly what these mutations mean in cancer. The implications of mitochondrial mutations are far better known in many metabolic diseases, but the implications for cancer pathogenesis and progression have always been somewhat in doubt.

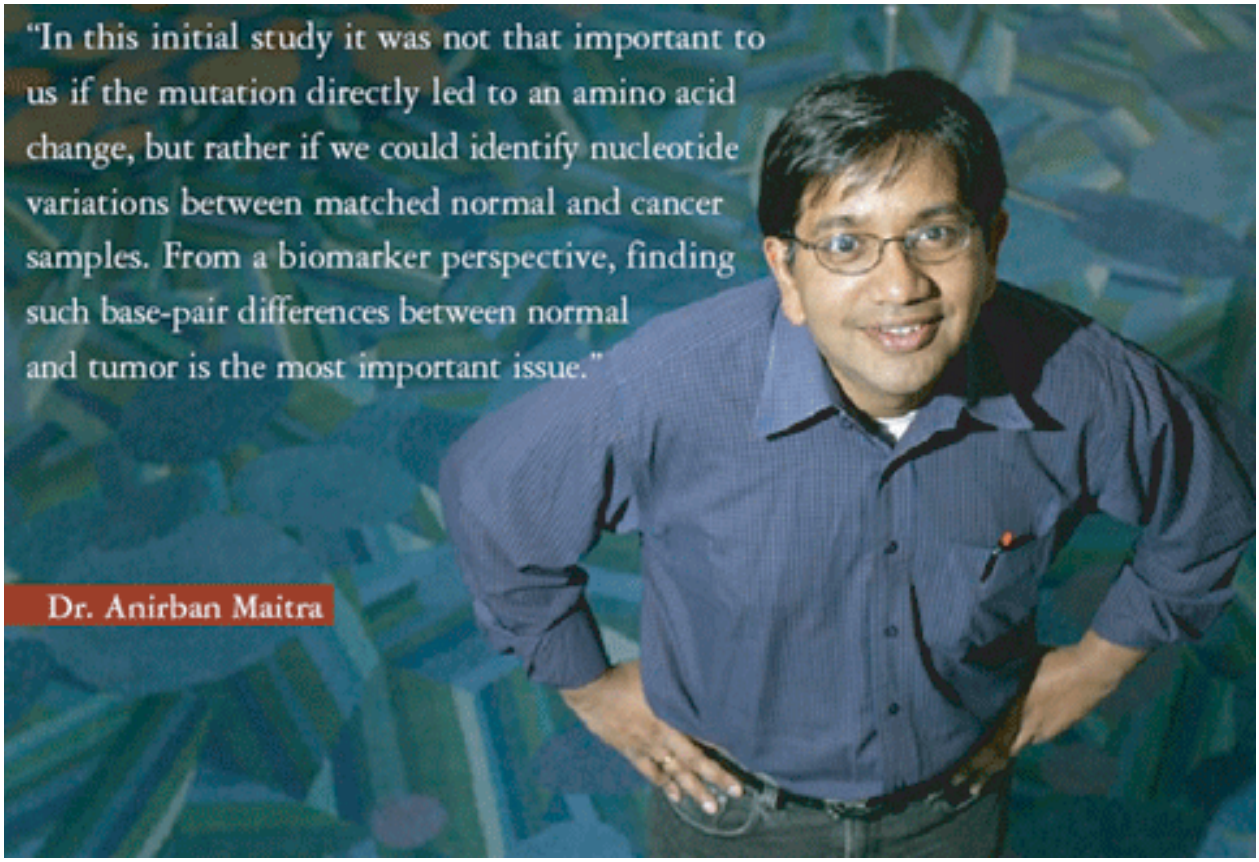
The first really conclusive evidence actually came from John Petros' work, that you obviously were involved in, which showed that when you over-express a mutant mitochondria in a cancer, you were able to make it grow faster in vivo, while if you express the wild type, you actually make it grow slower than it normally does.

I suppose the final sort of proof will be when we are able to overcome the technical challenge of making a transgenic mouse model of a mutant mitochondrial gene and show that these mice are more predisposed to cancer, aging and so forth. But just the demonstration that you can take a single point mutation in the mitochondrial gene and can actually increase in vivo tumorigenicity is, by itself, a very nice demonstration of functionality.

The "mass advantage" and heteroplasmy

Maitra: The 'mass advantage' of mitochondria means that you only have two copies of every gene in the nucleus, but you have many more copies of mitochondrial DNA. So if you're dealing with limited clinical samples, for example body fluids that have limited amounts of DNA, you have a better chance of detecting mutations with mitochondria than nuclear DNA. This has been shown by several groups at Johns Hopkins in a variety of human cancer types.

Baumann: But the quantity of DNA you're able to obtain with mitochondria may also create a disadvantage in distinguishing between different types of heteroplasmy. Within a mitochondrion you might have both wild type and mutant base pairs or maybe within a cell, you have different mitochondria that are wild type and mutant. Even within a tissue, you may have cells that are wildtype or mutant for certain mitochondrial DNA mutations. How are you able to distinguish between these different types of heteroplasmy?



Maitra: The problem at this point is technically challenging to answer: `Is there a difference if you have an intracellular heteroplasmy versus an intercellular heterocellular heteroplasmy?'

You'd have to fall back on in vitro studies where you basically make cells with designer mitochondria (so called "cybrids") within which you can actually titrate the dose of heteroplasmy. Then you can say, 'Okay, if I have this many mitochondria in this cell that are mutant versus this many that are wild type, is there a difference compared to mixed populations of mutant and wildtype cells?' These are questions that have to be answered in vitro using hybrid techniques that we have access to and gradually titrating the doses. But in terms of what happens in vivo, I don't think we have the expertise at this point to answer that question.

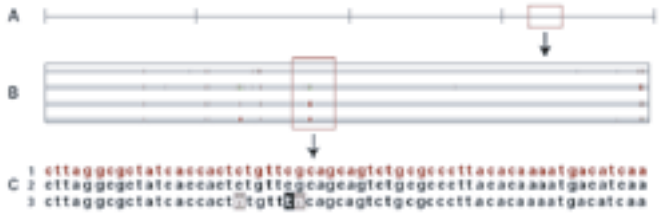
Baumann: I agree. The hybrid-hybrid cells are really the best in vitro system that we have, especially if you can get matched wildtype and mutant DNA, so you know the only nucleotide at which the genome is different.

In terms of the heteroplasmic patients we looked at, something we talked a lot about was whether it might actually be better—instead of obtaining a pure wildtype and mutant cybrid—to have a 90 percent and 10 percent wild type across the spectrum to find the threshold at which a mutation becomes dominant.

Maitra: The cybrid technology allows you to actually plan some very elegant experiments looking at mutational thresholds that put a cell over the edge. You know, with the nuclear DNA it's either heterozygote or homozygote, so you're dealing with basically a binary system. But with mitochondria it's sort of a continuous bell curve, if you will, and so the experiments you have to conduct are quite different than what you would do with a nuclear gene.

Array-based mitochondrial genome sequencing

Baumann: What does the MitoChip offer to researchers who are currently doing manual and automated DNA sequencing?



Maitra: We discussed several of the advantages of the MitoChip in our Genome Research paper, but there are three key ones worth recapping:

First, we cut down the number of polymerase chain reactions (PCR) required to sequence the entire mitochondrial DNA to three. As you know, most routine mitochondrial protocols will use anywhere from 20 to 40 individual reactions.

Second, we reduced the amount of starting template that you need. We cited 100 nanograms in the paper, but we were able to amplify mitochondria with as low as 10 nanograms of starting template. I think that's because on a per cell basis you're dealing with a lot more DNA than you would with nuclear DNA.

Third, of course, now you're able to sequence the entire mitochondria on a single chip without having to pour over 16,000 bases of chromatograms. The first generation MitoChip we described had the entire coding sequence short of the D-loop. Now, Affymetrix has developed a second generation chip that has the entire mitochondria on it, including the D-loop, as well as redundant tiling for deciphering mitochondrial haplotypes.

We've been collaborating with an investigator here at Hopkins, David Cutler, who has a lot of experience in designing high-throughput genotyping analysis software. Dr. Cutler has developed an algorithm for automated sequence analysis of MitoChips; his software is available as freeware online. Affymetrix also has its own software analysis platform that will basically spew out the mutations for you in a beautiful Excel format, with the exact position, the wildtype sequence and the base change.

Then the other thing is the time factor. It takes about one and a half days for a fulltime technician to get from starting template to a scanned chip with 16,000 bases of associated information. And you can multiplex eight chips at a time on most Affymetrix wash stations, so obviously that significantly cuts down on the time required as well.

Sequencing method sensitivities

Maitra: The sensitivity of the assay is particularly important when you're looking at this as a potential biomarker detection tool. We showed that this assay is more sensitive than some of the more conventional fluorescent dye-sequencing technologies that are out there. We've basically done sequential dilution studies and shown that in a mixed population of normal and mutant cells you can detect mutations—80 percent of the mutations in a sample—even at a dilution of 1:50 (i.e., when only two percent of the cellular population is comprised of tumor cells and 98 percent are normal, wildtype cells). I think the limit of detection for fluorescent dye sequencing technologies is about 10 percent. We've managed to achieve about five-fold better rates.

It's well known that allele-specific hybridization techniques are always going to be more sensitive than fluorescent dye-sequencing technologies. While it used to be the gold standard to prove mixed populations of cells existed within a sample, it's really time consuming and tedious. Affymetrix technology basically is allele-specific hybridization on a chip, so it's not surprising that it's more sensitive than regular fluorescent sequencing techniques.

What's interesting is that we have been collaborating with the National Institute of Standards and Technologies (NIST), located in Gaithersburg, Maryland, and they actually used a two stage PCR base-sequencing protocol that they developed to sequence many of the same tumor samples that we described in the original MitoChip paper. So they had these chromatograms, and when we gave them the MitoChip data, there were several mutations they had actually not found using their sequencing software. Anyone that has worked with chromatograms knows—especially with 16,000 bases worth of chromatograms—that you're bound to miss small heteroplasmies. When the NIST group went back, they found that every one of the heteroplasmies was present; they were just little blips that were not picked up by manual scanning or even by automated sequencing software. That paper just came out in the May issue of Journal of Molecular Diagnostics.

Baumann: I definitely agree. Those chromatograms can be very time consuming to read and it's often difficult to know what is background and what might be considered heteroplasmy. From just a data collection point of view, being able to verify and obtain more accurate results has to be one of the biggest advantages.

Sample preparation

Baumann: What types of samples are you able to sequence off the microarray?

Maitra: Well, we've used either snap frozen samples or body fluids like urine or pancreatic juice. Certainly, body fluids work very well; there's enough mitochondria floating around to amplify. We've actually never tried paraffin embedded or laser-capture microdissected samples.

We've designed the assay for long-PCRs and there are three fragments about 5 to 6 kB long. So, as long as the DNA quality is adequate enough to get those long products, the assay should work. If you can do laser capture of cryostat embedded sections that should do pretty well. But unless the existing long-PCRs are modified, I don't know at this time if paraffin embedded, archival material will work, so that's one drawback.

Baumann: We used about 18 sets of PCR fragments to span the whole genome. How did you choose the three sets of PCR fragments?

Maitra: I designed the primers, made sure they worked, and then we went to the MitoChip. I designed these mitochondrial DNA long PCR primers using a web based protocol called 'Amplify' that I found is the best long-PCR program on the web. It's free on the web and I highly recommend it; it really works very well.

I arbitrarily divided the mitochondrial DNA into three fragments and designed two alternative primers for each, such that we would have amplicons of 5 to 6 kB each. So we had six sets of primers, each of which was sort of an alternative pair covering three of the fragments. And of each pair, one of them worked beautifully.

We chose such long-PCR protocols for two reasons: number one was, of course, you wanted to cover the mitochondria in as few PCRs as possible, without having to do too many amplification reactions. And the way the chip works, you need to pool all of them in equimolar quantities. So if you have 20 PCR reactions, you'd have to clean them each up, you'd have to measure the DNA quantity, and you'd have to aliquot equimolar quantities from each fragment. The fewer reactions you have, the better it is for you to actually be working with the MitoChip, or for that matter, any DNA sequencing array.

The other more technical issue we had is that there are significantly long stretches of mitochondrial DNA insertions in the nuclear DNA—sort of like pseudo genes if you will. If you have too short an amplicon, you could potentially be amplifying a nuclear insert, as well as a true mitochondrial DNA. And if they are different in sequence, you'd think that you are getting a mutation or a heteroplasmy, but it is really nuclear DNA. To overcome that, we actually designed these long-PCR primers; we tested them on cells depleted of mitochondria; and we found that there was no product. So we know that these are mitochondria-specific primers.

We've also tested them on mouse DNA, and they do not amplify. So fortunately for us, they are also human specific. So, if you're looking at xenografted tumors, for example, you can safely use these primer sets to amplify, and you know that the only thing you are amplifying is human mitochondrial tumor DNA and not the mouse stroma.

Genome coverage

Baumann: You said that initially you had left out the D-loop in the MitoChip?

Maitra: Yes. However, the new second generation MitoChip has several differences.

For one, it has a smaller feature size; so it has about 50 kB of sequence as compared to the 32 kB that we had in the first generation chip. That 50 kB comprises the entire mitochondrial coding sequence—top to bottom, forward and reverse strand—as well as redundant tiling for mitochondrial haplotypes and known mutations from common websites such as what is referenced in MitoMap, and other online databases out there.

That gets me to one of the problems of chip-based hybridization chemistries, chips are very good at detecting single base changes, such as missense or nonsense mutations or single base polymorphisms; however, they're not as good when it comes to detecting insertion-deletions.

Affymetrix has included redundant tiling features in the new MitoChip that will enable researchers to detect potential insertion-deletion mutations especially in the non-coding D-loop region where these are not uncommon. The advantage of an iterative tiling strategy for regions that demonstrate frequent insertion-deletion mutations is that you can now actually look for the sequence that gives you the best hybridization and use that as a guide for detecting such mutations with greater confidence than you could in the older generation of arrays. In other words, having 50 kB worth of sequence to 'play with' has given Affymetrix the freedom to introduce additional features that I believe strongly enhance the utility of resequencing arrays.

We've now worked with this new MitoChip for about six to eight months using human cancer samples, and although we haven't published the data yet, it looks very good; for example, we are able to detect additional mutations that we missed using the older generation microarrays. A lot of the artifacts associated with, for example, dye bleeding through and so forth in the older generation chips have now gone as well.