Scientist Spotlight

Affymetrix®
Revolutionize life

Novel approach results in first ever rhesus macaque whole-genome array

Rob Norgren of the University of Nebraska Medical Center and Katja Nowick of Lawrence Livermore National Laboratory discuss the innovation behind the development of the Affymetrix GeneChip® Rhesus Macaque Genome Array.

Researchers at the University of Nebraska Medical Center (UNMC), in collaboration with scientists at Affymetrix, have discovered a novel way to derive sequences for microarrays, resulting in the first ever commercially available whole-genome rhesus macaque microarray. This advance in primate genome research may help researchers to better understand the evolutionary relationship between humans and other primates, and will increase the utility of rhesus macaque as a model organism for disease research.

Rob Norgren and his collaborators used human transcript annotation to design primers to amplify and sequence rhesus genes. In collaboration with Affymetrix, this strategy was extended to an *in silico* approach using information from the Baylor Rhesus Genome Project. This project resulted in a set of probes representing 18,296 rhesus/human orthologs, including transcript variants and more than 17,000 genes for the GeneChip® Rhesus Macaque Genome Array. The resulting human genome-derived macaque array was the first whole-genome rhesus expression array.

The array design was published in the January 23, 2007, issue of *BMC Genomics*. Experiments illustrating the reliability and validity of the array were published in the February 28, 2007, issue of *BMC Genomics*.

"The normal way a microarray is developed, you have to wait until you have a complete genome project and then you cluster the ESTs [expressed sequence tags] and come up with a consensus sequence to align with the genome project. By using our method, we were able to get the orthologous rhesus macaque sequence using human sequences to design primers," said Norgren. "It was a nice shortcut. As far as I know, this is the only Affymetrix microarray that has been made in this way."

To test the performance and reliability of the new array, Norgren sent aliquots of RNA from five different tissue sources to three different centers, selecting five of the most differentially expressed genes for analysis. He found that the reliability was nearly the same as that for human arrays and that qPCR confirmed the validity of the results.

Norgren recently spoke with Katja Nowick, a postdoctoral fellow at Lawrence Livermore National Laboratory in Livermore, California, about the development of the Rhesus Macaque Genome Array and the next steps for using microarrays in primate genomics. The two discussed:

- Using human genome information to develop the rhesus macaque array
- Validation of the GeneChip® Rhesus Macaque Genome Array
- The utility of rhesus macaque as a model organism for disease research and the study of primate evolution

Rhesus Macaques are model organisms with a close evolutionary relationship to humans, and are thus ideal for studying many diseases, including AIDS, cardiovascular diseases, and neurological diseases.

Robert B. Norgren, PhD, is a professor in the department of Genetics, Cell Biology, and Anatomy at the University of Nebraska Medical Center. He received his PhD from Columbia University and his research focuses on nonhuman primate genomics. He is primarily interested in using rhesus macaques as models for human diseases, particularly childhood neurological disorders such as Kallmann syndrome, ataxia-telangiectasia, and Lesch-Nyhan disease. His laboratory is developing novel approaches to microarray technology in hopes of better understanding the causes of these diseases and how best to treat them.

Designing the Rhesus Macaque Array

Nowick: You used an interesting and innovative approach to choose probe sequences for the Affymetrix Rhesus Macaque Genome Array. Would you talk a little bit more about why you chose this method and the selection process you used?

Norgren: When we started developing this array, there was very little rhesus sequence available. Because we didn't want to wait for the normal amount of sequence to be available, we leveraged the information from the Human Genome Project to target sequences for acquisition.

At the time, Affymetrix expression arrays were primarily aimed at the 3' ends of genes. Rather than wait for the genome data, we decided to target those regions in the rhesus genome and sequence them.

Initially, we ran a bioinformatics program and identified all of the last exons in the human genome. Then we used the Affymetrix probe selection region (PSR), which is primarily in that 3' end of the gene, and aligned those sequences with the last exon. Finally, we designed human primers that flanked the PSR and used them to amplify rhesus macaque genomic DNA.

The normal way a microarray is developed, you have to wait until you have a complete genome project and then cluster the ESTs and come up with a consensus sequence to align with the genome project. By using our method, we were able to get the orthologous rhesus macaque sequence directly from human genomic DNA. We didn't need a huge library of ESTs or a complete genome project; we could go directly after the sequences we wanted. It was a nice shortcut. As far as I know, this is the only Affymetrix microarray that has been made in this way.

We initially planned on sequencing 1,800 genes, but we expanded that to about 5,000 because the approach worked so well. Then we heard a rumor that some of the Baylor sequences might be available. They were at a very early stage of preparation. I contacted Baylor and it turned out that we were able to get some of their genomic sequences—a very early version of the rhesus macaque genome.

Then, in collaboration with Affymetrix, we did an *in silico* version of what we had been doing in the wet lab. We aligned the human PSRs with the available Baylor genome sequences to get a representation of pretty much all of the orthologous genes.

If we hadn't used this novel approach, there probably would not be a Rhesus Macaque Genome Array on the market today.

Nowick: What was your success rate with PCR and sequencing? I assume that not all of the primers worked the first time?

Norgren: We used a multi-step process. On the first pass-through, which was more medium than high-throughput, our success rate was 70 percent. Some primers failed because there was no band and some failed because there were too many bands. We grouped all the primers where we needed to raise the annealing temperature into one group and sorted all the ones that needed a lower annealing temperature into another group.

"We are just starting to explore the similarities in gene expression between rhesus and humans. When we compare the same tissues from humans on the human chip and rhesus samples on the rhesus chip, we get a very high degree of similarity in expression. The same genes are expressed."

Then we performed a second round of PCR and picked up more primers that worked. If any of those didn't work, we sometimes tried a third round of PCR. If by the third round, they didn't work, we designed new primers.

You are not going to get 100 percent success using human primers with rhesus because a mismatch between the human and macaque might keep a primer from working. Sometimes you have to design another set of primers.

Nowick: How does the Affymetrix Rhesus Macaque Genome Array differ from other rhesus arrays on the market?

Norgren: As far as I know, there is really only one other commercially available rhesus macaque array. Both are whole-genome arrays, so they both target all the genes. The main difference is in the technology that is used. Affymetrix uses multiple smaller probes instead of longer probe sequences. The Affymetrix array also features probes for a variety of pathogens.

In general, the Affymetrix technology allows you to look at alternative poly-As. Part of the reason there are more probe sets than there are genes on the Affymetrix Rhesus Macaque Genome Array is that there are more transcripts than there are genes.

Nowick: What were the biggest challenges or major pitfalls you experienced while developing this array?

Our NCRR funding was key. If we hadn't gotten the grant to get these sequences and develop the algorithms, we would not have been able to design this array.

Array performance and reliability

Nowick: How did you test the performance and reliability of the new array?

Norgren: We knew performance and reliability would be really important. There are eight different primate centers and we wanted to make sure that the array would yield reliable results between primate centers and among all the researchers using rhesus macaques.

We extracted RNA from five different tissue sources. We sent the same aliquots to three different labs. We independently hybridized and analyzed the data and did a variety of statistical tests to compare the reliability. The reliability of the arrays was high—about the same as has been published for human arrays at different centers.

To test validity, we performed qPCR with a subset of the genes to see if our results were similar. The results seemed to be valid. I have since heard from people who have used the chip that it's yielding both reliable and valid results.

Nowick: How did the rhesus macaque array performance compare to that of human expression arrays?

Norgren: We had data from a variety of people suggesting there was about 40 percent drop-off due to false negatives when using the human chip with rhesus samples. We expected the rhesus chip would detect significantly more genes with rhesus samples than the human chip did. That's what we found. Across all the tissues we looked at, there was a very significant increase in the percent present call.

Nowick: How will the development of this new microarray change your work? Will it affect your strategy for selecting and validating candidate genes?

Norgren: I think it's going to open up a range of new possibilities. We are interested in studying embryonic stem cells and working on redifferentiating fibroblasts in rhesus macaques. One of the things you can use this chip for is to get an index of "stemness." You can use it to profile a particular cell type. The more markers you have in terms of expression, the more accurate a target you have.

Rhesus macaque genetics and evolution

Nowick: How different are rhesus individuals from each other in sequence, expression, and physiology?

Norgren: Within rhesus macaques, there are a number of subgroups. In particular, there is a difference between rhesus

Katja Nowick, PhD, is a postdoctoral fellow in Lisa Stubbs' group at Lawrence Livermore National Laboratory. She received her PhD in biology at the Max Planck Institute for Evolutionary Anthropology in Leipzig. Her work focuses on human evolution, particularly using microarrays, bioinformatics tools, and statistics to compare humans to other primates in terms of sequence and expression differences. She works primarily on the functional characterization of a group of KRAB-Krüppel transcription factors that have changed dramatically during primate evolution.

macaques from India and those from China. Those differences are biologically important.

One of the major uses of rhesus macaques is for AIDS research. It turns out that when you infect the Indian macaque with a similar virus to HIV, called SIV, they get a disease that is more similar to the human disease than the Chinese macaques do. So there is a lot of interest in looking at expression differences between those two groups, with the hope that this difference might tell you something about response or resistance to viruses.

We are starting to look at the differences between individuals in another project involving SNPs. It will be interesting to look at those differences with respect to a variety of disease processes, including response to drugs.

Nowick: Rhesus is a model organism for many diseases, like AIDS, cardiovascular disease, and neurological diseases. Given the differences between rhesus and humans, how much do you think we can learn from this animal? How similar are they, for instance, in gene expression?

Norgren: I think we can learn a tremendous amount from rhesus. Of course, there are going to be some species differences between rhesus and humans. There will never be a perfect animal model, but of all the possible animal models that you can use for translational research, I think the rhesus macaque is the best. The chimpanzee would be even more similar to humans, but it's just not practical to do a lot of the experiments in chimps that you can do in rhesus. Rhesus is the closest practical animal model.

There are quite a lot of human diseases where the mouse is just too distant to yield the same kind of phenotype that you would see in humans. There really aren't very many good rodent models for AIDS, for instance. Rhesus is really the only model. As the emphasis on translational research grows, I think you will see increasing use of rhesus macaques.

We are just starting to explore the similarities in gene expression between rhesus and humans. When we compare the same tissues from humans on the human chip and rhesus samples on the rhesus chip, we get a very high degree of similarity in expression. The same genes are expressed.

Nowick: Rhesus is also interesting from an evolutionary standpoint, and is used as an outgroup to place human-chimpanzee sequence differences on the human or chimpanzee lineage. As someone interested in evolution, I think it would be interesting to do the same kind of experiments with regard to expression differences. Previous studies employing human expression arrays have been limited by the 5 to 8 percent sequence difference between humans and rhesus. How suitable would the rhesus array be for a project like this?

Norgren: I think it's important to redo some of those early studies with the rhesus array. I am concerned that some of the conclusions from those studies may be the result of false negatives. There was no way to control for that before.

The problems of using rhesus samples with human chips may have been underestimated by some researchers and might have skewed the results more dramatically than had been anticipated. It's important to use rhesus samples with rhesus chips if you are going to make statements about evolutionary differences in gene expression.

Future directions in primate arrays

Nowick: What are you working on now? Are you planning on developing more primate microarrays?

Norgren: Right now, we are working on getting sequences for African green monkey. I think the Rhesus Macaque Genome Array will probably work pretty well in African green monkeys, but they are different enough that it's probably worth getting African green monkey sequences.

We are also using the same primers that we developed for the microarrays to discover gene-specific SNPs for rhesus macaques. That's our big project.

Further reading

- Duan F., *et al.* Intercenter reliability and validity of the rhesus macaque GeneChip Array. *BMC Genomics* **8**:61 (2007).
- Ferguson B., *et al.* Single nucleotide polymorphisms (SNPs) distinguish Indian-origin and Chinese-origin rhesus macaques (*Macaca mulatta*). *BMC Genomics* **8**:43 (2007).
- Spindel, E. R., *et al*. Leveraging human genomic information to identify nonhuman primate sequences for expression array development. *BMC Genomics* **6**:160 (2005).

Affymetrix, Inc. Tel: +1-888-362-2447 ■ **Affymetrix UK Ltd.** Tel: +44-(0)-1628-552550 ■ **Affymetrix Japan K.K.** Tel: +81-(0)3-5730-8200
Panomics Products Tel: +1-877-PANOMICS www.panomics.com ■ **USB Products** Tel: +1-800-321-9322 www.usb.affymetrix.com

www.affymetrix.com Please visit our website for international distributor contact information.

For research use only. Not for use in diagnostic procedures.

P/N SS103 Rev.1