Scientist Spotlight

Affymetrix®
Revolutionize life

New technique improves the speed and precision of chromosome mapping

Yong-Jie Lu of Queen Mary, University of London and Colleen Elso of the Walter and Eliza Hall Institute of Medical Research discuss a groundbreaking method for rapidly identifying chromosome rearrangements and precisely mapping breakpoints.

Researchers at Queen Mary, University of London have developed a new technique that combines three previously available technologies to quickly and precisely identify DNA rearrangements associated with tumor formation and growth. This new approach, which utilizes M-FISH, high-resolution karyotyping, and exon array analysis, may help researchers identify critical candidate genes and genetic markers and better understand the role of complex genetic changes in human cancers.

Traditionally, researchers have used techniques such as g-banding karyotyping and 24-color fluorescence in situ hybridization (FISH) analysis to identify sections of a chromosome that differ between tumor cells and normal cells. However, these techniques are limited in their resolution and cannot pinpoint exactly where in the DNA sequence the changes occur or how they change gene expression.

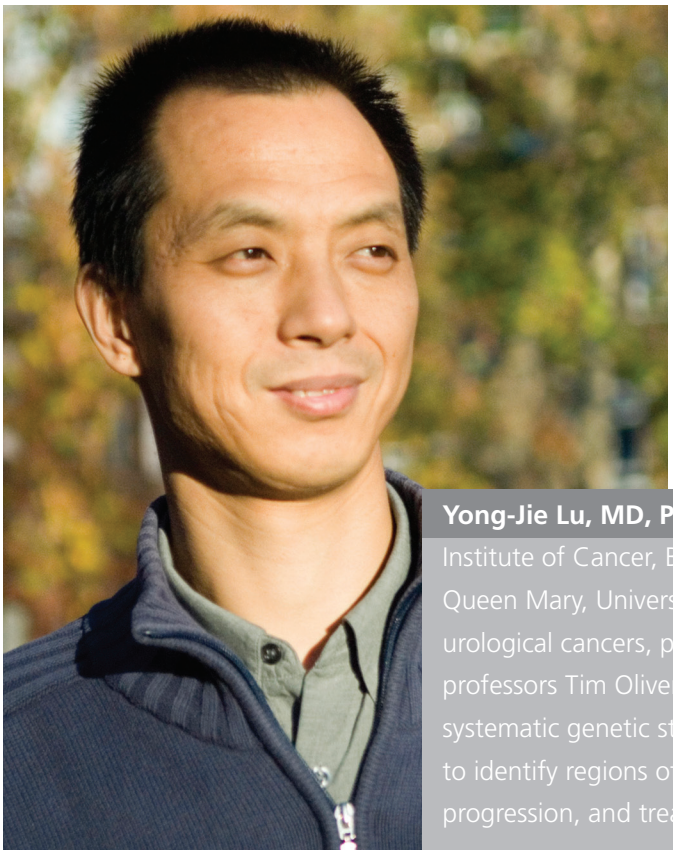
The researchers at Queen Mary, University of London, led by Dr. Yong-Jie Lu, were the first to combine a 24-color FISH analysis, called M-FISH, with high-density microarray analysis using the GeneChip® Mapping 500K Array Set for high-resolution karyotyping.

"We used the 500K SNP arrays in combination with M-FISH for chromosome rearrangement because it gave us a resolution 1,000 times higher than traditional karyotyping and 20 times higher than FISH mapping," said Dr. Lu, a senior lecturer in Medical Oncology at the Institute of Cancer, Barts and the London School of Medicine and Dentistry, Queen Mary, University of London.

Using this technique and in-house analysis software dubbed GOLF (Genome-Oriented Laboratory Filing system) to identify genomic copy number changes and define breakpoints, the team precisely defined the breakpoints of 27 translocations in three prostate cancer cell lines. They also identified 29 internal deletions, changes that can also lead to gene fusions.

The study was published in the July 2007 issue of *Genes, Chromosomes & Cancer.*

Dr. Lu's team followed up on their results using GeneChip Exon 1.0 ST Arrays to monitor gene expression changes at many of the chromosome breakpoints. These arrays include probes for 83 genes located at breakpoints in the three cell lines. Of the 83 genes investigated, 28 showed differential



Yong-Jie Lu, MD, PhD, is a senior lecturer and team leader at the Institute of Cancer, Barts and the London School of Medicine and Dentistry, Queen Mary, University of London. Lu's team is focused on the genetics of male urological cancers, particularly prostate and testicular cancer. Working with professors Tim Oliver and Bryan Young, Lu has created a research program of systematic genetic studies of prostate cancers and testicular germ cell tumors to identify regions of the genome that are significant in the development, progression, and treatment of these tumors.

exon expression in the cancer cells when compared to control samples, indicating that fusion genes and truncated genes may be expressed in these cells.

Dr. Lu recently spoke with Colleen Elso, a postdoctoral fellow at the Walter and Eliza Hall Institute of Medical Research in Melbourne, Australia, about the details of this new technique and the contributions it stands to make to molecular medicine.

The two discussed:

- Improving mapping resolution from 100 kb to 5 kb using Mapping 500K Arrays
- Using SNP signal density information to detect chromosomal gains and losses
- Future plans to use high-density arrays to understand the mechanisms behind DNA rearrangements and identify fusion genes

Resolution reigns

Elso: How much has including the Mapping 500K Array data with your M-FISH data increased the mapping resolution?

Lu: M-FISH makes identification of rearrangements of the chromosomes easier, but it does little to improve the definition of breakpoints. The resolution is still limited to several megabases. FISH using sequence-specific DNA clones is routinely used to map the breakpoints; however, many probes have to be used to really define the breakpoints and the resolution is still very limited by the size of the probes.

The most commonly used FISH probes are BACs, which give a resolution of about 100 kb. Because of their high density, 500K SNP arrays have an average resolution of 5 kb. We used the 500K SNP arrays in combination with M-FISH for chromosome rearrangement because it gave us a resolution 1,000 times higher than traditional karyotyping and 20 times higher than FISH mapping using BACs.

Another benefit is that SNP arrays simultaneously review many chromosome breakpoints in a single sample.

Elso: This kind of study has also been done using array CGH and M-FISH. What is the advantage in combining the SNP data with the M-FISH rather than array CGH data?

Lu: The array CGH study came out just a couple of months before ours. I think SNP arrays have a much better resolution than CGH arrays. CGH arrays use BACs or PACs. There is a limitation to their resolution due to the large size of the BAC or PAC DNA, which is usually about 100 kb.

In a similar study using CGH arrays, a resolution of 78 kb was achieved. Using 500K SNP arrays, we have achieved a resolution of 5 kb, which is much better.

Importantly, it is estimated that in the human genome, there is an average of one SNP every 200 base pairs (bp). I know Affymetrix is working to develop even higher-density SNP arrays. With further technical development, I think they may achieve 200 bp resolutions, which can be directly followed up by sequencing analyses.

A new role for SNP arrays

Elso: SNP arrays are traditionally used for genotyping. How did you use them to detect the DNA copy number changes that are associated with chromosomal rearrangements? What software did you use for the analysis?

Lu: It is true that SNP arrays were originally designed and used for genotyping. However, due to the increased resolution afforded by early SNP arrays, several research groups realized that they would be good tools for DNA copy number analysis.

Colleen M. Elso, PhD, is a postdoctoral fellow in the laboratory of Dr. Tom Brodnicki in the Molecular Medicine Division at the Walter and Eliza Hall Institute of Medical Research in Melbourne, Australia. Her current research focuses on using a mouse model to investigate the genetic and cellular mechanisms underlying the predisposition to type 1 diabetes. Elso was previously a postdoctoral fellow at the Lawrence Livermore National Laboratory in the laboratory of Dr. Lisa Stubbs, where she studied translocations in the mouse genome.

For the copy number changes, they ignore the A and B types and calculate the total value of combined A and B signal densities that reflect the copy number.

Professor Bryan Young, who is the head of our cancer genomics unit, developed a software program called GOLF—Genome-Oriented Laboratory Filing system. Using this software, the user can align all of the SNPs along the chromosomes using signal intensity data. It can also do the genotyping analysis, so it can combine genotyping and copy number changes together. We have published quite a few papers based on the use of this software.

Elso: What type of chromosomal abnormalities were you able to detect by looking at the DNA copy number changes using SNP arrays?

"We used the 500K SNP arrays in combination with M-FISH for chromosome rearrangement because it gave us a resolution 1,000 times higher than traditional karyotyping and 20 times higher than FISH mapping using BACs."

Lu: If you just look at the chromosomal copy number changes by the array method, you can measure chromosomal gain and loss, even in small, opaque chromosomal regions. But we can't say much about how the chromosome structure is changing. That's why we have combined it with M-FISH.

M-FISH shows us roughly what happens to individual chromosomes. Once the microarray data is combined with M-FISH, we can identify the chromosome fragments involved in translocations and deletions, including internal deletions. We can also detect chromosome duplications. The combined technologies allow us to determine whether the amplification is on the same chromosome or if it has moved to another location, like homogeneously staining region (HSR) or double minutes. It can also be used to quickly identify fusion genes resulting from these chromosome rearrangements.

Elso: How does the presence of subpopulations of cells in the sample affect the performance of this type of analysis?

Lu: In cancer research, heterogeneity is a major problem for cell population-based genetic analyses. In clinical samples there is also normal cell contamination. The Affymetrix array system is a very reliable system for detecting copy number changes. We can detect the genetic changes in about 30 percent of the cells. However, although we can detect the genetic change in a small population, the resolution for mapping the breakpoints will be significantly reduced. If there is single copy change in half of the cell population, the resolution of breakpoint mapping will be reduced to about 50 kb. That results in about a 10-SNP region we have to use, since it is difficult to determine exactly between which two SNPs the breakpoint occurs.

Following up with exon arrays

Elso: Your previous data enabled you to use exon arrays in a very focused way rather than using them for a "fishing expedition" type of experiment. Would you comment on this?

Lu: Exon arrays provide a fantastic opportunity to identify truncated genes. However, in these cancer cells, there are many genes that are alternatively spliced, making it complicated to analyze the truncated genes.

Most important, there is currently a lack of effective software for identifying all of the differentially expressed genes among the individual exons. We used exon arrays on several cell lines and two clinical samples. Data has to be extensively analyzed by the bioinformaticist to identify recurrent truncated gene expression. Until efficient, user-friendly analysis software is developed, the genome-to-expression approach is a quicker way to identify important and truncated genes in cancer.

Elso: Do you think the information from exon arrays feeds back into the system and helps map these breakpoints that haven't yet been mapped to a high resolution already?

Lu: Yes. If one gene with differential expression of exons was identified within a region, of course it can be used to further refine the genomic rearrangement. However, as I mentioned, many genes are alternatively spliced in cancer. If several genes with altered expression of exons are present in a region, most of them caused by alternative splicing, they have to be checked individually in order to determine which one is linked to DNA rearrangement. This can be time consuming.

Elso: You identified quite a few small internal chromosomal insertions and deletions in your study that could easily have been missed with standard cytogenetic techniques. Do you think this type of aberration will be found in many other types of cancer by using high-resolution techniques like this?

Lu: Yes. In our leukemia genetics program, led by Professor Bryan Young, we have found many small chromosome gains and losses by 500K SNP array analysis. Interestingly, my own team performed 500K SNP array analysis on some normal karyotyping cancer cell lines, which we confirmed by a few low-resolution techniques. In each of these cell lines, we found microdeletions or duplications of subtle chromosome regions.

So, I believe that internal chromosome insertions, duplications, and deletions commonly exist in many of the human tumors.

We simply could not detect these changes before because the resolution of our technique was too low.

Future directions

Elso: That's really amazing. Do you think this work brings us closer to being able to characterize tumors from individual patients?

Lu: Yes. Combining the high-resolution karyotyping with the gene expression approach reveals many more details of the genetic aberrations in the individual tumors. We are currently using this approach to study prostate cancer clinical samples. However, there are a few technical issues that have to be sorted out to make this clinically applicable.

We need to improve the cell culture system in order to make M-FISH analysis generally applicable to all cancers, particularly solid tumors, from which it is very difficult to get metaphase chromosomes. In many tumors, microdissection or other methods are required to purify tumor cells. We end up with a limited number of cells to work with. If we can develop a technique to further reduce the number of cells and DNA or RNA required for array analysis, this technique will become more useful.

Elso: The mapping of these genomic rearrangements in your prostate cancer cell lines clearly identified a number of genes that could play a role in the development of prostate cancer. What are your future plans for this project?

Lu: Since we used only three cell lines, we would like to do more samples, particularly clinical samples, to confirm that there are some recurrent changes in those genes. Another thing is, of course, we would like to look at fusion genes because they are very important in tumorigenesis. This has been proven in leukemia and the soft tissue sarcomas. And recently, a high frequency of fusion genes has been found in prostate cancer. I believe there are many more fusion genes remaining to be discovered in solid tumors.

So, we are planning to use RT-PCR to clone all of the fusion transcripts. We have actually done some RT-PCR to clone the fusion transcripts already. For some of those, we already know there are fused genes, but we can't get the transcripts at the moment. So, we are also trying a technique that was recently published in *Nature Methods*, called molecular copy number counting, or MCC.

We are using MCC to try to quickly clone the genomic fusion sequence, based on the SNP array high-density mapping. Eventually, we should be able to reveal all of the sequence at the breakpoints. Once we know that, it may help us understand the mechanisms behind why DNA rearrangement happens and how it affects the genes.

Further reading

- Mao X., *et al.* Rapid high-resolution karyotyping with precise identification of chromosome breakpoints. *Genes, Chromosomes and Cancer* **46**(7):675-83 (2007).
- Strefford J. C., *et al.* The use of multicolor fluorescence technologies in the characterization of prostate carcinoma cell lines: a comparison of multiplex fluorescence in situ hybridization and spectral karyotyping data. *Cancer Genetics and Cytogenetics* **124**(2):112-21 (2001).
- Mao X., *et al.* The application of single nucleotide polymorphism microarrays in cancer research. *Current Genomics* **8**(4):219-28 (2007).
- Daser A., *et al.* Interrogation of genomes by molecular copy-number counting (MCC). *Nature Methods* **3**(6):447-53 (2006).

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 SS105 Rev.1