AFFYMETRIX
MICROARRAY
BULLETIN

ADVANCES IN HIGH RESOLUTION STUDIES OF THE GENOME
REGULATION ■ REPLICATION ■ MITOCHONDRIAL SEQUENCE
VOLUME 1 ■ ISSUE 4 ■ $10.00

www.microarraybulletin.com

October 2005

IN THIS ISSUE

Regulation

Harvard's Kevin Struhl and Dana-Farber's Myles Brown discuss ChIP-on-chip as an unbiased method for transcription factor binding site identification *pg1*

Dana-Farber's Shirley Liu and Lawrence Berkeley National Laboratory's David Nix weigh in on improved algorithms for ChIP-on-chip experiments *pg5*

London Research Institute's Frank Uhlmann and University of Geneva's Patrick Descombes discuss what it takes to design a successful ChIP-on-chip experiment *pg9*

Replication

University of Virginia's Anindya Dutta and UC Berkeley's Michael Botchan discuss the use of whole-genome tiling microarrays to study DNA replication at various times during S-phase *pg12*

Sequence Analysis

Johns Hopkins' Joseph Califano discusses the link between mitochondrial mutations and oral cancer with Duke's Amanda Baumann *pg1*

Maastricht University's Bert Smeets and Amanda Parrish discuss mitochondrial resequencing arrays to identify mutations responsible for mitochondrial diseases *pg21*

National Institute of Standards and Technology's John Jakupciak and Genesis Genomics' Ryan Parr discuss array-based resequencing for detecting heteroplasmies in mitochondrial DNA and applications in the clinic *pg23*

University of Connecticut's Linda Strausbaugh and the California Department of Justice's Martin Buoncristiani discuss mitochondrial sequencing arrays in forensics *pg27*

RIKEN's Tadafumi Kato and the Japan National Institute of Neuroscience's Ryota Hashimoto discuss altered expression and sequence variation of mitochondria-related genes in bipolar disorder patients *pg32*

REGULATION

Thousands of Previously Unknown Transcription Factor Binding Sites Mapped Using Tiling Microarrays

Harvard's Kevin Struhl and Dana-Farber's Myles Brown discuss ChIP-on-chip as an unbiased method for transcription factor binding site identification

Scientists at the Dana-Farber Cancer Institute and Harvard Medical School are using chromatin immunoprecipitation (ChIP) combined with tiled microarrays, or ChIP-on-chip, to discover thousands of previously unknown transcription factor binding sites that regulate gene expression. Their findings reveal new information about the basic processes of transcription and potential regulatory targets for breast cancer treatments.


Myles Brown

Myles Brown's research team in the Department of Medical Oncology at the Dana-Farber Cancer Institute is using ChIP-on-chip to define the gene expression pathways controlled by

continued on page 2


Joseph Califano

SEQUENCE ANALYSIS

Mitochondrial Genome Scan Finds Cancer Mutations in Saliva DNA Samples

Johns Hopkins' Joseph Califano discusses the link between mitochondrial mutations and oral cancer with Duke's Amanda Parrish

Cancer researchers at Johns Hopkins Medical Institute have discovered a series of mitochondrial mutations linked to oral squamous cell carcinoma by analyzing the saliva from cancer patients with a new microarray, providing an initial step toward non-invasive early-detection screens.

continued on page 17

steroid hormones and their receptors, including the estrogen receptor in breast cancer and the androgen receptor in prostate cancer.

"When you get data using this technique, it's much more powerful than purely computational methods that can be based on mistaken assumptions," said Brown. "I think this technique tells us a lot more about how proteins find their targets in DNA and shows that the previous models were oversimplifications."

Dr. Kevin Struhl, in the department of Biological Chemistry and Molecular Pharmacology at Harvard Medical School, studies gene regulation in yeast and has recently begun to focus his studies in humans. He believes that ChIP-on-chip methods lead to relatively unbiased results.

"I think a lot of the previous predictions were based on data from quite biased experiments," said Struhl. "When you look at things in an unbiased fashion with a relatively unbiased technique, you find a lot of things that you didn't expect."

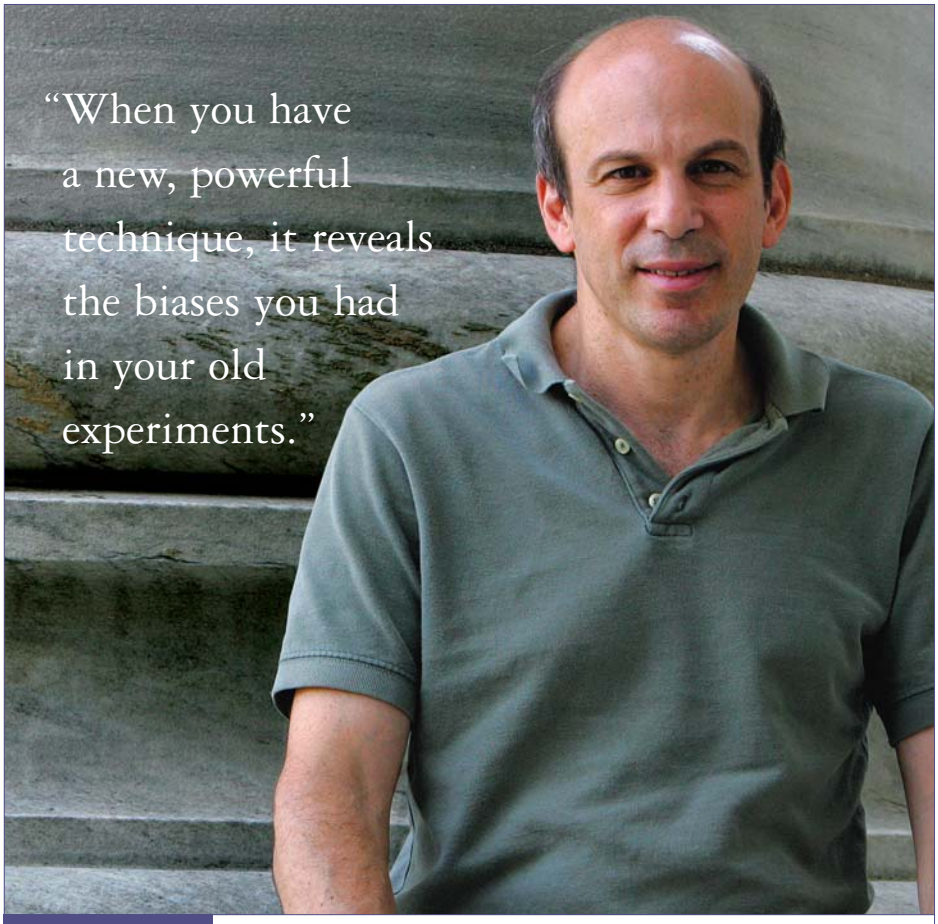
The AMB recently spoke with Brown and Struhl about ways in which ChIP-on-chip technology are changing ideas about binding sites and gene regulation. The two discussed:

- Discrepancies between the predicted number of binding sites and what they find in microarray experiments
- The role of DNA motifs in transcription factor binding
- Limitations and advantages of different technologies for studying binding sites

Surprising Numbers for Real Binding sites

Both Struhl and Brown have found major differences between the actual and predicted number of transcription factor binding sites using ChIP-on-chip technology.

Struhl: Chromatin immunoprecipitation is a powerful technique for locating binding sites, especially when combined with unbiased, tiled microarrays. It allows you to see where the proteins are associating and it's quite open to detecting interactions even if they're just associated indirectly with the target region. The nice thing about a tiled microarray, is that it allows you to ask where the proteins are associating in an unbiased fashion. We found that for some factors, like Sp1 and Myc, there are a large number of sites. We didn't find that many sites for p53. Until you do the experiment you really don't know.



Kevin Struhl is the David Wesley Gaiser Professor in the Department of Biological Chemistry and Molecular Pharmacology at Harvard Medical School. His research focuses on gene regulation in yeast and mammalian cells. His lab has recently been collaborating with Affymetrix to use ChIP-on-chip experiments to investigate human transcriptional regulation.

Brown: What we found for the estrogen receptor is similar to what you found for p53. I think it may suggest that when different transcription factors have their authentic sites mapped by this technique, they will fall into different categories: those that seem to have relatively few targets and those that have a lot more targets.

Struhl: The classical view is that you have a DNA sequence motif and that motif is recognized by a protein, and once that protein is bound there, it does something transcriptionally. The reason that classic view exists is because people have studied genes, during which they mapped protein-binding sites and found motifs. So that's the view from a very biased perspective. Only a few motifs look like the "good sequences." These are sequences that have a good-looking motif and bind *in vitro*. But they don't bind *in vivo*. Only a small percentage of good sequences are bound *in vivo*. The flip side of that is that not every real *in vivo* binding site actually has the motif. So the original dogma is dramatically wrong on both sides.

Binding Requires More than Motifs

Struhl: We've done a little bit of work with *E. coli*. Unlike in eukaryotes, in *E. coli*, if you have a good motif, you are bound virtually 100 percent of the time. Interestingly, even in *E. coli*, it turns out

that about half the sites that we mapped don't have the motif at all. They don't bind *in vitro*, but they are clearly bound *in vivo*. And so even in an organism that you think would be the bioinformaticist's dream, roughly half the sites for the one factor we've studied in detail don't have the motif, but still bind.

There is an overrepresentation of motifs in the real *in vivo* targets. So, it's not like the motif is irrelevant, but it certainly is only part of the story.

Brown: That's what we found as well. In fact, we found that most of our sites had either perfect estrogen receptor binding sites or half sites, suggesting that the motif is a necessary component of the binding we saw, or nearly necessary component, but it's not sufficient information to find where the true binding sites are. And that's probably going to depend on the factor. Some factors are going to care more about their motif than others. For ER, we found that it required not only its own response element, but also the forkhead factor FoxA1 to define a true binding site.

Searching the Whole Genome for Regulatory Elements

Struhl and Brown discuss the types of experiments needed to further elucidate the role of DNA binding and the contribution of microarray analysis to those studies.

Struhl: We are currently part of the ENCODE Consortium, which was formed to identify all of the functional elements of the human genome. The first stage of the Human Genome Project was, of course, sequencing the human genome, but you can now imagine a project where you basically did ChIP-chip on a lot of transcription factors to map the regulatory elements of the human genome. I'm sure both of us believe we could do it right now. But convincing the powers that be that one can do this and getting them to devote serious funding to it, is a different matter. So, this ENCODE Consortium includes a number of labs that hope these studies can be done on a pretty large scale. There aren't all that many transcription factors. There may be a thousand or so. I don't think we could do a thousand. That's probably too expensive. We don't have all the antibodies, but certainly one could do a couple of hundred at this point. And that would be incredibly useful.

Brown: I think the technology is available to do the whole genome, and so we are in the middle of a collaboration with Tom Gingeras at Affymetrix on ER binding across the whole genome. It is the time to do one's transcription factor of choice. I think this, coupled with expression arrays, is going to be the direct output of transcription factor action. If we want to understand regulation, this is what's needed.

Comparing Approaches to Binding Site Studies

ChIP-on-chip experiments generate detailed genome-wide data that was previously impossible to obtain. Struhl and Brown weigh in on the advantages of using microarrays with complementary technologies like siRNA for whole-genome discovery of gene regulation.

Struhl: With polymerase, one of the nice things this array tells you is which mechanism is used more often and which genes use which mechanism. That information would be difficult to get otherwise.

Brown: And that's been a problem in the estrogen receptor field. People have made large conclusions studying a very small number of genes. For any given transcription factor, only a few targets have been studied in any detail. This array allows us to study all the targets in detail.

Struhl: Something that's nice about chromatin IP, is that it can be used in any organism. And what's particularly nice is that it's a quantitative assay for the wild type cell. That's a rare thing to be able to do. One of the problems in the mammalian field is that, unlike in yeast, doing good genetics is hard. In yeast, making a gene knockout is trivial and it's a powerful approach. It is doable in mammalian cells, but it's difficult.

Brown: I think siRNA has allowed that to be done more. Approaches like you're using—doing siRNA followed by an expression profile—is also extremely useful for mammalian cells. It's just surprising how few transcription factors have many known direct target genes.

Struhl: I think most of the other *in vivo* assays that have been used are things like transient transfections, and those are prone to artifacts for two reasons. First, the template you're putting in doesn't look like a normal template—it's not real chromatin. And second, when people are looking at the effect of their favorite factor, they will toss in a plasmid that expresses their favorite factor, and they're generally grossly over-expressing it. So you see what your protein can do if you grossly over-express it, but not what it does in the real world.

Brown: I think transient transfection is an important technology, but to a large

Myles Brown is chief of the Division of Molecular and Cellular Oncology at the Dana-Farber Cancer Institute. His research focuses on the roles of estrogen receptor in breast cancer and androgen receptor in prostate cancer. Researchers in his laboratory have been using ChIP-on-chip technology to study the details of gene regulation in cancer cells.



degree, chromatin immunoprecipitation and ChIP-chip certainly gives superior answers in terms of actual targets and the actual proteins involved in regulation. Chromatin IP has allowed people to look at endogenous proteins at their normal levels, and people have even looked at them from tissues, so you really can get a true look at what's happening under normal physiology.

Limitations of Using Antibodies

While ChIP-on-chip studies have provided high-resolution views into the genome, a major limitation is the quality of antibodies available for these studies.

Brown: Unfortunately, having good antibodies is often the limiting factor in IP experiments. But I think a lot of companies and a lot of people are developing good antibodies to mammalian transcription factors. You can use tagged proteins, and we have done that for chromatin IP. In mammalian biology in general, it's still surprising how often having good antibodies to your protein is a rate limiting step.

Struhl: Yes. I think there's another issue with antibodies, which is a difficult one to assess. You assume an antibody is specific for your protein, but it's difficult to know that. You can certainly show that it pulls your protein down nicely and that it works for a ChIP, but it's difficult to know whether you're pulling something else down that you don't know about.

Brown: We've been concerned about that issue too. Early on, we used multiple different antibodies for the estrogen receptor. Another limitation of ChIP is the possibility that certain epitopes might be masked under certain conditions, because these proteins are in large complexes. You could imagine that an antibody that can recognize estrogen receptor might not work well in ChIP because the epitope is masked.

Advantages of Using Microarrays

Struhl: I think reviewers often mention these theoretical possibilities to give you trouble. The technique is not perfect. It does have its theoretical issues about what you're really measuring. But by and large, I would say that compared to most techniques, it's quite robust.

Brown: I agree completely. For us it's revolutionized what we can look at.

> "People have made large conclusions studying a very small number of genes. For any given transcription factor, only a few targets have been studied in any detail. This array allows us to study all the targets in detail."

Struhl: We're probably among the first people to do ChIP-chip for transcriptional regulation. And in our yeast work, every time we did a new experiment, just using a ChIP on a standard protein, we found new things. Either we answered a major question or we found something completely unexpected. That happened frequently. When you have a new, powerful technique, it reveals the biases you had in your old experiments.

Brown: People have made good use of the promoter arrays, but the unbiased whole-genome tiled arrays eliminate the bias you have when you look only at particular regions of the genome. And what we found, for example, is that ER is mostly not at the promoters. So promoter arrays would have missed most of the sites of estrogen receptor binding and functional regulation.

Struhl: And for us, we found a similar thing in a slightly different way. Especially for things like Sp1 and Myc, which are often thought of as fairly proximal factors, we found that only about 20 percent of all the sites were in classical promoters. Now a lot of that is turning out to be because there is a lot more transcription than people thought. And a lot of these sites are in places that drive these so-called unusual transcripts. There's no question that if you just use a biased array, you're going to confirm your own biases. You're going to miss a lot of important stuff.

High Background in ChIP-on-chip Experiments

Scientists have consistently been faced with high background problems in ChIP-on-chip experiments. Figuring out ways to lower that background and increase sensitivity are critical to discovering regulatory sites that are currently dismissed as noise.

Brown: We were lucky to have wonderful collaborators who helped us enormously with these studies. We have an extremely good Microarray Core at the Dana-Farber headed by Ed Fox that provided us with very high quality data. Pam Silver's group in our Department of Cancer Biology has been a pioneer in the use of ChIP-chip in both yeast and mammalian cells and were instrumental in getting us started in this area. In addition, Shirley Liu's group in the biostatistics department at Dana-Farber has done a lot of work on developing methods for data analysis for ChIP-chip experiments. Using her methods, we came up with a list of binding sites. Then we did some experiments, because I'm an experimentalist. I like having a way of validating the list, so we got an ordered list of sites and then we determined where we thought the boundary between real and background might be. We did directed ChIP experiments to define where the cut-off should be. But, ultimately in these types of experiments, you're always going to have some false discovery rate. It's inevitable when you go to these genome-wide analyses.

Struhl: I think this issue is important, but it makes you too conservative. The real problem is in the control sample. You're trying to get correct hybridization, so you must be able to pick it up at one part in 10^8, and that's asking a lot for a biochemical experiment. That's what causes a lot of the background. As a result, you miss more things. In other words, when you see something that's significant, it's usually pretty significant. So a lot of times we'll see something that passes a cut-off, at least by a computational method. Then you do an experiment, you find its 30-fold enriched. It is a non-trivial issue. It's relatively easy to pick out the best sites. If you want to go down on the list, then it gets a little more tricky. And I think the most serious issue occurs when you want to compare different samples and you want to do more quantitative stuff. So, if you have a situation where you're looking at condition X and condition Y and want to know whether something is down 3-fold here or there, that's where it gets a little more difficult.

FOR MORE INFORMATION

Contact Information

■ **Kevin Struhl, Ph.D.**
Department of Biological Chemistry and Molecular Pharmacology
Harvard Medical School
Building C1, Room 315
240 Longwood Avenue
Boston, MA 02115
Kevin@hms.harvard.edu

■ **Myles A. Brown, M.D.**
Department of Medical Oncology
Dana-Farber Cancer Institute
44 Binney Street
Dana 730
Boston, MA 02115
myles_brown@dfci.harvard.edu

Companies

■ **Affymetrix Inc.** –
http://www.affymetrix.com

Organizations

■ **Harvard Medical School**
http://hms.harvard.edu

■ **Dana-Farber Cancer Institute**
http://dana-farber.org

■ **The ENCODE Project**
http://www.genome.gov/ENCODE

■ **The Human Genome Project**
http://www.ornl.gov/sci/techresources/Human_Genome

Further Reading

■ Cawley, S. *et al*. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of non-coding RNAs. *Cell* 2004; 116(4): 499-509.

■ Geisberg JV, Struhl K. Quantitative sequential chromatin immunoprecipitation, a method for analyzing co-occupancy of proteins at genomic regions in vivo. *Nucleic Acids Res.* 2004; 32(19):e151.

■ Grainger DC, Overton TW, Reppas N, Wade JT, Tamai E, Hobman JL, Constantinidou C, Struhl K, Church G, Busby SJ. Genomic studies with Escherichia coli MelR protein: applications of chromatin immunoprecipitation and microarrays. *J Bacteriol.* 2004; 186(20):6938-43.

■ Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, Eeckhoute J, Shao W, Hestermann EV, Geistlinger TR, Fox EA, Silver PA, Brown M. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* 2005; 15;122(1):33-43.

People

■ **Xiaole Shirley Liu**
xsliu@jimmy.harvard.edu

■ **Tom Gingeras**
Tom_gingeras@affymetrix.com

■ **Ed Fox**
http://chip.dfci.harvard.edu/lab/services.php

■ **Pam Silver**
http://research.dfci.harvard.edu/silverlab/

REGULATION

New Bioinformatics Algorithm Predicts Forkhead Protein Plays a Role in Breast Cancer Estrogen Response

Dana-Farber's Shirley Liu and Lawrence Berkeley National Laboratory's David Nix weigh in on improved algorithms for ChIP-on-chip experiments

Dr. Shirley Liu, assistant professor of biostatistics at the Dana-Farber Cancer Institute, has developed a new algorithm for analyzing ChIP-on-chip data which enables the unbiased mapping of estrogen receptor (ER) binding sites along chromosomes 21 and 22. Sequence analysis shows that a forkhead protein works together with estrogen receptor to mediate steroid response in breast cancer cells.

Dr. David Nix, a computational scientist working on the Berkeley *Drosophila* Transcription Network Project at Lawrence Berkeley National Laboratory has also been using ChIP-on-chip technology to look at binding sites for transcription factors that are expressed in early *Drosophila* development.

"Eventually, our goal is to be able to predict the expression of genes based on sequence," said Nix. "To do that, we need to understand the rules that govern cis-regulatory modules. Based on sequence so far, we have found it most difficult to make meaningful predictions in many areas. Our hope is that we can take an unbiased approach by using ChIP-chip technology to look at where things are actually binding in the embryo."

Liu recently spoke to Nix about the complementary roles of ChIP-on-chip experiments and bioinformatic algorithms. The two discussed:

- Methodologies used by Liu and collaborators
- Developing standards for processing and interpreting tiling microarray data
- Commercial and open source software available for motif finding

Analysis of ER Chromatin IP on Tiled Microarrays

Nix: Your algorithm tried to estimate probe behavior by obtaining an empirical measurement of noise. How well did the algorithm work?

Liu: It worked pretty well at least for our analysis of estrogen receptor ChIP-chip, and re-analysis of the p53 data from Cawley's *Cell* paper, out of Affymetrix. There were 48 sets of microarray data from Cawley's paper alone. Dana-Farber generated another 20 datasets on chromosomes 21 and 22. One thing we tried to improve on the Cawley analysis, is to consider all probe pairs, instead of only those with higher perfect match (PM) than mismatch (MM). Sometimes the mismatch might hybridize to other sequences in the genome, so it could have a much higher value. Half of the probes on the array have higher mismatch than perfect match values, so we could loose a lot of information if we don't consider them.

Nix: We've tried to use mismatch probes in our analysis, but find that they introduce a lot of noise into the data. In cases where you don't have a good control, like in some of the transcriptome experiments where you are putting mRNA/cDNA onto a chip, you don't have a good control that can be used to normalize for differential probe response. With ChIP-chip data, we can use input chromatin. By taking a simple ratio of treatment to input control chromatin, we can avoid the use of mismatch oligos for the ChIP-chip data.

Liu: We used PM-MM for the analysis just published, but you can get pretty good results from PM alone. Once we identified the ChIP-enriched regions, we tried to look for the transcription factor binding motifs in these sequences. These regions are theoretically bound by transcription factors *in vivo*, so every sequence that's pulled down should have a binding site. We performed *de novo* motif finding, without any knowledge of what the motif might look like or checking any motif databases, to find enriched sequence patterns of widths 5 to 17 bp in the ChIP-enriched regions.

From the ER ChIP-regions, we found a very strong palindrome motif of width 15 which turned out to be the canonical ER binding motif. When we masked out this motif to look again, another motif came out very strong around 9 to 10 bases. From my experience in analyzing yeast ChIP-chip experiments, I told our collaborator that there seemed to be a forkhead protein working with the ER to bind these regions.

The biologists checked a cancer microarray database called Oncomine, and found FoxA1— a member of the forkhead protein family—to be highly co-expressed with ER. They used an antibody against FoxA1 to pull down ChIP DNA and did site-directed quantitative PCR to check for the ER ChIP regions that contain the FoxA1 binding site. This procedure verified that those sites are indeed bound by FoxA1. When they used siRNA to knock out the FoxA1 protein, and conducted ChIP-chip against ER, the ER binding site was lost. So our recent *Cell* paper not only located the ER-bound region on chromosomes 21 and 22, but found a cofactor FoxA1 which is required by ER to bind to these regions.



Shirley Liu is an assistant professor in the Department of Biostatistics at the Dana-Farber Cancer Institute. Dr. Liu's research is focused on using computational methods to predict sites of transcription factor binding. Recently, Dr. Liu has been collaborating with Dr. Myles Brown's laboratory, using their microarray data to help make accurate predictions of transcription factor binding sites for the estrogen receptor.

Analysis of p53 Chromatin IP on Tiled Microarrays

Liu: We used a similar method to reanalyze the p53 ChIP-chip data in Cawley's *Cell* paper. We were able to identify many p53 bound regions, including all 10 previous qPCR-validated regions. In addition, we found a motif from the predicted p53-bound regions similar to the canonical p53 binding motif. This motif and the canonical p53 binding motif are both highly enriched in the predicted p53 ChIP regions compared to the genome background. They are also more enriched in our predicted p53 ChIP regions than in Cawley's predicted p53 ChIP regions. We thought this was pretty reassuring.

We found that p53 binding may not be conserved between humans and rodents. A lot of the regions we identified including some of the qPCR-validated regions do not align to the rodent genome at all. We talked to some colleagues, and they were not surprised. Rodents may have quite different p53 regulation from human.

Nix: There is not a whole lot of conservation in the bound regions we've been looking at either. They are not conserved and that's actually fitting with what people are reporting elsewhere. Several groups have looked at the eve locus in several *Drosophila* species, and really teased apart the enhancers. There's not a heck a lot of

conservation at the sequence level, but functionally, they still work. It suggests that there are very different mechanisms by which these things are evolving.

A Question of GC-rich Content

Nix: We've noticed disturbingly in the last few weeks that our identified regions are enriched for GC content. And this is even found in the mock IP/IgG controls. The problem is that if you have a non-specific enrichment for GC right under the binding peaks, it's really going to skew your ability to detect novel "words." Have you seen this? And if you have, have you tried to control for it in any way?

Liu: GC-rich probes often give higher probe intensity, which might cause them to look "enriched." But these probes should also give high intensity in genomic input or mock IP controls, so the algorithm would know they are not really enriched.

Nix: Well, it does. For some reason, we are pulling down regions that are GC-rich. This may very well be an experimental artifact. We are using formaldehyde as a crosslinker and that is known to have a GC bias. Maybe these regions amplify better by PCR. They certainly hybridize better. We can get rid of most of the differences in hybridization by doing a ratio of treatment versus control. But it still doesn't get around the fact that we are seeing those increases.

Liu: GC-rich sequences and repeats often interfere with downstream motif finding. After retrieving the ChIP sequences, we ran three rounds of repeat masking. The first is RepeatMasker from the University of Washington, which most people use. The second is the tandem repeat masker, for which we downloaded the repeat coordinates from UCSC. The third is our own simple repeat masker which gets rid of single base (e.g. AAAAAAA) or double bases (e.g. ACACACACACA) repeats. With these repeats removed, we haven't had a problem with repeats or GC-rich sequences, at least in the ER and p53 data.

Creating Standards to Compare Methods

Nix: I think we need to have some standards by which we can compare different data analysis methods. We are finding that we have to go back to ground zero to try to figure out how to process the data. We do stuff very differently than what folks have done before.

We recently produced some spike-in data from an experiment where we took BACs containing large 200 kb inserts and mixed these at varying concentrations with input chromatin. These were then hybridized to the tiling microarray. Then we measured the output of various tests to see how well they differentiated between calling a true positive and a true negative.

We think that by using such spike-in experiments, we can also get a handle on some of the pre-processing steps that might help with the data analysis.

We are hoping that the Hidden Markov Model that you folks have developed will be a step up. Because right now, we are not using a traditional statistical test. We are basically looking at simple ratios. And that gets into a difficult issue of how to establish a false positive rate. We create a list where we have all these regions of the genome that we think are bound and we rank them by the highest intensity ratio. Transcription factor binding sites tend to cluster at the top, but we don't know how far down the list we can go.

Liu: The only concern I have is that spike-in DNA may not represent real ChIP DNA. In the real experiment, we are dealing with the whole genome DNA population and sonication is random.



David Nix is a computational scientist in the Department of Genome Sciences at Lawrence Berkeley National Laboratory. He is part of a large multidisciplinary group called the Berkeley *Drosophila* Transcription Network Project **(http://bdtnp.lbl.gov)** that is currently working with scientists at Affymetrix to use ChIP-on-chip technology to identify binding sites for transcription factors involved in early *Drosophila* development.

Nix: I should have mentioned that we did everything the same with the BACs as with an IP. They were sonicated, random primed, and amplified. There's going to be bias for fragment sizes, amplification, etc.

I think such BAC spike-in experiments are a good first step towards developing standardized datasets for use in evaluating different tests and data processing methods, but clearly they are not the end.

Liu: Eric Lander's group published a ChIP-chip paper using tiled arrays (Bernstein *et al.*) this spring, in which they did qPCR to validate their ChIP-enrichment predictions. Many regions below the prediction cutoff still have pretty high qPCR fold enrichment. So current methods might still give many false negatives, but we will see many improvements on correct fold change prediction in the future. The most important part now is probably to estimate probe behavior by considering probe sequence.

Open-source Bioinformatics Software

Nix: Is your Hidden Markov Model (HMM) software available?

Liu: We have a package already which people could download. At the same time, Stratagene, a designated third party software developer for Affymetrix chips, is working to incorporate our analysis algorithm into their software. The light version of the Stratagene software is free, but they charge some money for the advanced version. A third effort is to work with John Quackenbush, a professor DFCI recently recruited from TIGR, to incorporate our algorithm into TM4, an open source Java software package. Compared to the R open-source package BioConductor, TM4 has a more graphical interface and does not require any coding from users. It's more for biologists to use but informatics people can look at the source code and modify it to do whatever they want.

Nix: We also have developed a large Java open-source package. It's called TiMAT for Tiling Microarray Analysis Tools. The whole package is available as open-source. I built it to process our *Drosophila* tiling data but haven't had an opportunity to test it with other datasets. It would be great to add an HMM option into TiMAT.

Liu: We would love to incorporate our software into your system. The only concern right now is we may not have enough experiments to estimate the probe behavior for the algorithm to work in *Drosophila* or the whole human genome. Instead of using other experiments, since the whole human genome is 44 million probes, we hope to estimate probe behavior by looking at other probes in the same experiment that have similar probe sequence or hybridization quality. This way, we can just do the estimate from a single experiment instead of using other experimental data that may or may not be there. Once this is ready, it'll be easier for us to integrate our software into TiMAT.

Nix: We can also provide you with a lot of *melanogaster* chip information. We've produced a data set from 30 to 40 individual chips from several different ChIP-chip experiments. Data from another 30 odd chips are also available from a set of experiments performed in collaboration with Tom Gingeras' group. The whole genome of *Drosophila* is covered on a single chip featuring about 3 million oligos. It's nice because it covers the entire spectrum of where transcription factors bind on the genome level.

FOR MORE INFORMATION

Contact Information

■Xiaole Shirley Liu, Ph.D.
Department of Biostatistical Science
Harvard School of Public Health /
Dana Farber Cancer Institute
44 Binney Street, Mayer 1B22,
Boston, MA 02115
xsliu@jimmy.harvard.edu

■David Nix, Ph.D.
Lawrence Berkeley National Laboratory
Department of Genome Sciences
1 Cyclotron Road, Mailstop 84-171
Berkeley, CA 94720
DANix@lbl.gov

Companies

■Affymetrix Inc. – http://www.affymetrix.com

■Stratagene http://www.stratagene.com

Organizations

■Dana Farber Cancer Institute – http://dana-farber.org

■Lawrence Berkeley National Laboratory – http://www.lbl.gov/

■Berkley *Drosophila* Transcription Network Project – http://bdtnp.lbl.gov/

■UCSC Genome Bioinformatics – http://genome.uscs.edu

Tools

■Oncomine Cancer Microarray Database – http://www.oncomine.org

■Bioconductor http://www.bioconductor.org

■TM4 http://www.tm4.org

■TiMAT http://bdtnp.lbl.gov/TiMAT

■RepeatMasker http://www.repeatmasker.org

Further Reading

■Cawley, S. *et al*. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of non-coding RNAs. *Cell* 2004; 116(4): 499-509.

■Hong P, Liu XS, Zhou Q, Lu X, Liu JS, Wong WH. A boosting approach for motif modeling using ChIP-chip data. *Bioinformatics*. 2005; 21(11):2636-43.

■Li W, Meyer CA, Liu XS. A hidden Markov model for analyzing ChIP-chip experiments on genome tiling arrays and its application to p53 binding sequences. *Bioinformatics* 2005; 21 Suppl 1:i274-i282.

■Nix, D.A, Li, X.Y, Moses, A.M., Eisen, M.B., Biggin, M.D. Development and optimization of tiling microarray analysis tools for chromatin immunoprecipitation. Manuscript in preparation. 2005

■Bernstein B. *et al*. Genomic Maps and Comparative Analysis of Histone Modification in Human and Mouse. *Cell* 2005; 120(2): 169-181.

People

■Myles A. Brown http://www.dana-farber.org/abo/danafarber/detail.asp?PersonID=22&RD=True

■Eric Lander http://web.mit.edu/biology/www/facultyareas/facresearch/lander.shtml

■John Quackenbush http://www.hsph.harvard.edu/faculty/JohnQuackenbush.html

REGULATION

ChIP-on-chip Experiments Reveal Unique Mechanism Underlying Yeast Chromosome Segregation

London Research Institute's Frank Uhlmann and University of Geneva's Patrick Descombes discuss what it takes to design a successful ChIP-on-chip experiment

Led by Dr. Frank Uhlmann, researchers using high-density tiling microarrays made the surprising discovery that transcriptional activity of the genome determines where sister chromosomes attach during cell division. They found that cohesin—the glue that holds chromosomes together—binds DNA where transcription occurring in opposite directions converges. The faithful segregation of chromosomes during cell division is necessary to keep cells from becoming cancerous, and Uhlmann's work provides one of the first high-resolution views of how that happens.

"Using this higher-than-ever possible resolution, we saw very striking localization to converging transcriptional terminator regions," said Uhlmann. "We didn't understand it at first sight, but it immediately told us there's something unique going on, and we have to follow up on this."

Dr. Patrick Descombes, is a core lab director at the Genomics Platform of the NCCR Frontiers in Genetics at the University of Geneva. Descombes provides counseling and access to a variety of genomic technologies to a large number of investigators and is actively engaged in developing new methods and improved protocols.

Uhlmann and Descombes recently spoke with AMB Editor Rachel Shreter about the methodologies that form the basis of successful ChIP-on-chip experiments. Topics covered include:

- Production and assessment of antibody quality for co-immunoprecipitation
- Developing proper negative controls and optimal experimental design
- DNA amplification methods
- Background normalization and loci identification

Production and Assessment of Antibody Quality for Chromatin Immunoprecipitation

Shreter: What considerations related to antibody affinity for the target protein must be taken into account for ChIP-on-chip experiments?

Uhlmann: For ChIP-on-chip experiments, it's a major issue to be able to efficiently and cleanly immunoprecipitate the protein under investigation. Unlike working directly with RNA and DNA, which independent of sequence have more or less similar properties, every protein is really different. Therefore, the availability of good antibodies for each protein is absolutely key to the success of the experiment.



Frank Uhlmann is a Group Leader of the Chromosome Segregation Laboratory at the Cancer Research UK. He received his Ph.D. from Memorial Sloan-Kettering Cancer Center, New York and the University of Tübingen, Germany. Uhlmann later performed his postdoctoral training at the Research Institute of Molecular Pathology in Vienna. He now heads the Imperial Cancer Research Fund in London. Uhlmann's main research interest lies in understanding the mechanisms that cells employ to ensure the faithful inheritance of their genomes, and he has authored nearly 50 publications studying the molecular pathways governing genome duplication and inheritance.







Patrick Descombes is the manager of the Genomics Platform of the NCCR Frontiers in Genetics at the University of Geneva. He received his Ph.D. in Molecular Biology in 1993, in the laboratory of Professor Schibler at the University of Geneva. His postdoctoral work focused on the regulation of DNA replication and the cell cycle in the laboratories of Professor Tim Hunt at the Cancer Research UK and Professor Erich Nigg at the University of Geneva.

In budding yeast, we fuse our protein of interest to epitopes that have been well characterized and have a very high affinity for available monoclonal antibodies. There are very high-affinity antibodies available for the HA or PK epitopes, and also, the Myc epitope. Some of the commercial anti-GFP antibodies are quite successful in precipitating the protein as well. This just has to be optimized depending on the protein.

A good polyclonal antibody is still probably better than a good monoclonal antibody—just for the reproducibility and supply. The epitope fusion to the protein of interest is, in itself, a concern. One has to validate that the active fusions are fully functional. But if they are, then this can be useful for successful immunoprecipitations.

Descombes: Yeast is fantastically favorable in this respect because you have the ability to use recombination in order to add a tag. Usually the tagged protein is still expressed at normal levels and remains functional. This is not always easy to achieve in higher eukaryotes, so the need for an antibody specific to the individual protein becomes more of an issue. In this case, additional work would be needed to clearly characterize the immunoprecipitated material to make sure that the immunoprecipitate is specifically enriched for your protein of interest and nothing else.

Descombes: Before going on to work on amplification and processing of the chip, I would check if the immunoprecipitated material contains the DNA that is expected to be present. And this, of course, is only possible if you know a potential target. If this is the case, I strongly advise using a technology like real-time PCR to amplify the potential target. Until this is confirmed, I wouldn't proceed to chip hybridization.

Developing Proper Negative Controls and Optimal Experimental Design

Shreter: Frank, you included several controls in the design of your recent experiments. These included duplication of each experiment, confirmation using different tags, different subunits of the same complex, and different yeast strain backgrounds. How do you decide on what controls to use?

Uhlmann: You never really know whether your controls are sufficient, but actually every one of the controls that we conducted brought up most of the 28 cohesin binding sites that we found. The good thing was that the controls confirmed each other. Once we saw the same pattern coming up in all instances, we had confidence that we were looking at the real sites of association of the cohesin complex.

Descombes: I think that specifically for ChIP-on-chip, you want to make sure that the signal you see coming from the array really means that the protein of interest is sitting on that piece of DNA. One obvious first thing that comes to mind is that it is difficult, if not impossible, to define whether the protein is directly binding to the DNA or whether it's an indirect interaction caused by the cross-linking of proteins with other proteins. So, if it's a transcription factor, which has a DNA binding motif, it becomes obvious that it is binding onto the DNA directly.

But, for example, many people are currently studying the interactions of the nuclear pore complex with chromatin. Despite the fact that some of the components of the complex do not directly interact with DNA, they may still give a positive signal. Complementary work may thus be required to confirm the ChIP data.

Uhlmann: Controls should not be designed to confirm your result, but to dispute it. If you can't dispute it, and it's supported by independent lines of research, a hypothesis can be derived.

Descombes: The way that you design a tiling array experiment is different from a classical expression study. The choice of the control, as we just discussed, and the preparation of the reagent is a key issue. And I would say that the major problem is that there is no way to standardize these experiments. With expression studies, once you have a good solid method to isolate your samples, you apply it to every case, irrespective of the species or whatever. But specifically for ChIP-on-chip, this is not possible. Therefore, the issue of a good reference is also important in the design.

I think one way to address this is exactly what Frank has been doing, which is to compare the data that you obtain with your antibody against hybridization of whole genomic DNA. You then calculate a signal log ratio. If there is specific enrichment of DNA in the ChIP, the ratio is

positive, whereas in the absence of enrichment, the ratio is negative.

DNA Amplification Methods

Uhlmann: Because we immunoprecipitate a smaller quantity of DNA than is needed for hybridization on the microarray, it has to be amplified and labeled. That this amplification procedure is completely unbiased and is a true reflection of the immunoprecipitated material is not exactly straightforward. These steps have to be carefully evaluated and possible alternative methods should be considered. That's especially true for larger genomes that show more repetitive features and where the currently standard PCR amplification techniques may run into some problems.

Descombes: Yes. Actually, what Frank has been using is a PCR-based method, which works very well for yeast, but for higher eukaryotes it's not as easy and there's lots of noise. One alternative, which seems to be quite successful, is to use a T7 linear amplification procedure, which basically requires that you add the T7 promoter at the end of the DNA fragment that you immunoprecipitate. Then you use the tagged DNA as a template for an individual transcription reaction that will generate cRNA, a little bit like expression arrays. And then you either directly hybridize this cRNA, or convert it into cDNA again and hybridize.

Background Normalization and Loci Identification

Shreter: What methods have you used for background normalization and loci identification?

Descombes: The data analysis method that Frank's collaborator, Katsuhiko Shirahige used, is the same methodology that Affymetrix is proposing to normalize and treat the data from expression arrays. I think the key point here will be to challenge the data obtained with different methods, and to see what the results are. An additional point is what we call the band width. If you have a tiling array, you have an oligonucleotide spaced every 35 base pairs. How many neighboring oligos should be turned on for you to believe that it's a real signal versus noise? If you're immunoprecipitating material of an average size of 500 base pairs, you expect at least eight or ten oligos to be turned on for it to be a real signal. So if you have only two or three oligos neighboring and then a gap, it's likely to be noise. But this, again, applies to immunoprecipitation with fragments of roughly 500 base pairs.

This is different from mRNA analysis—for mRNA with 20 bases, you have to count on a single oligonucleotide. There is no rule for loci identification at the moment. What we can hope is that this technology will evolve like expression arrays have evolved with such an enthusiasm that many people, including companies like Affymetrix, will try to provide analysis tools that suit these applications.

Uhlmann: The data on the microarray gives you a readout for each individual oligonucleotide. For each probe set, there are 16 oligonucleotides, each with 25 bases. The first straightforward or intuitive criteria is that for a 300 base pair region, if a probe set is a positive score, then all 16 oligonucleotides that are in that probe set should agree that it's a positive score. And there are, of course, always exceptions—particularly a nucleotide might or might not hybridize in an experiment for various, really technical issues. An important criterion for significance of binding is p value that takes into account all 16 oligonucleotides. And the other criterion is then, as Patrick also says, if the resolution of the experiment is on the order of 500 nucleotides. If there is protein association at a particular place in the chromosome, and we have 300 nucleotide probe sets, then you expect that every association site is at least reflected in several neighboring probe sets. So if just one probe set stands out, in a background of all negative probe sets, that's also unlikely to be a real signal. So that was an additional criterion that was important for us.

And I can only agree with Patrick that in the future when the data set is very large and there are huge amounts of information, I'm sure that our statistician and bioinformatics colleagues that we really depend on for the analysis of this data, will have other good ideas for us to help evaluate the signal.

Descombes: I agree. The study of protein-DNA interactions has been going on for a long time. This fantastic new dimension, which is brought about by the tiling arrays is that you can go genome-wide. The need to map, genome-wide, these types of interactions has been in the air for quite some time. Now we have the tool to do it, and, of course, this enables us to think of new experiments and new ideas about biology.

FOR MORE INFORMATION

Contacts

■Frank Uhlmann, Ph.D.
Chromosome Segregation Laboratory
London Research Institute
Lincoln's Inn Fields Laboratories
44 Lincoln's Inn Fields
London WC2A 3PX
frank.uhlmann@cancer.org.uk

■Patrick Descombes, Ph.D.
NCCR Frontiers in Genetics
University of Geneva CMU
1 rue Michel Servet
1211 Genève 4
Suisse
patrick.descombes@frontiers-in-genetics.org

Companies

■Affymetrix Inc.
http://www.affymetrix.com

Organizations

■London Research Institute – http://www.london-research-institute.co.uk/

■University of Geneva – http://www.unige.ch/

■Frontiers in Genetics – http://www.frontiers-in-genetics.org/

Further Reading

■Lengronne A, Katou Y, Mori S, Yokobayashi S, Kelly GP, Itoh T, Watanabe Y, Shirahige K and Uhlmann F. Cohesin relocation from sites of chromosomal loading to places of convergent transcription. *Nature* 2004; 430(6999): 573-578.

■Higuchi T and Uhlmann F. Stabilization of microtubule dynamics at anaphase onset promotes chromosome segregation. *Nature* 2005; 433(7022):171-176.

■Gauthier BR, Brun T, Sarret EJ, Ishihara H, Schaad O, Descombes P, Wollheim CB. Oligonucleotide microarray analysis reveals PDX1 as an essential regulator of mitochondrial metabolism in rat islets. *J Biol Chem*. 2004;279(30):31121-30.

REPLICATION

Highest Resolution Study of Genome Replication Surprises Scientists with Unexpected Timing in S phase

University of Virginia's Anindya Dutta and UC Berkeley's Michael Botchan discuss the use of whole-genome tiling microarrays to study DNA replication at various times during S phase

Scientists led by Dr. Anindya Dutta of University of Virginia (UVA) discovered that DNA replication initiates during all quarters of S phase in some chromosomal segments in cancer cells, providing critical information about how the cancer genome replicates and how it might be restored to healthy growth. The team made their findings through the most detailed whole-genome scan of origins of replication ever performed using high-resolution tiling microarrays that measured every 35th base of chromosomes 21 and 22.

Dutta recently published this work in the May 2005 issue of the *Proceedings of the National Academy of Sciences, USA.*

"The pan-S phase pattern was a complete surprise," said Dutta, professor of biochemistry and molecular genetics at UVA. "I wanted to ignore it, but it sort of jumped out at us. The result is hard to believe given the general consensus that everything replicates in a temporally specific manner."

Dr. Michael Botchan, a professor of biochemistry and molecular biology at the University of California at Berkeley, is using a similar approach to study DNA replication in *Drosophila*. Botchan is using a microarray tiled with the *Drosophila* genome to explore how the mechanisms and regulation of DNA replication are coupled to the cell cycle.

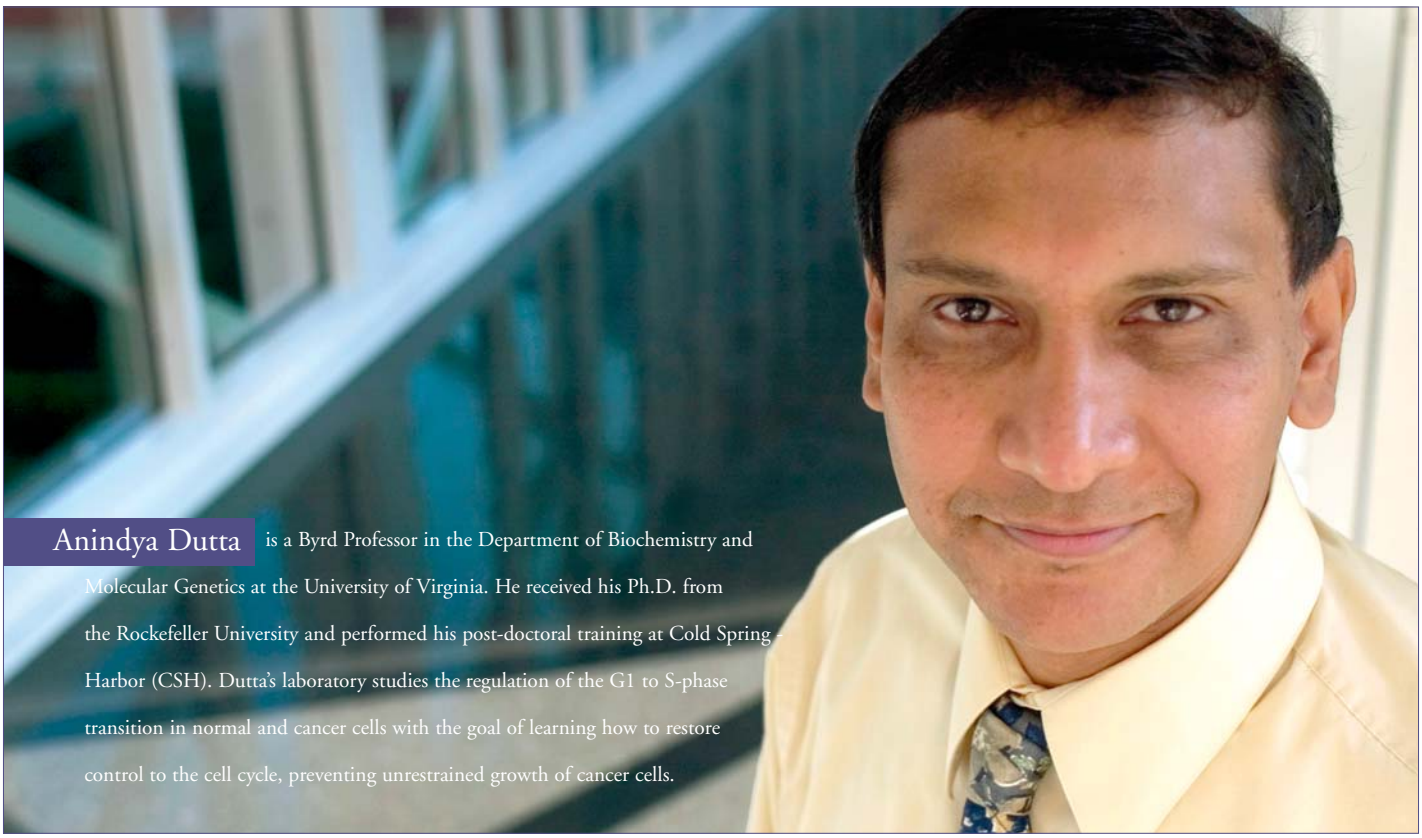
Botchan recently spoke with Dutta about the challenges they both face in exploring the genome for origins of replication using these new tiling microarrays. Topics covered:

- The discovery of pan-S phase replication in HeLa cells and ways to validate those tiling microarray results
- The role of chromatin structure and associated proteins regulating the timing of replication in early or late S-phase
- The methodologies used to validate and create standards for ChIP-on-chip data

Discovery and Data Validation of pan S-phase Replication

Botchan: I'd like to know more about the surprising observation that you made with Tom Gingeras' group at Affymetrix, that in HeLa cells, at least 60 percent of chromosome 21 and 22 is replicated pan-S phase—it might replicate early, it might replicate late, it might replicate in the middle of the S phase. What are your thoughts about that 60 percent of the chromosome?

Dutta: That's a great question. The pan-S phase pattern was a surprise. It was hard for us to believe. So we have reanalyzed



Anindya Dutta is a Byrd Professor in the Department of Biochemistry and Molecular Genetics at the University of Virginia. He received his Ph.D. from the Rockefeller University and performed his post-doctoral training at Cold Spring Harbor (CSH). Dutta's laboratory studies the regulation of the G1 to S-phase transition in normal and cancer cells with the goal of learning how to restore control to the cell cycle, preventing unrestrained growth of cancer cells.

the data in a different way. Instead of looking at just the p value of replication in a given time interval, we have done TR50 calculations—the time at which a given segment is 50 percent replicated. Using that, we can decrease the pan-S phase area to 30 to 40 percent. So it's probably not 60 percent.

Botchan: I still have an open mind that origin selection may be random for some unknown fraction of the genome, depending on the cell type. We know that in early embryos that is the case, and even for the hard-wired phage, the T-phage, for example, you can delete an origin and you can still get by. So it may be that we don't understand why there is specific origin selection or when it occurs. The default may be that when there is no reason to replicate at a specific time, timing is essentially random.

Dutta: Right. Let me tell you about the experiment we have done. We saw 30 to 40 percent pan-S replication. If you look at the area that the ENCODE Project works on, which is a completely different segment, the number is 20 to 30 percent.

One thing we were concerned about was whether our results were due to the microarray we used. So, we confirmed our result by interphase FISH experiments, where you can count the number of dots produced by a given chromosomal segment. So you are sure that you're not seeing cross hybridization. And the pan-S phase pattern was not an artifact of the microarrays.

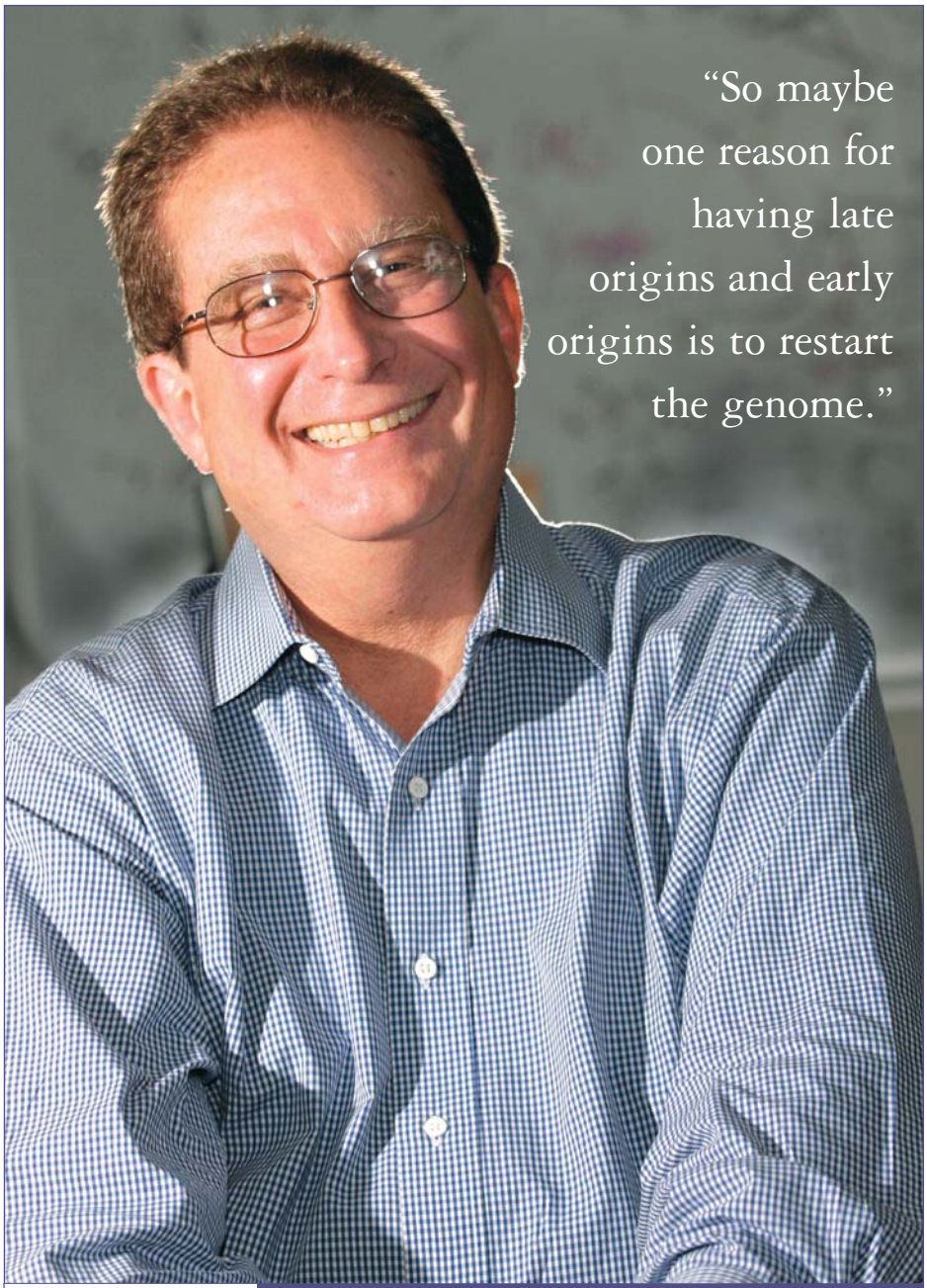
We were also worried about whether our results were due to the type of block we were imposing, because we did thymidine-aphidicolin block and release. So we repeated the experiment by using a nocodazole block and release, and we still saw the pan-S phase pattern of replication in HeLa cells.

Our last concern is whether this is something unique to HeLa cells. These are cancer cells and maybe they have been around for such a long time that they are completely messed up. So we are in the process of repeating the experiment in different cell lines to see if this pan-S phase pattern holds up.

Botchan: Even though HeLa cells are terribly aneuploid and there are lots of rearrangements, the domain of control may be reasonably local. That domain would divide a chromosome into a hundred different segments at least. And I don't think there are that many rearrangements in the HeLa cell nucleus. Of course, as the number goes down from 60 percent to 30 percent, it becomes more reasonable to think that it could be a result particular to that cancer cell.

Regulation of DNA Replication Timing

Botchan: In your paper you discuss that the early origins are in euchromatin and late origins are in heterochromatin. Where do you envision the rest of the origins—the 30 to 40 percent that replicate randomly? Are they sort of schizophrenic? Do they have a hard time deciding whether they're heterochromatin or euchromatin?



Michael Botchan **is a professor of biochemistry and molecular biology and an affiliate of the Division of Genetics and Development at the University of California at Berkeley. His laboratory is using whole-genome tiled microarrays for *Drosophila* to investigate the mechanisms and regulation of the initiation of DNA replication in eukaryotes. They are particularly interested in how these processes are coupled with the cell cycle.**

Dutta: When we did the interphase FISH studies, we could determine whether the pan-S phase replication was due to intercellular variation or interallelic variation. It turned out to be due to interallelic variation. That makes us think that the pan-S phase areas are schizophrenic in the sense that the two alleles are in different chromatin environments. Some of them are in the open chromatin, so they replicate early and other alleles are in the heterochromatin, so they replicate late. When we look at the global replication signature, we see replication in multiple quarters of S phase.

Botchan: Your observation that one allele has one phenotype and the other allele has another phenotype would help to explain the data, but another way to look at it is that the origins chosen for a particular time have particular types of proteins associated with them and those proteins specify whether they will replicate early or late. If an origin doesn't have those proteins, then the replication is random.

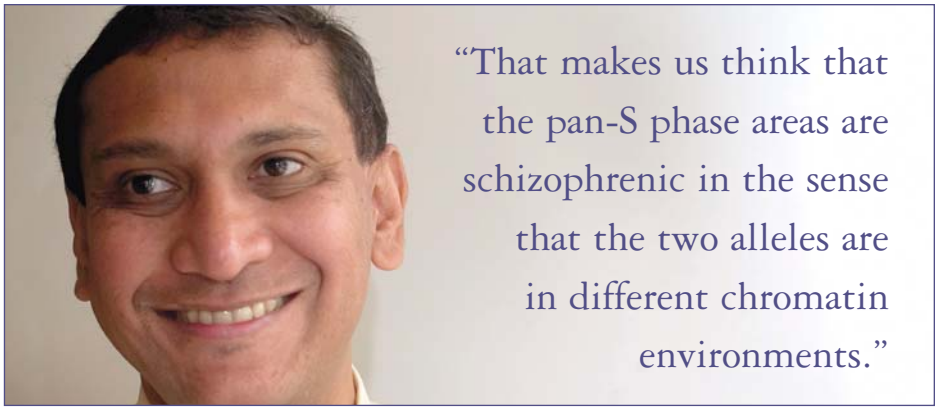
That would give you some kind of differentiation beyond euchromatin and heterochromatin, because there are data that show there are things that replicate very early in S phase and things that replicate a little bit later and things that replicate even later. That might require specialized proteins.

Botchan: What is the reason for having early versus late replication?

Dutta: One possibility is that early replication is a by-product of the fact that some segments of the chromosomes have very active transcription. So it's not that early replication is necessarily required, but the active transcription allows these parts of the chromosomes to be more open and attract replication factors earlier. So, the early replication is just a correlative feature rather than something required for the cell's survival.

Botchan: So you don't hold to the possibility that replication origins are at least sometimes important cis-acting genetic elements? What I'm thinking of is our own work with chorion amplification. And there, we know that replication can be uncoupled from transcription because the amplification occurs before the transcription.

Dutta: I don't believe that replication origins contain important cis-acting elements



that interact with replication factors in a temporally specific manner regardless of where the origin is located in the chromosome. I'm slightly discouraged that the data is tracking almost entirely with chromatin structure, gene density and RNA levels. If there is a cis-acting genetic element that dictates an origin, it's possible that it's pretty heterogeneous, like promoters are. When people first looked at promoters, they thought they were very defined—there was a TATA box and a CCAT box and all sorts of things like that. And gradually that definition is becoming weaker and weaker. Now it appears that different combinations of factors can bind to make a promoter.

Botchan: Yes. But the details are important in understanding the heterogeneity and selection mechanisms. I think it is important to keep the possibility open that there really is a mechanism that determines at least a few of the origins.

So going back to some old ideas, in yeast it certainly appears as if you can get by with very few origins. I believe that there was a suppressor of CDC7, an MCM mutant, that had almost no phenotype except that late origins were not fired and the cells became very sensitive to radiation. So maybe one reason for having late origins and early origins is to restart the genome.

In prokaryotes, you have primosomes and very elaborate restart mechanisms, so that if a replication fork is stalled, it can get going again. Now maybe in eukaryotes, you have that kind of restart mechanism, but you also have late origins.

Dutta: Yes. That's a possibility. So, there would be very strong selection to specify at least a few origins that were discretely late. That way, when the early origins get stuck, and they inevitably do, you can finish off the chromosome. You probably wouldn't see the phenotype until after many divisions.

Botchan: The other possibility, of course, is the old idea that you replicate a region to clear all the proteins off the DNA so that you can reprogram and get new factors on. The assembly of those factors is coupled to the growing fork.

As you say in your paper, we now have hundreds of examples we can use to understand the rules that determine the specificity. And once we start getting these details, we'll be able to make more sense out of what the mechanisms are and why they are there in the first place.

Validating and Creating Standards for ChIP-on-chip data

Dutta: Do you worry about the fact that there is a lot of noise in the data? Why is it so jumpy and why do you need the smoothing-out algorithms to get a reasonable data set?

Botchan: It's an issue, but I don't worry about it too much, because I think in the end, the important issue is what we have learned from the pattern and to determine whether there is anything we can do to allow us to go further. So, for example, I look at the smoothed data in your paper and I see a few regions that look like well-defined early origins and well-defined late origins. Then we can use that to go further and define what's special about those DNA sequences. The data becomes useful if important conclusions can be followed up.

Dutta: I have some interesting results from the recent ENCODE work we've done. The ENCODE Project is looking at the same DNA segments, but they are using different techniques. We find that

our early-replicating areas are co-segregating with areas rich in histone modifications. And the late-replicating areas are relatively depleted of them.

Botchan: Right, so you have a completely independent mechanism of experimentation that allows you to validate your conclusions.

Dutta: We are still looking at the temporal profile, but once you come to the narrow 10 to 30 kb area where we think the origins are, how much of that has been validated? For example, in *Drosophila*, have you gone back and looked at the 60 origins that David MacAlpine has mapped and seen whether they are origins by any other criteria?

Botchan: No. I think that's really where we are now. Even in very discrete amplification zones, if you look very carefully at where replication initiates and where the origins are, the way we used to do with viruses, it turns out that it's not a discrete point. So, I think that we are looking at something that we don't yet understand. Even the deposition of ORC (origin recognition complex) may not independently control where the first nucleotides are going to be laid down in a given cell.

Dutta: One thing I learned during these experiments is that ChIP-on-chip studies, at least in the mammalian cells, give widely differing results from platform to platform. How do you think we can best validate data from ChIP-on-chip experiments?

Botchan: We're going back and very carefully taking the ChIP-on-chip data and validating it with quantitative PCR. We're trying to get just classes we want to study. So, the ChIP-on-chip data tells us that in repression, sometimes Myb is a silent partner and E2F2 is doing the targeting. We have other examples where E2F is the silent partner. E2F is not required, but Myb is. We couldn't have found those without the ChIP-on-chip data, but then we go back and validate it with quantitative PCR. Using MacAlpine's data, we're getting close to 100 percent validation.

Dutta: In mammalian cells we're not getting that kind of validation rate.

Botchan: A lot of it may depend on reagents and how the thing is done. I would guess that the more you're trying to do, the more errors are going to be made. Let me give you a simple example. If you try to do ChIP-chip with ORC, at some sites, it turns out that it's very hard to see ORC1, and at other places, you can see ORC1, but it's very hard to see ORC5.

We have the same thing with the Myb complex. When we were first getting the ChIP-chip data with Myb complex, where Myb is a silent partner, Myb itself doesn't give nearly as good a signal as E2F2. But when we go to places where E2F2 is a silent partner, Myb gives us a much better signal than E2F2. So why is that?

Our working hypothesis is that it has to do with the conformation of the complex at these different sites. So at site one, where Myb is a silent partner, Myb is not terribly exposed and E2F2 is, and it changes at the other site.

Dutta: I'd think that if it's just biological variation, then E2F in Berkeley and E2F at MIT would produce the same patterns.

Botchan: Yes. That depends on whether they're using the same kind of polyclonal antibodies with the same affinities and using the same cut-offs with the same p values.

Dutta: One of the ways that I'd like to follow up on the replication experiment is to do the ORC ChIP-on-chip, but I've been sort of put off by the fact that groups in the ENCODE project have compared data on the same region in mammalian cells and they're finding lots of differences between the data sets.

Botchan: Have people exchanged reagents?

Dutta: They are doing that now.

Botchan: Well, I'm sure it will all sort itself out. I'm still a big proponent of using the genomic approach to get some good examples to understand the details.

FOR MORE INFORMATION

Contacts

■Anindya Dutta, M.D., Ph.D.
Department of Biochemistry &
Molecular Genetics
University of Virginia Health System
Jordan Hall Box 800733
Charlottesville, VA 22908
ad8q@virginia.edu

■Michael Botchan, Ph.D.
Department of Molecular & Cell Biology
University of California, Berkeley
16 Barker Hall
Berkeley, CA 94720
mbotchan@berkeley.edu

Companies

■Affymetrix Inc. –
http://www.affymetrix.com

Organizations

■University of Virginia
http://www.virginia.edu

■University of California, Berkeley
http://www.berkeley.edu

■The ENCODE Project
http://www.genome.gov/ENCODE/

Further Reading

■Jeon Y, Bekiranov S, Karnani N, Kapranov P, Ghosh S, MacAlpine D, Lee C, Hwang DS, Gingeras TR, Dutta A. Temporal profile of replication of human chromosomes. *Proc Natl Acad Sci* USA 2005; 102(18):6419-24

■Machida YJ, Teer JK, Dutta A. Acute reduction of an origin recognition complex (ORC) subunit in human cells reveals a requirement of ORC for Cdk2 activation. *J Biol Chem*. 2005; 280(30):27624-30.

■Bell S P. and Dutta A. Initiation of DNA replication in eukaryotic cells. *Annu Rev Biochem*. 2002; 71:333-374.

■Bandura JL, Beall EL, Bell M, Silver HR, Botchan MR, Calvi BR. humpty dumpty is required for developmental DNA amplification and cell proliferation in *Drosophila*. *Curr Biol*. 2005; 15(8):755-9.

■Remus D, Beall EL, Botchan MR. DNA topology, not DNA sequence, is a critical determinant for *Drosophila* ORC-DNA binding. *EMBO J.* 2004; 23(4):897-907.

■MacAlpine DM, Bell SP. A genomic view of eukaryotic DNA replication. *Chromosome Res.* 2005; 13(3): 309-26

People

■David MacAlpine
http://www.macalpine.org/research

Trends in Whole–Genome Tiling Array Studies

Recent whole genome array analysis at near nucleotide resolution has drastically altered functional models of RNA transcription, DNA replication and the protein-nucleic acid interactions that regulate these processes. Scientists have revisited the notion of "junk" DNA as recent studies of the human transcriptome have demonstrated that large regions, beyond just coding segments, are transcribed. These studies revealed that up to ten times more expression than predicted is evident in the human genome. Other studies that have used chromatin immunoprecipitation followed by microarray hybridization (ChIP-on-chip) have demonstrated that transcription factor binding is not restricted to the canonical upstream promoter regions. Genome-wide methylation and acetylation analysis have revealed patterns that may globally regulate chromatin accessibility. Even our understanding of the long studied, basic cellular process of replication has been altered by high-resolution analysis of the timing of initiation events.

1. Carroll J, *et al*. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell*. 2005; July 12; 122(1): 33-43.

Little is known about the organization of the cis-regulatory elements in the human genome. In light of estrogen receptor's (ER) dual role in normal female development and reproduction, and in the proliferation of cancer cells in the reproductive organs, the identification of targets for this ligand-dependent transcription factor has been of great interest. In this paper, Carroll and colleagues mapped the sites of ER binding on all non-repetitive regions of human chromosomes 21 and 22 by combining chromatin immunoprecipitation (ChIP) with hybridization to tiled microarrays. Their results suggest that ER binding requires the binding of a forkhead protein in close proximity. The repression of FoxA1 expression blocks estrogen response in breast cancer cells, further confirming the microarray results. This adds a potential target to the growing list of breast cancer therapeutic strategies.

2. Cheng J *et al*. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science*. 2005; 308(5725):1149-54.

With an interrogation resolution at least seven times as high as that of previous studies, Cheng and colleagues examined approximately 30 percent of the human genome encoded in 10 chromosomes (6, 7, 13, 14, 19, 20, 21, 22, X, and Y) and mapped the sites of transcription for polyandenylated and non-polyadenylated transcripts distributed in the cell nucleus and cytosol. They found that a much higher percentage of the genome is transcribed than previously assumed and that many of these transcripts have not been annotated. Highlighting the enormous complexity of the human genome, many of these transcripts have no known function and come from overlapping sense and antisense strands.

3. Bernstein B, *et al*. Genomic maps and comparative analysis of histone modification in human and mouse. *Cell*. 2005; 28;120(2):169-181

Spanning an unprecedented 37 MB of the human genome, as well as 2 MB of orthologous mouse genome, Bernstein and coworkers mapped histone H3 lysine 4 di- and trimethylation and lysine 9/14 acetylation using chromatin immunoprecipitation combined with oligonucleotide tiling arrays. This work revealed punctate methylation sites with trimethylation correlating with transcription start sites and dimethylation elsewhere in the vicinity of active genes. The Hox clusters display strikingly different patterns with broad H3K4-methylated regions that could be many kilobases in length, and likely play a significant role in their regulation. Methylation patterns were conserved between orthologous human and mouse loci at locations where the underlying sequence is only modestly conserved.

4 Jeon Y, *et al*. Temporal profile of replication of human chromosomes. *Proc Natl Acad Sci USA*. 2005; 3;102(18):6419-24.

The work described by Jeon and colleagues significantly challenges models describing the location and firing of origins of replication. Using timed replication pools of DNA harvested from synchronized human cancer cells, they found that a significant percentage of chromosomal segments interrogated on high density tiling arrays replicated in all quarters of S phase. This provides evidence that replication forks and origins, at least in aneuploid HeLa cells, are variably located and asynchronous.

Additional Articles

■Kapranov P, *et al*. Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. *Genome Research.* 2005; 15:987-997.

■Cawley S, *et al*. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell*. 2004; 116 (4):499-509.

■Kampa D, *et al*. Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res.* 2004; Mar;14(3):331-42.

■ENCODE Consortium, The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science.* 2004; (306): 636-640.

■Kapranov P, *et al*. Large-scale transcriptional activity in chromosomes 21 and 22. *Science*. 2002; 296(5569): 916-9.

Reviews

1. Mockler, TC *et al*, Applications of DNA tiling arrays for whole-genome analysis. *Genomics*. 2005; 85(1):1-15.

from page 1

Joseph Califano, associate professor of otolaryngology at Johns Hopkins, has studied the effects of carcinogens on the mitochondrial genome for the past four years. His team previously found that both tumors and saliva from head-neck cancer patients contained an increased amount of mitochondrial DNA content. Califano's hypothesis is that mitochondrial mutations result in an increase in mitochondrial mass, which could be due to nuclear-encoded factors that are translocated into the mitochondria, resulting in proliferation.

"It's a very, very old observation that mitochondria are increased in head-neck squamous cell carcinoma and pre-malignant diseases," said Califano. "But there's a little bit more of the molecular biology that may be understood today. We're finding a measurable rate of mitochondrial mutation in head-neck cancer."

Califano recently sat down with *AMB* contributor, Amanda Parrish, a former member of the Emory University research team that made the discovery that mitochondrial DNA mutations increase tumorigenicity in cancer. The two discussed the challenges facing cancer researchers studying mitochondrial genetics. Topics covered included:

- The discovery of mitochondrial mutations in "hot spots"
- The functional consequence of mitochondrial mutations and their link to cancer
- The use of saliva as a diagnostic fluid for cancer biomarker detection



Joseph Califano **is an associate professor of otolaryngology at the Sydney Kimmel Comprehensive Cancer Center at Johns Hopkins Medical Institute. He received his M.D. from Harvard Medical School and performed his postgraduate training at The Johns Hopkins Hospital. Califano's current research interests include investigating the molecular basis of head and neck cancer development, with a particular interest in molecular alterations in premalignant lesions, genetic instability, mitochondrial dysregulation, as well as molecular detection and surveillance.**

Detecting Mitochondrial Mutations

Parrish: Is the increased ratio between mitochondrial and nuclear DNA related to specific mutations or changes in other specific proteins?

Califano: Right now we're looking at the *in vitro* effect of tobacco smoke extract and tobacco smoke exposure on mitochondrial function and mitochondrial content in both an acute and chronic setting. We have research that indicates that mitochondria from head and neck cancer cannot maintain appropriate depolarization of the mitochondrial membrane. There is evidence of decreased mitochondrial function within head and neck cancers. In addition, we have unpublished data which indicates that there is a defect in depolarization-induced apoptosis when head-neck squamous cell carcinoma cell lines are given depolarizing agents for example. There's no apoptotic response, which there obviously should be, despite release of commonly known apoptotic-inducing factors into the cytosol.

Parrish: Are some of these mutations, or other proteins, involved in affecting the ratio of the DNA? Are you using the MitoChip to detect these mutations?

Califano: We're using the MitoChip to do a fairly complete sequencing of a large cohort of head-neck squamous cell cancers and pre-malignant lesions. To date, we've sequenced about 85 head-neck tumors and the white blood cell DNA from matched pairs of those patients, so that we can get an accurate representation of the normal genetic polymorphisms unique to an individual, versus the somatic mutations occuring within the tumor itself. So far, we've found a mutation rate of about 50 percent, which is pretty impressive.

There was a paper in the June 2005 issue of *Nature Genetics* that did deep sequencing of the entire kinome for breast cancer, and it was remarkable how infrequent point mutations, and other mutations, were. So, this gives us some indication that mitochondria are, at least for this tumor type, probably a hot spot for mutations. Of those mutations, most of them are non-conservative amino acid changes. So now we're looking at the functional effects of the mutations of various mitochondrial-encoded proteins, using mitochondrially-targeted transfection schemes. We've found some functional differences, so it's been fruitful from that point of view.

Mitochondrial Mutations are in Hot Spots

Parrish: Have you found that the mutations are more localized to certain proteins in the genome, or that it is randomly interspersed between the D-loop and proteins?

Califano: Well, it's not stochastic, it's not random, and there's clearly evidence for selection of specific mutations. There are mutations spread throughout the mitochondrial genome, but there's clearly at least one or two hot spots that we've found so far.

As you know, the D-loop is commonly reported to be a mutation hot spot and one of the advantages of the second generation MitoChip is that it includes D-loop sequences. When we look at the ability of the second generation MitoChip to detect mutation, there is an obvious increase in the number of mutations that can be detected. So it has confirmed a lot of those general impressions that have been out there in the literature.

But even in areas of the mitochondrial genome that are covered by both chips, the second generation chip has a higher sensitivity and identical specificity when compared with the first generation chip. So, in general, it's been helpful because we've detected more mutations using the second generation. There are other hot spots, but I'd prefer not to share them right now.

Parrish: That's definitely exciting for potential biomarkers.

Califano: The key is to be able to sequence a large number of tumors. Eighty-five doesn't seem like a lot, but to sequence a whole mitochondrial genome in 85 primary tumors and matched normal tissues—it's a lot of data. We did a pilot experiment, where we looked at something like 15 to 20 tumors and matching normals using both the first generation and the second generation chip. And, like I said, the second generation has better sensitivity and picked up quite a few more mutations. We confirmed them all 100 percent of the time with standard cycle sequencing, so it's been very, very reliable. Then we used the second generation chip on the 85 tumors and all 85 matched normal white blood cell fractions. That's where our 50 percent mutation rate comes from. And all those mutations were also confirmed using cycle sequencing as well. It's impressive data.

I'm excited about the study we're doing in terms of mitochondrial mutations because quite frankly, I'm always a little skeptical about deep sequencing projects because they seem something like a fishing expedition. In this case—I'm not sure how deep you'd call the sequencing—but it's deep enough that it demonstrated that this may be a real hot spot, at least for our particular tumor type.

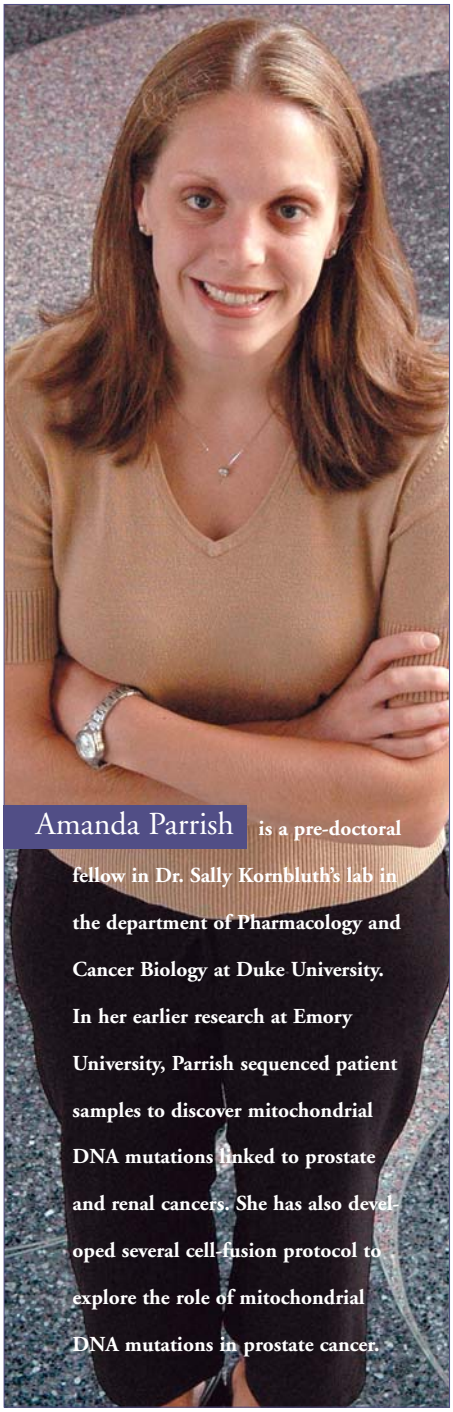
Linking Mitochondrial Mutations to Cancer Progression

Parrish: Determining the functional outcome of any mutation can be a challenge. What strategies are you using to figure out how these mutations relate back to the patient's disease?

Califano: Once you find coding abnormalities in mitochondrial DNA, it's fairly straightforward to use targeted transfection strategies to tease out what those functional effects are. I think the real challenge though is to figure out what the effect of protein trafficking in and out of mitochondria—particularly for nuclear-encoded mitochondrial genes—is on mitochondrial function.

There's no question that mitochondria, both in terms of reactive oxygen species production, involvement with apoptosis and perhaps some other functions, probably have a lot to do with cancer. But it's been very difficult to understand how nuclear-encoded genes are imported into mitochondria, how they become functional and what may interfere or change those processes.

We have interesting data that examine this when you look at hereditary paragangliomas, for example. These tumors are caused by mutated nuclear-encoded genes that result in mitochondrial alterations of the Krebs Cycle and subsequently result in neoplastic development, angiogenesis and invasion.



Amanda Parrish is a pre-doctoral fellow in Dr. Sally Kornbluth's lab in the department of Pharmacology and Cancer Biology at Duke University. In her earlier research at Emory University, Parrish sequenced patient samples to discover mitochondrial DNA mutations linked to prostate and renal cancers. She has also developed several cell-fusion protocol to explore the role of mitochondrial DNA mutations in prostate cancer.

Teasing out the effects of gene dosage can be tough, but it's an important issue, it's critical when you're evaluating what coding alterations in mitochondrial genes really do because it probably has a big effect on function.

It requires interesting mental gymnastics to figure out what results in increased selection for mitochondrial proliferation as

opposed to cellular proliferation. Presumably without some influence on cellular proliferation, an alteration in the mitochondrial function or mitochondrial genome wouldn't result in persistence of a clone with an altered mitochondrial population within it. But, nonetheless, it's very difficult to try to tease out how mitochondria act within the context of a single cell as opposed to how it affects function and the competitive advantage of the cell, in which that mitochondrial population resides. So it's a bit of a double think that you always have to do when you're looking at these issues, and I think it would be interesting to try to sort out which one of those is the more important effect, or how they balance.

Parrish: Absolutely. Increased mitochondrial mass can lead to increased cellular proliferation, but it's hard to tease out the specifics of what's causing what and if they're directly, or just indirectly related. What transfection procedures are you using to look at specific mitochondrial DNA mutations?

Califano: Well, you can use cybrid models, you can use mitochondrial localization sequences to force a protein of interest to become transported into mitochondria. But I think those can be crude tools when we look at cell lines that are rho-nulls, they still have mitochondrial-like structures in them, and it's unclear what kind of biochemical processes occur there. To select these cells, we subject them to incredibly toxic and long-term regimens, which probably have effects on the nuclear genome as well. The other question is, if you have a plasmid-encoded gene or nuclear-encoded gene that codes for a mitochondrial protein with an artificial localization sequence that's driven under a promoter that's just driving it like crazy—is that really an accurate reflection of the normal transcription level and mechanism by which the protein gets translocated to the mitochondria? Even with cybrids, you can make the argument that unless you do complete sequencing of the mitochondrial genome, you don't know what your cybrid is doing.

Using Salivary DNA Samples

Parrish: How accurate is a disease profile that you might obtain from saliva as related to head and neck cancer?

Califano: It just depends on what kind of marker you are using. People have tried all types of things. You can look at proteins or nucleic acids—two general classes. And people have started to look at RNA, but I think the data on that are quite preliminary. It's certainly challenging, because there are issues of collection, and there are issues intrinsic to the sensitivity and specificity of the assay itself. We are using fairly normal PCR-based molecular techniques and we can routinely detect a signal of one abnormal allele, if you will, in the background of about 1000, upwards of 10,000, wild type alleles. That's where most mutation detection strategies fall. The issue, in terms of clinical presentation, is where a tumor is. For example, we found—we haven't published this either—that oral cavity tumors are more easily detected than those in the oropharynx. There are some phenotype-specific alterations—for example, human papilloma virus detection is much easier for oropharynx tumors, because they have a higher rate of human papilloma virus integration. So there's some tailoring in terms of which marker will be appropriate and there's other tailoring in terms of being able to have some measurable quantity of abnormal signal—be it DNA, protein or RNA—that you can quantitate in the sample that isn't completely overwhelmed by normal shed cells.

Parrish: Can you explain the method you used to correlate the tissue and the saliva?

Califano: When you look at the scatter plots, you see a fair amount of scatter on the measurement and you realize that within an individual population, there can be significant variation. So, the real key was to look at the clinical variables.

In these types of analyses, the tumor group, or the group of interest, is almost not as important as your control group, because the real devil in any detection methodology is confounding factors that exist within a control population.

FOR MORE INFORMATION

Contact Info

■Joseph Califano, M.D.
Department of Otolaryngology
Johns Hopkins University School of Medicine
720 Rutland Ave
Baltimore, MD 21205
http://www.hopkinshospital.org/Faculty/ENT/Califano.html

■Amanda Parrish
Department of Pharmacology & Cancer Biology
Duke University Medical Center
Box 3813
Durham, NC 27710
amanda.baumann@duke.edu

Organizations

■Johns Hopkins University School of Medicine – http://www.jhu.edu/

■Sydney Kimmel Comprehensive Cancer Center – http://www.hopkinskimmelcancercenter.org/index.cfm

People

■John Petros
http://www.emoryhealthcare.org/departments/wci/sub_menu/Prostate_Cancer.html

Companies

■Affymetrix Inc. – http://www.affymetrix.com/

Further Reading

■Maitra A, Cohen Y, Gillespie SE, Mambo E, Fukushima N, Hoque MO, Shah N, Goggins M, Califano J, Sidransky D, Chakravarti A. The Human MitoChip: a high-throughput sequencing microarray for mitochondrial mutation detection. *Genome Res.* 2004;14(5):812-9.

■Kim MM, Clinger JD, Masayesva BG, Ha PK, Zahurak ML, Westra WH, Califano JA. Mitochondrial DNA quantity increases with histopathologic grade in premalignant and malignant head and neck lesions. *Clin Cancer Res.* 2004; 10(24):8512-5.

■Petros JA, Parrish AK, Ruiz-Pesini E, Amin MB, Sun CQ, Hall J, Lim S, Issa MM, Flanders WD, Hosseini SH, Marshall FF, Wallace DC. mtDNA mutations increase tumorigenicity in prostate cancer. *PNAS.* 2005;102(3):719-24.

■Stephens P, Edkins S, Davies H *et al*. A screen of the complete protein kinase gene family identifies diverse patterns of somatic mutations in human breast cancer. *Nat Genet*. 2005; 37(6):590-2

SEQUENCE ANALYSIS

Genome Scan Finds Mutations in at Least a Quarter of Patients with Complex Mitochondrial Diseases

Maastricht University's Bert Smeets and Duke University's Amanda Parrish discuss using mitochondrial resequencing arrays to identify mutations responsible for mitochondrial diseases

Dr. Bert Smeets and his research group at Maastricht University have identified known and new mutations in at least 25 percent of patients who suffer from mitochondrial disorders. Screening for these mutations with resequencing microarrays could simplify the diagnosis and prognosis of complex mitochondrial diseases which often have conflicting clinical presentations.

"A thousand, or even several thousands of genes might be involved in the mitochondrial disease system," said Smeets, director of the Genome Centre Maastricht. "One of the reasons we went to the resequencing MitoChip for these disorders is that the data gives you a quick and comprehensive clue to whether the mitochondrial DNA or the nuclear genome is involved. Especially for genomes like the mitochondrial DNA, where it is possible to generate the template with a single PCR and where the vast majority of the point mutations are nucleotide substitutions, it will become the method of choice."

Smeets recently spoke to Amanda Parrish, a pre-doctoral fellow in Duke University's department of Pharmacology and Cancer Biology about the challenges and promise of diagnosing mitochondrial-related disease directly in the clinic. Formerly at Emory University, Parrish worked with Dr. John Petros to prove that mitochondrial DNA mutations increase tumorigenicity in prostate cancer.

The two discussed:

- Locating mitochondrial DNA defects and correlating them to the large variety of mitochondrial-related diseases
- The challenge of determining a threshold of heteroplasmy that produces a disease phenotype
- Counseling for individuals who have mitochondrial defects and new reproductive options for those individuals

Mitochondrial-related Diseases

Parrish: Is there a particular mitochondrial disease that your lab focuses on or is your lab interested in mitochondrial-related diseases in general?

Smeets: Our interest is in mitochondrial diseases in general, but we study mostly pediatric patients with neurological syndromes. But, because the Maastricht University is strong in cardiovascular research, we extended our research to cardiomyopathies with a focus on defects in oxidative phosphorylation.

We have done a number of gene expression studies on mouse models with a genetic or induced cardiomyopathy, and they, in fact, led us to a defect in mitochondrial function.We're trying to extend our work to patients with cardiomyopathies and, indeed, we found a number of mitochondrial DNA mutations. If a patient presenting with neurological symptoms has a mitochondrial DNA mutation, it is strongly advised to investigate the heart, as patients often die from heart problems, which may not be the predominant or presenting symptom.

We also moved into the field of diabetes because some papers last year reported the coordinated decrease in expression of OXPHOS-genes in muscle of individuals with type II diabetes. We will investigate mitochondrial DNA, gene expression and oxidative phosphorylation in a selected group of type II diabetes patients.

So we aren't looking at only one group of patients. Preventing the transmission of mitochondrial DNA mutations is another key objective of ours because you can't cure mitochondrial diseases, but at least you can try to prevent them from spreading through families.

Locating Genes Involved in Mitochondrial-related Disease

Parrish: What are some of the other big challenges you are facing in the field of mitochondrial genetics and sequencing?

Smeets: Well, for us, the problem with mitochondrial disease is that the clinical picture is extremely heterogeneous and that many genes can be involved. It can manifest as specific syndromes, but you can also have symptoms that might relate to a defect in the mitochondrial system, but they might not be specific. You can have type II diabetes, deafness or kidney disease. Any of those could be caused by mitochondrial defects by other factors entirely or a combination.

But if there are a number of possible mitochondrial symptoms in a patient, or his or her family, and additional investigation suggests a mitochondrial defect, you have to decide whether to look at the mitochondrial DNA or screen a substantial number of nuclear genes. A thousand, or even several thousands of genes might be involved in the mitochondrial disease system. One of the reasons we went to the resequencing MitoChip for these disorders is that you can get a quick and comprehensive clue to whether the mitochondrial DNA or the nuclear genome is involved. Especially for genomes like the mitochondrial DNA, where it is possible to generate the template with a single PCR and where the vast majority of the point mutations are nucleotide substitutions, it will become the method of choice. Given the clinical and genetic heterogeneity, there is often no direct line from the clinical picture to a gene. So you are stuck with the problem of where to look for the defect. This problem is common for many genetic diseases. The Human Genome Project has yielded a huge potential for genetic testing, but DNA-testing is not able to keep up with the pace due to the lack of smart protocols based on genotype-phenotype correlations, and the limitations of current mutation detection technology. Massive sequencing platforms, like resequencing chips, are essential to remove this bottleneck and provide rapidly complete genetic information to patients for all candidate genes. For example, if you put all nuclear Complex I

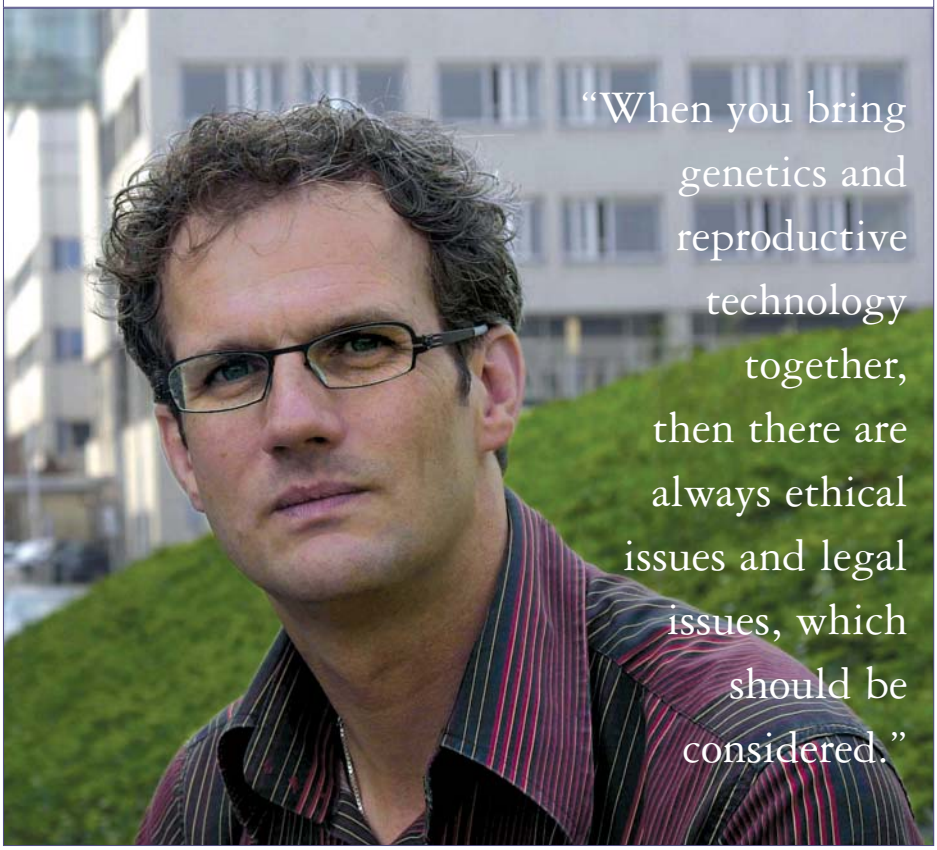
genes on a chip, you could analyze all the genes that might be involved in Complex I deficiency.

Parrish: You were just discussing characterizing different nuclear genes that affect mitochondrial DNA and disease. How do you plan on going about that? Do you have some target genes or are you planning on doing a lot of resequencing analysis?

Smeets: I think in the end, if you don't have a good idea where to look, you would probably have to resequence substantial numbers of genes that are involved in the clinical entity. At this point, that's only feasible for mitochondrial DNA. So we are using more classical approaches.

We look at familial studies, where you can do linkage analysis, or homozygosity mapping in the case of autosomal recessive disease. We are also doing biochemical complementation studies. The advantage of studying mitochondrial diseases is that you may have a cellular phenotype. You can have a measurable enzyme deficiency and if you fuse cells, you can see whether they complement each other or not. You can make complementation groups of patients with for example, Complex IV deficiency, and go from there to the gene defect either by linkage or by subcloning chromosomes to see if specific parts can complement the defect. In Europe, we have big mitochondrial networks, one of which is MitoCircle, and we try to gather patients to see if we can extend groups and find the defects. We are considering using as many patients from within the EU as possible. Then we could do homozygosity mapping with the 10K or 100K SNP-chips on patients from different countries and see if there are regions that are homozygous within groups of patients to help us localize the gene defect.

Bert Smeets is an associate professor in both Clinical Genetics and Population Genetics at Maastricht University in the Netherlands. He is also director of the Genome Centre Maastricht, which he founded in 2000. He received his Ph.D. from the University of Nijmegen in the Netherlands, where he studied the genetic basis of myotonic dystrophy. His laboratory at Maastricht University is focused on using high throughput genomic technologies to screen for mitochondrial and nuclear genes involved in mitochondrial-related disease. They are currently using the MitoChip to screen patients for mitochondrial mutations with the goal of one day preventing the transmission of mitochondrial disease to future generations.



Parrish: Did you actually obtain mitochondrial preps from different patients and then try to fuse the cells together and see if they complement each other?

Smeets: Yes. We collaborate with the group of Zeviani and Tiranti in Milan, Italy and they have done that very nicely to clone Surf1, a Complex IV gene involved in Leigh's disease. They found a large complementation group with Complex IV deficiency. From that point on, they knew that it must be the same gene involved and they did linkage analysis and created a map that led quite directly to the gene.

Parrish: What do you think are some of the best available *in vivo* models to examine the role of mitochondrial DNA mutations or mitochondrial-related diseases?

Smeets: Well, the models for mitochondrial DNA mutations are a problem to generate. Over the last decade several mouse models have been developed for mitochondrial diseases, but only two mouse models with mitochondrial DNA mutations exist. We are concentrating on mouse models that have cardiomyopathies to see whether defects of the OXPHOS system are a consequence of the pathology or an early event in the disease pathology or just a temporary change. So our emphasis is more on existing models of diseases than on creating new models for primary mitochondrial disease.

Using the MitoChip to Find Mutations

Parrish: How are you using the new microarray sequencing technique to identify mutations associated with human diseases?

Smeets: We use the chip for groups of patients where we don't see a common mutation. We generally screen for the most common ones, and then, if we strongly suspect a mitochondrial disease, we do a screening of the entire mitochondrial DNA.

We originally screened by HPLC analysis to look for heteroplasmic mutations, because we considered that the hallmark of pathogenic mutations, but it's obvious that there's also a considerable number of homoplasmic mutations that might cause mitochondrial disease. You would not pick those up, so we were looking for a system where you could do an analysis of the entire mitochondrial DNA, which would yield the complete sequence, but would also have substantial sensitivity to be able

to detect heteroplasmic levels as low as about five percent.

We do these analyses only on tissues where the disease is expressed. So if we analyze only muscle, you should have a mutation percentage that is usually above 30 percent or so, but it can be lower.

We sequenced a number of patients, then we mixed those who were different from each other at 60 positions in all kinds of percentages to establish heteroplasmy at those 60 positions. Then we looked to see how well the software was able to detect these low level mutations. We're still working to optimize that. We pick up some of the mutations at five percent, but others only at 50 percent, so that's quite a range. Basically, however, the results in our patients were quite good. We detected a large number of variants, solved at least 25 percent of the cases and we detected new heteroplasmic and homoplasmic variants that we have to validate to prove a primary or modifying role in the disease.

Correlating Mitochondrial DNA Mutations to Disease

Parrish: When I was in John Petros' lab, we talked a lot about how you like to see heteroplasmy in order to correlate a mutation with a disease. What experiments or pieces of data do you think are really critical to have in order to demonstrate that a mitochondrial DNA mutation is related to or causative of a specific disease?

Smeets: When you screen the mitochondrial DNA with the complete screening technology and find a mutation that has been published before, then it's quite straightforward. You know the mutation is pathogenic and you can inform the families. It's still difficult to give a prognosis for these families, because often these mutations have been published only once. And the description of the phenotype may be different from the clinical phenotype of the patient you're looking at. So you're stuck with being able to follow those mutations within the families, but not being able to predict the consequence of the mutation.

On the other hand, we find a lot of new ND5 mutations in patients with Leigh's disease or MELAS syndrome. If these mutations affect conserved codons in the gene, have a drastic consequence on the protein and relate to the biochemical defect, then we can correlate the mutation to the disease. If you find heteroplasmy that is co-segregating in the family, and you see that the higher the heteroplasmy levels are, the more severely people are affected, then you have another reason to argue that that mutation is causative for the disease. But you cannot always be sure and you have to refer to some mutations as unclassified variants.

Thresholds for Heteroplasmy

Parrish: You have said that heteroplasmy probably has to be at some kind of threshold level before a phenotype is actually observable. How do you determine that threshold?

Smeets: It's something that people have worked on with tRNA mutations, for example. They looked at percentage levels and ATP production to see how they correlated. You can see that beyond a specific threshold, ATP production drops, but it is difficult to do that in a patient. It's not really black and white. That's a problem for the families, in the sense that if you identify the level of mutation in muscles of maternal relatives, you can give only empirical data on their chances of developing a disease or transmitting it to their children.

We're also doing gene expression profiling in patients with mitochondrial DNA mutations to see if we can figure out if gene activity influences which patients will develop symptoms. We found some pathways, like oxidative damage, which might be involved. It's a complex picture as mutations do not work in isolation. You must consider the mutation in different tissues and the genetic background of the patient. And behavior makes a difference. Exercise is good. It's not a therapy, but it improves the status of some mitochondrial DNA patients. So many genetic and environmental factors play a role in how the disease manifests.

Testing and Counseling Families for Mitochondrial Defects

Parrish: You've mentioned a couple of times that you'd ultimately like to be able to decrease the number of OXPHOS and mitochondrial-related diseases that are transmitted. Would you elaborate on how you're going about it, and what you'd like to do in terms of counseling patients?

Smeets: When a family with an affected child wants to have additional children, you can test the mitochondrial DNA. If it is not a mitochondrial DNA defect, then autosomal recessive, dominant or X-linked disease will be more likely. The parents and their relatives can then be informed about the recurrence risks and reproductive choices.

If it is a mitochondrial DNA mutation, then you have a maternally segregating defect in the family. That is important to know in order to counsel that family in an appropriate way. The problem with mitochondrial DNA mutations is that if they are transmitted to the offspring, the point mutations go through a genetic bottleneck. That means that if the mother carries the mutation at say 25 percent, then she can produce oocytes, which can be between zero and 90 percent. Based on the percentage of her mutation load, she might have 10 oocytes, of which eight might produce severely affected children and the other two might produce healthy children.

So, what are the options for preventing the transmission of these diseases? You could do conventional prenatal testing, but you can't do that for all mitochondrial DNA mutations because we don't know whether there is a good correlation between the percentage of the mutation load in chorionic villus and the phenotype.

By looking at opportunities like preimplantation genetic diagnosis, you might be able to address that problem. You do an IVF procedure to get the oocytes fertilized, and then in an eight-cell stage of the embryo, you pick out two cells and type them for the mutation. Then you select the embryos with mutation loads the furthest below the threshold of expression. In that way, you can at least try to produce healthy children. At the moment the procedure is only reliable for a few mutations because data for most mitochondrial DNA mutations are so scarce.

When you bring genetics and reproductive technology together, then there are always ethical issues and legal issues, which should be considered. But I think the research we are performing creates possibilities for people who carry a mitochondrial DNA mutation and want to have children. They can become pregnant and take their chances or use options, like oocyte donation. But preimplantation genetic diagnosis may help select embryos that do not carry mitochondrial mutations and allow these people to have healthy children of their own.

FOR MORE INFORMATION

Contacts

■Hubertus (Bert) J.M. Smeets, Ph.D.
Dept. of Genetics and Cell Biology
Research institutes GROW and CARIM
Maastricht University
P.O. Box 616
6200 MD Maastricht
The Netherlands
bert.smeets@molcelb.unimaas.nl

■Amanda Parrish
Duke University Medical Center
Box 3813
Durham, NC 27710
amanda.baumann@duke.edu

Companies

■Affymetrix Inc. –
http://www.affymetrix.com

Organizations

■Maastricht University
http://www.unimaas.nl

■MitoCircle
http://mitocircle.unimaas.nl

■Duke University
http://www.duke.edu

■University of Nijmegen
http://www.ru.nl

Further Reading

■Jacobs LJ, de Coo IF, Nijland JG, Galjaard RJ, Los FJ, Schoonderwoerd K, Niermeijer MF, Geraedts JP, Scholte HR, Smeets HJ. Transmission and prenatal diagnosis of the T9176C mitochondrial DNA mutation. Mol *Hum Reprod.* 2005; 11(32):223-8.

■van den Bosch BJ, van den Burg CM, Schoonderwoerd K, Lindsey PJ, Scholte HR, de Coo RF, van Rooij E, Rockman HA Doevendans PA, Smeets HJ. Regional absence of mitochondria causing energy depletion in the myocardium or muscle LIM protein knockout mice. *Cardiovasc Res.* 2005; 65(2);411-8.

■van den Bosch BJ, de Coo IF, Hendrickx AT, Busch HF, de Jong G, Scholte HR, Smeets HJ. Increased risk for cardiorespiratory failure associated with the A3302G mutation in the mitochondrial DNA encoded tRNALeu(UUR) gene. *Neuromuscul Disord.* 2004 Oct;14(10):683-8.

■Petros JA, Baumann AK, Ruiz-Pesini E, Amin MB, Sun CQ, Hall J, Lim S, Issa MM, Flanders WD, Hosseini SH, Marshall FF, Wallace DC. mtDNA mutations increase tumorigenicity in prostate cancer. *PNAS* 2005; 102(3):719-24.

■Aldridge BA, Lim SD, Baumann AK, Hosseini S, Buck W, Almekinder TL, Sun CQ, Petros JA. Automated sequencing of complete mitochondrial genomes from laser-capture microdissected samples. *Biotechniques* 2003; 35(3);606-7, 609-10, 612.

People

■John Petros
http://www.emoryhealthcare.org/departments/wci/sub_menu/Prostate_Cancer.html

SEQUENCE ANALYSIS

New Microarray Improves Sensitivity of Cancer Detection by Pinpointing Small Changes in Mitochondrial DNA

NIST's John Jakupciak and Genesis Genomics' Ryan Parr discuss array-based resequencing for detecting heteroplasmies in mitochondrial DNA and application in the clinic

Researchers at the National Institute of Standards and Technology (NIST) have developed a new way of catching cancer in its earliest stages by using a mitochondrial resequencing microarray that has ten times more sensitivity than traditional sequencing methods. By scanning for these mutations in patient biopsies or body fluids, physicians could identify and treat the disease before irreversible cancer symptoms arise.

Dr. John Jakupciak, principal scientist at NIST, develops and evaluates protocols and technologies, such as the Affymetrix mitochondrial resequencing array, for validating cancer biomarkers and developing DNA diagnostics.

"I think the community understands that 20 to 25 percent is the limit to confidently detect mixtures using traditional fluorescence sequencing," said Jakupciak. "We've found that this array has a detection limit down to one to two percent."

Jakupciak recently spoke with Dr. Ryan Parr, vice president of Research and Development at Genesis Genomics. Genesis Genomics is currently developing a new universal detection tool that will identify the early stages of cancer. The two discussed the merits of microarray analysis in detecting heteroplasmies in mitochondrial DNA and the challenges of transferring those technologies to the clinic in assays that can detect cancer earlier than ever before.

The two discussed:

- The differences between microarray-based and fluorescence-based sequencing when looking for heteroplasmies in mtDNA
- The challenge of detecting and avoiding mitochondrial pseudogene interference from the nucleus
- The use of surrogate tissues, like blood or urine, in microarrays developed for diagnostic tests in the clinic

Heteroplasmy Detection

Parr: I think all of us who are working in early cancer detection and diagnostics are really interested in the issue of heteroplasmy. Can it be detected at low levels? And if it is present at low levels, does it have meaning? If you're going to use this microarray and resequencing technology for any kind of clinical application you

absolutely have to know the answers to those questions.

Jakupciak: Yes. I think the community understands that 20 to 25 percent is the limit to confidently detect mixtures using traditional fluorescence sequencing, but we've found that the MitoChip has a detection limit down to one to two percent. When you're talking about these low levels of heteroplasmy, technology that can detect them at a lower level is attractive. In our recent work, we've used traditional "gold standard" fluorescent sequencing and we get electropherograms where it just isn't possible to see the heteroplasmic call that we see on the MitoChip. Detection at a low level is important because of the fast progression of tumorigenesis. We'd all love to understand the biological meaning of heteroplasmy, but certainly we're just at the tip of the iceberg. I think that the community would also be surprised to hear that heteroplasmic calls are the majority.

Parr: We find both heteroplasmy and homoplasmy as well, but heteroplasmy is the more common lesion. When we find these low levels of mutations, it makes us want to know what's really happening at these low levels.

Are we going to find out that even normal cells have an underlying level of heteroplasmy that needs to be teased out from those heteroplasmic sites that might be disease-associated? Are we going to find that it's a dynamic mitochondrial molecule as opposed to what, I think, we've all been taught, that mitochondria are embedded in post-somatic tissue and, therefore, do not undergo genomic mutations?

Jakupciak: There's some confusion between using mitochondrial DNA for haplotyping the human races around the world and using mitochondrial DNA mutations as biomarkers for disease. Those who work on haplotyping look at more or less fixed mutations, whereas those of us who study disease look at more dynamic mutations. And although these molecular markers and DNA mutations, especially in the mitochondrial DNA, are considered molecular markers for disease, trying to distinguish between normal and diseased individuals is still going to be a challenge. I'm happy that people like you are presenting data that show there is a difference between the normal and the disease state.



John Jakupciak is a principal scientist in the Biotechnology Division at the National Institute of Standards and Technology. He received his Ph.D. in genetics from the Biochemistry Department at Texas A&M. His laboratory is tasked with developing standard methods to measure biomarkers for disease, including markers for cancer diagnostics. Their work suggests that a relatively simple diagnostic test based on a DNA microarray resequencing chip could promote more accurate and cost-effective early detection of some solid tumors.

Covering the Coding Region

Parr: As you've said before, we really need to look at the entire genome. Traditionally, people have used HVI and HVII from the mitochondrial DNA as markers in population and forensic studies. We've noticed in our work that many of the hypervariable sites in these regions seem to be mutational hotspots when looking at disease pathology and its relationship to somatic mutations in the mitochondrial genome. I think more information is to be had if you look at the gene coding regions, because that's where the damage really counts—particularly if the changes are non-synonymous.

Jakupciak: Yes, I agree. We found that full gene amplification can cover all of the sequence information, but trying to cover

the full genome using traditional fluorescent sequencing was a very intensive process. We needed to manually go through, and amplify and sequence all of the information. And because there's no infrastructure to realign all of this data, we had to enter all that information and have it aligned with the help of different software. Each mitochondrial genome is about 16,500 base pairs, harboring 37 densely-packed genes. So, you can imagine if we processed 78 samples, then tried to align the resulting several hundred thousand base pairs of DNA and find those few mutations, it would be very difficult and time consuming. There are only a couple of mutations per genome, after all. That's why we liked the mitochondrial array. In one analysis, we had all the information in a single report.

Parr: I know what you mean. We sequenced mitochondrial genomes from a series of individuals using a commercial process, and it took one full-time person six months to edit all of those sequences. That's time someone could really spend doing something more meaningful than cleaning up sequencing data

Jakupciak: In addition, when we had the same sequence or the same sample separately amplified, but analyzed on the MitoChip, we found that there were heteroplasmic calls. Then we went back to the fluorescent sequencing data and we found that we had overlooked that low level of heteroplasmy, because you can imagine the monotony of just visually scanning all that sequence data. After 16,000 base pairs you're not very keen to observe those low levels of heteroplasmy anymore.

Mitochondrial Pseudogenes

Parr: One of the things that people are going to be working on is archived formalin-fixed paraffin-embedded tissues. Currently, one of the limitations of using that source material is that it's damaged, so the amplicon size is fairly low. So when we look at complete mitochondrial genomes, we have to amplify 625 base pair segments. Now, I don't see that as too much of a problem, but when you get down to that size, once again nuclear mitochondrial pseudogenes become a concern. And since heteroplasmy is what these disease-associated mutations are called, we must be very careful, and define exactly what these heteroplasmies are such that there is no associated paralogous nuclear sequence. What are your thoughts on that?

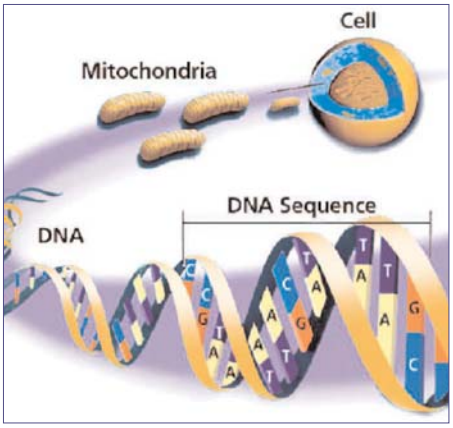
Jakupciak: Certainly going down lower and lower represents a problem because you're going down to a smaller amplicon size and what are you really amplifying then? Is it unique mitochondrial DNA or nuclear DNA? There are a number of primers that have been rigorously shown to not amplify nuclear DNA and, being from the Institute of Standards, I would like to see people use more standardized primers, although I agree that anyone can use a primer of their choice and that may be suitable for the results that they're obtaining. I guess your other question was what's the implication of using those smaller primer sets?

Parr: There seems to be a tendency to avoid the pseudogene issue, perhaps a reasonable solution may be to clone and characterize co-amplification products containing pseudogenes, then creation of a library may enable mapping of their nuclear locations. I agree that the best way to go is to use pseudo-free primers. On the other hand, as we move into the realm of mito-

"Diagnostics just tells you whether or not you have a disease. That's not much help if you have it, but if you can detect early, monitor, and follow, then I think we're making advances."



Ryan Parr **is Vice President of Research and Development at Genesis Genomics in Thunderbay, Ontario, Canada. He received his Ph.D. in biological anthropology from the University of Utah. Genesis Genomics is working on methods to use mitochondrial DNA as a biosensor for early detection and monitoring of cancer, particularly breast and prostate cancers.**



chondrial medicine and begin to understand the role of heteroplasmies and their relationship to disease, it is important to characterize and ensure that these signals are real and not nuclear artifacts.

Clinical Applications

Jakupciak: Since studies seem to show that mitochondrial DNA can be a biomarker for tumor detection, it would be nice to be able to use that knowledge to alleviate the need to do biopsies. So the next step will be to show or demonstrate that these other tissues could be used in place of collecting primary tissue.

I think that resequencing microarrays could be introduced to the clinic. As we mentioned, fluorescent signaling is very labor-intensive, and microarray analysis has the ability to shorten that time-frame and also save cost in terms of labor. But before anyone talks about clinical applications, there must be more cross-validation work; work where you exchange sample sets and independent labs can validate the results. These factors will be important to address in advancing any biomarker, as well as mitochondrial DNA, to the clinic.

Parr: Well, John, I agree whole-heartedly with that. With bladder cancer, for example, one could use urine that is already collected in the clinic and subject it to one more test, looking at mitochondrial DNA from bladder cells purified out of the urine, to find those tell-tail signs of developing cancer. I think we may have a tool that allows us to detect cancer and then monitor it. There's a big difference between that and what we're doing now with diagnostics. Diagnostics just tells you whether or not you have a disease. That's not much help if you have it, but if you can detect early, monitor, and follow, then I think we're making advances.

And, if you can detect these conditions before there are clinical symptoms, that has application to drug companies, for example, because now they have a whole new group of people that they can develop treatments for. So I think it's a win-win situation all around, and the future looks very bright and promising in this area.

There are some challenges, however, to developing these types of non-invasive procedures intended to avoid the need for biopsy. For example, if you take bladder cancer once again, there are a variety of cells in the urine. So the ability to purify bladder-specific cells away from other epithelial cells is vital for enabling a correct genetic signature. Obtaining appropriate fluids for testing, for example, for lung cancer, would be another challenge. How do you get fluid from the lungs? It is an issue we're grappling with.

Jakupciak: Yes, certainly the most universal body fluid is going to be whole blood or serum, and people have demonstrated that you can detect circulating DNA. But we're using the blood as our control in our studies to show how that is different from the tumor. So, although we have some limited evidence that perhaps the blood could be used as an early detection fluid, it doesn't seem like that would be feasible. If you're going to choose a body fluid then it must be one that's going to be appropriate for the type of cancer you're going to try to find. And another facet is that, in our hands, the fluid samples are more difficult to amplify and the percentage of sequence coverage is lower for the body fluid. With traditional fluorescent sequencing we could typically obtain a range of 95 to 100 percent sequence coverage. With the GeneChip microarray we would get, on average, about 98 percent, but then with the body fluid, it would drop down to 91 percent.

In the future, we must address the errors of each type of technology—to report on the accuracy and reliability of each so that anyone can determine whether to use GeneChip technology, fluorescent sequencing technology or other types of DNA analysis to answer their particular question.

FOR MORE INFORMATION

Contacts

■John Jakupciak, Ph.D.
Chemical Science and Technology Laboratory
Biotechnology Division
National Institute of Standards
and Technology
Building 227
Gaithersburg, MD 20899
john.jakupciak@nist.gov

■Ryan L. Parr, Ph.D.
Genesis Genomics
3rd Floor, 1294 Balmoral Street
Thunder Bay, Ontario
P7B 5Z5
Canada
ryan.parr@genesisgenomics.com

Companies

■Affymetrix Inc. – http://www.affymetrix.com

■Genesis Genomics – http://genesisgenomics.com

Organizations

■National Institute of Standards and Technology – http://www.nist.gov

Further Reading

■Jakupciak JP, Markowitz M, Ally D, Srivastava S, Wang W, Maitra A, Sidransky D, O'Connell CD. Mitochondrial DNA as a cancer biomarker. *J. Mol. Diagnostics*. 2005; 7, 258-267.

■Jakupciak JP, Markowitz M, Ally D, Srivastava S, Wang W, Maitra A, Sidransky D, O'Connell CD. High-throughput early detection of mitochondrial DNA mutations in cancer. *Nature Biotechnology.* 2005; (manuscript in preparation)

■Dakubo GD, Parr R, Costello LC, Franklin RB, Thayer RE. Altered metabolism and mitochondrial genome in prostate cancer. *J. Clin Path.* 2005; (in press).

■Parr R, *et al*. Use of mitochondrial DNA as a means of early cancer detection. *Human Genome.* 2005; (Invited review submitted).

SEQUENCE ANALYSIS

Microarrays Provide Fast and Sensitive Genetic Fingerprint for Forensic Investigations

University of Connecticut's Linda Strausbaugh and the California Department of Justice's Martin Buoncristiani discuss mitochondrial sequencing arrays in forensics

Researchers at the University of Connecticut (UConn) are using a new microarray to find mitochondrial mutations that genetically identify unknown samples in forensic studies more quickly and accurately than traditional sequencing.

Forensic science has stringent requirements for DNA technologies since sample sizes are often very small and difficult to isolate. Dr. Linda Strausbaugh, director of the Center for Applied Genetics and Technology at the UConn, hopes this new technology will save time and money and still deliver the high quality data needed to identify individuals using DNA evidence.

"We are looking at new technologies and learning to apply new instrumentation to forensic problems," said Strausbaugh. "We even investigate the use of DNA profiles of plant and soil in the analysis of evidentiary material."

Martin Buoncristiani, assistant director of the California Department of Justice's DNA Laboratory, is tasked with developing new technologies for forensics, and evaluating and validating new technologies for use in crime labs.

Strausbaugh recently spoke with Buoncristiani about her experiences using microarrays to look at mitochondrial DNA sequence for forensic studies. The two discussed:

- The limits of sensitivity for very small sample sizes
- The ability to pick up heteroplasmies and detect heterozygosity in DNA samples
- The potential scientific and cost-saving benefits in forensic laboratories

Limits of Sensitivity and Sample Size

Buoncristiani: Right now, the requirement is for 5 to 10 nanograms of high-quality DNA to run a MitoChip experiment. That is a lot of DNA for a forensic case or missing person identification. Are you working on increasing the sensitivity?

Strausbaugh: We think it can be much more sensitive than that, even as the protocol is currently configured. The mitochondrial resequencing chip, version 2, which is the one that we're using from Affymetrix, was primarily developed for mostly biomedical applications. If you read the original protocol, it recommends using 100 nanograms of DNA. So one of the first things we were interested in was learning whether this technology could be used with more forensic-style amounts of DNA. We began by trying 10 nanograms. That worked very well. We've used 5 nanograms. That also works very well. We think we can get below that. We don't know how far below that, but we're going do a dilution series soon to see how low we can go.

Buoncristiani: Great. We will soon be working with the Affymetrix mitochondrial microarray, and one of the first things we want to look at is sensitivity. Five to 10 nanograms of DNA is a lot for us. In a fair number of cases, maybe 20 to 40 percent, we have only trace amounts of DNA (less than 100 picograms). So, sensitivity is a very big issue for us. That's why we do mitochondrial sequencing to begin with, because the copy number is much higher than the nuclear copy number and we usually get a result.

I understand you're doing whole genome amplification for your experiments. So theoretically you could start with very little DNA, assuming that the amplification is efficient and accurate. What's your experience with the whole genome amplification? Is it usually pretty robust?

Strausbaugh: It's quite robust. This is a three-fragment amplification using two to three nanograms per amplification. In most of the samples, we're seeing pretty much equal representation of all of the three fragments. We've been doing it with three individual amplifications only because we wanted to be confident of what we were dealing with before we multiplexed them. But I can't see any reason that they shouldn't be done in a multiplex.

> "The microarray has built-in verification in its redundant tiling strategy."

Methods for Looking at Degraded DNA Samples

Buoncristiani: We would like to see if we can get it into a multiplex. As you know, one of the issues includes optimizing for degraded DNA, which we often have with forensic type samples. Do you think the amplicons are going to have to be shortened in order to make the whole genome amplification work with degraded DNA?

Strausbaugh: I think they probably are. These amplicons are large. The largest one, I think, is eight kilobases or so, so they will have to be shortened. One of the things we're going to do is to use the 58-primer set that John Butler designed for conventional sequencing of the coding region. Those primers generate smaller fragments. So we think using smaller amplicons will be feasible and will be a big improvement. One thing I like about the array technology is that multiplexing is not a problem. In talking with Affymetrix, 50 to 58 primer sets are not going to be a problem at all.

Also, in the array assay, one has to take the PCR fragments that are generated and take them through a fragmentation step, because smaller fragments are desirable for the assay. We've been wondering whether it is possible to design some very small amplicons for looking at degraded DNA. By very small, I mean one to two hundred bases. Amplicons this size would already



Linda Strausbaugh is a professor in the Department of Molecular and Cell Biology at the University of Connecticut and Director of its Center for Applied Genetics and Technology. She received her Ph.D. from Wesleyan Unviersity and did post-doctoral work at Johns Hopkins and the University of Pennsylvania. Strausbaugh is heavily involved in what she calls the "cornerstone" of the Center for Applied Genetics and Technology — a laboratory for non-traditional DNA typing, which works in support of crime labs around the country. The laboratory is tasked with finding ways to incorporate new information and technologies into forensic investigations. The lab has recently been exploring the potential use of microarrays for DNA-based identification.

be appropriately sized for the array labeling step and would allow us to avoid the fragmentation step altogether.

Detecting Mitochondrial Heteroplasmy and Heterozygosity

Buoncristiani: How sensitive do you think the detection of the low-level component in a mixture or heteroplasmic sample could be?

Strausbaugh: Well, we included a heteroplasmic individual in our trial. That heteroplasmy was detected, so we know that we can pick up at least the one heteroplasmic individual that we included. We feel that we still need to do a good deal of work with this.

We were talking with one of the Affymetrix field specialists who was at UConn this week about whether it would be possible to detect minor components in mixtures. We think that if we had a mixture of mitochondrial genomes in a sample, it should be possible to quantify the relative intensity of each of those four bases at each SNP. For example, if we have a minor mixture that's contributing 25 percent, the question is whether we could see that 25 percent at each of the SNPs that we called. If we could, we might be able construct a linkage map, in essence.

We think the sensitivity is actually better than 25 percent. I was just using that as an example. We have some very interesting samples in our Connecticut collection and one of the things that we intend to do is construct a mixture of genomic DNA using individuals for whom we already have sequence data and know exactly what the differences are. When we talked to Affymetrix about that, based on the quick look we took at our data, one might be able to get down to 25 percent pretty easily. I wouldn't be surprised if it goes lower than that.

A paper by Maitra *et al.* from *Genome Research* set out to look for heterozygosity. It's remarkable because in ten weeks, a single investigator sequenced and analyzed more than 2 million bases and did a study on the detection of heterozygosity. And he was able to detect heterozygosity at two percent, according to this paper.

Buoncristiani: Most impressive.

Trouble with Miscalled SNPs

Buoncristiani: In your pilot study, you sequenced mitochondrial DNA from 14 samples. I noticed that two out of 14 samples had incorrect calls in HV1, and for both samples it was the same base that was miscalled. Any idea what is going on there?

Strausbaugh: We're not sure yet, but let me bring you up to date. We now have three samples that have the same SNP miscalled. So in addition to the two we ran in our pilot study, we have a third that has the exact same SNP as those two individuals. Our guess is that it has something to do with the surrounding sequence. We're designing experiments now to try to look at that miscalling.

What I can tell you is that all three of these individuals share the same genetic, geographic history. So we think that we're

looking at some kind of ethnic-specific or ethnic-enriched combination of bases that's simply causing the miscalled base. Those are the only miscalls that we've found in all of the sequence that we've verified so far.

One of the things that is intriguing to us about the MitoChip resequencing array is that there is so much excess capacity, it's quite conceivable that one could add a haplotype specifically designed for this problematic area.

> "We think the real technical advantage of using an array approach over standard, automated sequencing is the capacity to multiplex."

Buoncristiani: That's encouraging that the only miscalls you've had have been for the same SNP. Short of putting another haploytype on there, you'd have to be suspicious about that position.

Strausbaugh: I think tiling more haplotypes is a good solution and will be required for a future forensics chip, in any case. You may recall from our data that although the chip performs well for most individuals, there is one person with a number of bases that are not called by the program. He is of Jamaican ancestry and his mitochondrial DNA sequence has many changes from the reference sequence. We know that the difficulties in this sequence arise from a stretch of 20 or 30 bases, in which there are four SNPs that are different from the Cambridge reference sequence, and therefore different from the common haplotypes that are tiled. We believe that these numerous mismatches in a small region compromise the technology, but can be corrected.

High-throughput Cost and Time-savings

Buoncristiani: What's the throughput like?

Strausbaugh: It's really impressive. The actual work for our project was done at Affymetrix, because the version 2 chip was still in final test stages and not yet available. Now it is available, so we are conducting research with it through the early access program. Our 14 samples were sequenced in their entirety in a matter of weeks.

Buoncristiani: Do you know what the actual experiment throughput time is from start to finish, after PCR?

Strausbaugh: We are now conducting experiments of our own, and we're talking about a day and a half, or two, to generate four complete mtDNA sequences after the amplification is completed. Basically it



Martin Buoncristiani is assistant director of the forensic DNA laboratory at the California Department of Justice, where he supervises the Method Development group. He received his MPH in Forensic Science from the University of California, Berkeley in 1989 and has been working to develop DNA technologies ever since. His group at the DOJ develops new technologies for forensic application and evaluates and validates new technology for use in forensic laboratories. He hopes that one day, forensic laboratories will be able to use microarrays for fast, accurate, DNA-based identifications.

takes one long day to do the amplification, the fractionation, and quick quantification. We do our hybridization overnight for exactly 16 hours. Then we go into the automated hydraulics system to rinse the chips; they can be scanned the same day. It is important to note that the actual hands on time for the scientist is considerably less than this since there is a lot of automated instrument time.

Buoncristiani: What is the relative cost of Affymetrix microarray resequencing compared to conventional di-deoxy sequencing?

Strausbaugh: We've thought about the reagent cost for obtaining an entire sequence using the Affymetrix GeneChip system versus doing a standard sequence with, let's say 58 primer sets. It looks to us like the reagent costs are going to be very, very similar, even when you include the cost of the chip.

We think the real technical advantage of using an array approach over standard, automated sequencing is the capacity to multiplex. When you're doing standard automated sequencing, you're limited to running one sequencing reaction on one column. So, it's a huge advantage to be able to multiplex reactions in the array system. And if you're doing the entire mitochondrial genome in a conventional way, remember you have to do both strands, because you have to verify the sequence. The microarray has built-in verification in its redundant tiling strategy.

The combination of savings in time with costs and multiplexing efficiency makes the microarray approach, I think, very cost effective. As I said, we run a small lab, but it is absolutely clear that the amount of time that one of our scientists spends getting four complete sequences from the Affymetrix chip is a whole lot less than it would take to obtain four full sequences by standard means.

FOR MORE INFORMATION

Contacts

■Linda Strausbaugh, Ph.D.
University of Connecticut
Molecular & Cell Biology, BH 350
354 Mansfield Road, Unit 2131
Storrs, CT 06269-2131
linda.strausbaugh@uconn.edu

■Martin R. Buoncristiani, M.P.H.
California Department of Justice
Jan Bashinski DNA Laboratory
1001 W. Cutting Blvd., Suite 110
Richmond, CA 94804
martin.buoncristiani@doj.ca.gov

Companies

■Affymetrix Inc. –
http://www.affymetrix.com

Organizations

■University of Connecticut –
http://uconn.edu

■California Department of Justice –-
http://caag.state.ca.us/

Further Reading

■Maitra A, Cohen Y, Gillespie SE, Mambo E, Fukushima N, Hoque MO, Shah N, Goggins M, Califano J, Sidransky D, Chakravarti A. The Human MitoChip: a high-throughput sequencing microarray for mitochondrial mutation detection. *Genome Res* 2004; 14(5):812-9

■Swango, Katie L., Timken, Mark D., Date-Chong, Mavis, Buoncristiani, Martin R. A quantitaive PCR assay for the assessment of DNA degradation in forensic samples. *Forensic Science International*. 2005 (in press)

■Timken, Mark D., Swango, Katie L., Orrego, Cristian, and Buoncristiani, Martin R. A duplex real-time qPCR assay for the quantification of human nuclear and mitochondrial DNA in forensic samples: implications for quantifying DNA in degraded samples. *Journal of Forensic Science*. (2005); 50(5)

■Date-Chong, M., Calloway, C.D., Klein, S.B., Orrego, C., Buoncristiani, M.R. Optimization of a duplex amplification and sequencing strategy for the HVI/HVII regions of human mitochondrial DNA for forensic casework. *Forensic Science International*. 2005 (in press)

■Crayton ME 3rd, Ladd CE, Sommer M, Hampikian G, Strausbaugh LD. An organizational model of transcription factor binding sites for a histone promoter in D. *melanogaster*. *In Silico Biol.* 2004;4(4):537-48.

■Strausbaugh L., J. Schienman, D. Kampa Bailey, R. Chiles, S. Walsh, G. Kennedy, J. Suhl, I. Ovchinnikov, L. Li, D. Renstrom, J. Tye, A. Berdine, C. Ladd, and H.C. Lee. Forensic Applications of GeneChip® Mitochondrial Resequencing Arrays (abstract) Proceedings of the Sixteenth International Symposium on Human Identification. 2005

■Hill, J., L. Strausbaugh and J. Graf, Soil DNA Typing in Forensic Science, in Forensic Non-Human DNA Typing, edited by Heather Miller-Coyle (in press)

■Schienman J. Techniques for DNA Fingerprinting, in Forensic Non-Human DNA Typing, edited by Heather Miller-Coyle (in press)

■Carita, E., Chapter 15: Classical and Future DNA Typing Techniques for Plants, in Forensic Botany: Principals and Applications to Criminal Casework, pp. 253-274, CRC Press, 2005

■Date-Chong, M., Calloway, C.D., Klein, S.B., Orrego, C., Buoncristiani, M.R. Optimization of a duplex amplification and sequencing strategy for the HVI/HVII regions of human mitochondrial DNA for forensic casework. *Forensic Science International*. 2005 (in press)

People

■John M. Butler, Ph.D.
DNA Technologies Group
National Institute of Standards and Technology
john.butler@nist.gov

Why use microarrays in forensic science?

AMB: What are the main challenges that people in the forensics community are facing that make them interested in using microarrays?

Buoncristiani: Our missing persons program tries to identify individuals based on DNA analysis of remains. These remains may have been exposed to their particular environmental conditions for as little as a few hours or as long as decades. It can sometimes be a real challenge to get information from these samples.

After DNA extraction we quantitate the DNA with a multiplex qPCR assay that we developed to amplify two different size targets. This allows us to assess the level of degradation present. From there, we make a decision whether to do STR typing or mitochondrial sequencing based on the quality and quantity of DNA recovered.

Mitochondrial sequencing is always the last resort, because it's not as discriminating as STR typing and it's fairly labor intensive to do. From start to finish, it's two or three days.

The Affymetrix chip technology is impressive because this approach is potentially much faster and yields more information. Currently, we sequence the mtDNA control region only. The Affymetrix resequencing chip provides sequence information for the entire mitochondrial genome, about 16 kb. This extra information can be very important for our application, because we often don't get discriminating results from the control region alone. Some of the haplotypes are fairly common. I think there's one that is found in seven percent of Caucasians. At that level of discrimination we get a lot of adventitious hits when we search a database of reference samples.

In addition, the FBI recently implemented a requirement that all reference samples searched with STRs on their CODIS Missing Person's database include mitochondrial sequence information as well. They use the mitochondrial sequence information to discriminate when they get adventitious hits. We have probably 700 or 800 reference samples in the laboratory right now that will need to be sequenced because of this requirement. So the higher throughput is very appealing.

AMB: What would make a transition to microarray technology possible for the forensics community?

Strausbaugh: When I talk with Affymetrix about this issue, I've suggested a true forensic chip. It would allow us to query a single array for mitochondrial DNA, a small set of autosomal SNPs and some Y SNPs. What an incredibly powerful tool that would be. I know it's going to be a while before the SNP technology filters down to routine case work, but the promise is just amazing. I think the applications for mitochondrial DNA are more immediate, especially for reference samples and others that are not as limited in amounts.

Buoncristiani: I agree. I've talked to Affymetrix about that type of a chip as well. I think part of the obstacle for adopting new technologies in this field is the investment that people have already made into STRs. It's been huge, and we've built all of these large databases. And the courts routinely accept STR typing results now. People are reluctant to change to another technology, even if it might be more automatable and have advantages in terms of throughput.

Strausbaugh: Right. When I talk with my collaborators at the state crime lab, what they point out is that every conversion to a new technology faces some resistance. But, I think just as there were opportunities to move from RFLP to the 377 platform with STRs, and then to the capillary electrophoresis platforms with STRs, I think there's going to be a future in microarrays too, assuming the costs come down.

Buoncristiani: It has occurred to me that as a transition for the forensic community, we could identify SNP haplotypes for each of the STR alleles for the loci that we currently type. So, you could do STR typing with SNP haplotypes and vice-a-versa. I think that might provide a transition that people in the forensic community would be more comfortable with, because all the efforts that have been put into the databases wouldn't be lost.

Strausbaugh: That is quite doable and it's an excellent idea. Along that similar line, I've been thinking that the MitoChip provides an opportunity for people to start thinking about the technology. Although I don't think you could interest people in autosomal SNPs to replace STRs right now, it's going to be clear that having additional SNP data from the mitochondrial genome (beyond HVI and HVII) is valuable in improving the power of discrimination. So, people will begin to get comfortable with the technology, the language, the software. I try very hard not to over-promote the microarray as a solution of immediate use to the crime lab. But I do see the MitoChip as something that might be of tremendous value for regional FBI labs and regional labs like yours, where you can begin to harness some of the power in a way that would be of immediate value to the forensics community.

Buoncristiani: Yes, that could be a good way to get people to start thinking about the advantages of SNPs. It's always a slow process in forensics because we have the courts to convince as well.

SEQUENCE ANALYSIS

Microarrays Detect Altered Expression of Mitochondria-Related Genes in Brains of Patients with Bipolar Disorder

RIKEN's Tadafumi Kato and the Japan National Institute of Neuroscience's Ryota Hashimoto discuss the altered expression and sequence variation of mitochondria-related genes in bipolar disorder patients.

RIKEN scientists led by Dr. Tadafumi Kato are using large-scale microarray analysis to determine how expression of mitochondria-related genes differs between healthy subjects and those diagnosed with bipolar disorder.

"Last year, a Harvard group did a DNA microarray analysis on samples from bipolar disorder patients and, surprisingly, reported that the expression of a number of nuclear genes related to mitochondria was down-regulated," said Kato. "We analyzed our data from a large-scale microarray analysis of independent samples and were able to replicate their findings, plus take them a step further."

Kato recently spoke with Dr. Ryota Hashimoto, section chief in the Department of Mental Disorder Research at National Institute of Neuroscience, National Center of Neurology and Psychiatry in Japan. Dr. Hashimoto's laboratory is interested in the genetic basis of psychiatric disorders and the molecular mechanism of action of psychiatric drugs. Hashimoto and Kato discussed:

- The regulation of mitochondrial-related genes in bipolar patients
- Polymorphisms in the mitochondrial genome of patients with biopolar disorder
- The potential benefits of mitochondrial resequencing microarrays for studies of psychiatric disease

The Role of Mitochondria in Bipolar Disorder

Hashimoto: You have been studying the genetic basis of bipolar disorder. What made you think the mitochondria might be involved?

Kato: About fifteen years ago, we conducted a study in which we used magnetic resonance spectroscopy (MRS), to look at brain metabolism in bipolar patients. The purpose of the study was to measure the metabolism of inositol DG. However, the most significant finding was an abnormality in energy metabolism.

We were thinking about why that might be when Francis McMahon at Johns Hopkins University showed that bipolar disorder is inherited maternally. In other words, it is inherited in a way that is consistent with mitochondrial inheritance. With this piece of information and our own data, we hypothesized that bipolar disorder might be a mitochondrial disorder.

Analysis of Gene Expression in Postmortem Bipolar Brains

Hashimoto: You have been doing extensive analysis of gene expression in postmortem brains. Did you find any differences related to mitochondria in bipolar disorder patients?

Kato: In the expression analysis from postmortem brains that we published in 2004, we found very few changes related to mitochondria. We left it at that. But last year, a Harvard group did a DNA microarray analysis on samples from bipolar disorder patients, and, surprisingly, reported that the expression of a number of nuclear genes related to mitochondria was down-regulated. So, we analyzed our data from a large-scale microarray analysis of independent samples and we were able to replicate their findings, plus take them a step further.



Tadafumi Kato is head of the Laboratory for Molecular Dyanamics of Mental Disorders in the Aging and Psychiatric Research Group at the RIKEN Brain Science Institute in Japan. He received his Ph.D. from Shiga University of Medical Science and his M.D. from the University of Tokyo. His laboratory studies the molecular basis of bipolar disorder with the goal of creating better diagnostics and new treatment options. They recently used microarrays to investigate gene expression in mitochondria-related nuclear genes of bipolar disorder patients.

However, from our detailed confounding factor analysis, we concluded that these were medication effects, so we don't think they really reflect the disease itself. In medication-free patients, some of the mitochondria-related gene expression was actually up-regulated, so that was contrary to the Harvard findings.

Hashimoto: How do the different drugs affect expression of mitochondria-related genes?

Kato: Antipsychotic drugs down-regulate the overall expression of mitochondria-related genes. Valproic acid has similar effects. It has been reported that both

antipsychotic drugs and valproic acid affect mitochondrial function adversely, so these findings are biochemically consistent.

Hashimoto: How does the Harvard study look if you separate medicated and non-medicated groups?

Kato: They said that medicated subjects had down-regulated gene expression, however, the gene expression was not significantly down-regulated in medicated schizophrenic patients, so they concluded that the down-regulation was due to the disease.

Hashimoto: So these genes were not down-regulated in schizophrenics even with medication?

Kato: They are down-regulated in schizophrenia as well, but the extent of down-regulation is more significant in bipolar disorder. However, in our study, the expression of mitochondria-related genes is most affected by the pH of the sample. The sample pH is strongly affected by the physical conditions before the patient dies.

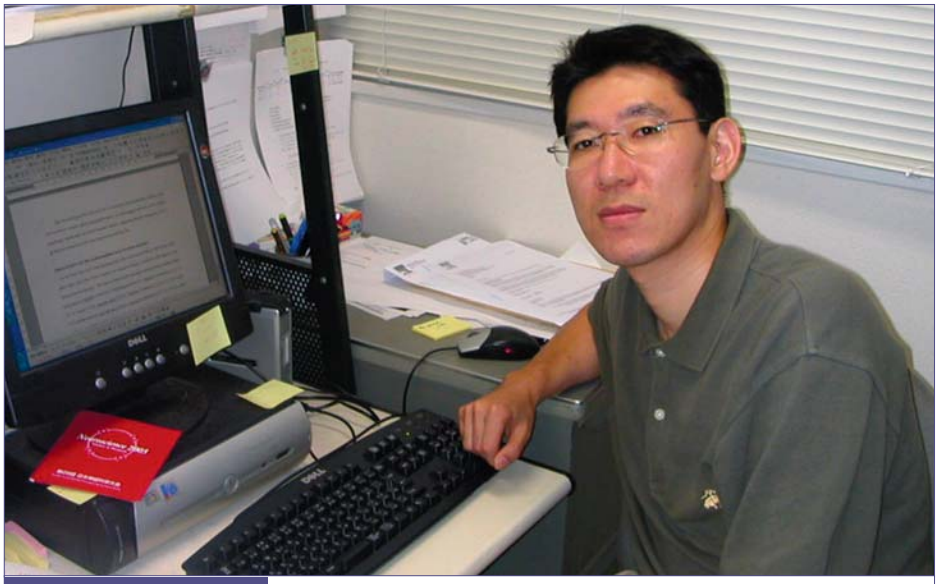
In postmortem samples, like those they used in the Harvard study, pH is significantly lower in bipolar disorder patients. They reported that in their paper. So, the down-regulation of mitochondria-related gene expression in their study may be due to the pH to some extent.

Up-regulation of Genes in Unmedicated Bipolar Patients

Hashimoto: You mentioned that if you take out the drug effect, gene expression may be in fact up-regulated. What kind of mitochondria-related genes are up-regulated?

Kato: We reanalyzed the microarray data we generated and identified up-regulated genes. The only gene that we are sure is up-regulated is LARS2.

LARS2 encodes a modifying enzyme for leucine transfer RNA. So we hypothesized that there might be a mutation in the mitochondrial leucine transfer RNA. When we analyzed cells with mutations in the mitochondrial leucine transfer RNA, we found that the LARS2 level was higher than normal. When we went after the leucine transfer RNA mutations, we found the famous 3243 mutation was present, albeit in levels less than one percent, in patients with elevated LARS2.



Ryota Hashimoto is section chief in the Department of Mental Disorder Research at the National Institute of Neuroscience at the National Center of Neurology and Psychiatry in Tokyo, Japan. He received his Ph.D. in clinical neuroscience and his M.D. from Osaka University School of Medicine. His laboratory's research focuses on the pathophysiology underlying schizophrenia, the genetic basis of psychiatric disorders and the molecular mechanism of action of psychiatric drugs.

This 3243 mutation is associated with mitochondrial diseases with a variety of physical symptoms, including mitochondrial encephalomyopathy, so it is possible that a low-level mutation in the brain might lead to mental symptoms.

Hashimoto: What about sporadic bipolar disorder patients or healthy subjects?

Kato: You can't detect the mutation in healthy subjects. However, we analyzed only 14 samples, so we have to look at a larger number and evaluate the prevalence in the general population. Among the 60 Stanley Foundation samples we used, two samples with bipolar disorder, one with schizophrenia contained this mutation.

Hashimoto: What is the detection limit of this mutation?

Kato: The minimum detection limit is 0.25 percent, so anything below 0.25 percent was regarded as absence of the mutation.

Mitochondrial Polymorphisms in Bipolar Patients

Hashimoto: Could you elaborate a little more on what kind of polymorphisms are in the mitochondrial DNA of these patients and how those changes are thought to be linked to disease?

Kato: Based on his observation that many hereditary cases of bipolar disorder are inherited maternally, Francis McMahon, then at Johns Hopkins University, analyzed the mitochondrial DNA from the probands of maternally inherited bipolar disorder families.

In contrast, our group analyzed the mitochondrial DNA from all the bipolar disorder patients.

McMahon reported that four variants showed association with bipolar disorder, one of which results in an amino acid substitution. That variant is called the 10398 variant. We analyzed variants associated with bipolar disorder with a different method, and also came up with this 10398 variant. In both studies, this single-nucleotide substitution, from G to A, is a risk factor.

Hashimoto: So two independent studies showed that A is a risk factor?

Kato: McMahon concluded that there is no association, because the difference was not statistically significant after correction of multiple testing. However, among the Japanese population, G is more prevalent, Caucasians are more likely to have A at that position, so there is a racial difference, but both our group and McMahon's came to the same finding.

However, when we did follow-up studies with more Japanese samples, we couldn't replicate our finding. This suggests that we cannot always come to the same conclusion with any sample set.

Hashimoto: Do you know anything about how this variant affects function?

Kato: In our study with lymphocytes, we found that the mitochondrial calcium concentration is different with this variant. Calcium levels are higher in patients with the A variant. We don't have too many details yet. At this point, we think there is a change in calcium metabolism.

Hashimoto: Are there any other mitochondrial mutations associated with bipolar disorder?

Kato: Yes, there are. Mitochondrial diseases have very characteristic symptoms, such as blepharoptosis, cardiac myopathy and muscle weakness. So we selected patients with these symptoms and sequenced the entire mitochondrial genome. Based on this study, we reported that a variant called 3644 might be associated with bipolar disorder. It's hard to determine whether this is really a variant or a mutation, but it appears to be specific to less than one percent of the Japanese population.

> "In medication-free patients, some of the mitochondria-related gene expression was actually up-regulated, so that was contrary to the Harvard findings."

Hashimoto: Did you observe any functional difference with this mutation or variant?

Kato: We know the mitochondrial membrane potential decreases with this variant.

The Potential of Mitochondrial Resequencing Microarrays

Hashimoto: When you are studying variants, do you usually sequence the entire mitochondrial genome? That sounds very costly and time-consuming. How long does it take?

Kato: We directly sequence the PCR products. It takes more than half a year to sequence several dozen samples. I admit it is time-consuming and labor-intensive. I hear mitochondrial resequencing arrays will be available in the near future, so we are looking forward to that and hoping it will make our experiments faster and more cost-effective.

Hashimoto: No doubt the resequencing arrays will expedite the process compared to the traditional direct sequencing method. What about the quality, though? For instance, mitochondrial variants differ among organs. With that in mind, what are your thoughts on the array technology?

Kato: As far as I know, this array can detect heteroplasmy at as little as 2 percent. Some mutations are present at only one percent in the brain. So we'll have to see whether we can detect them using arrays or not.

Hashimoto: How sensitive is direct sequencing to heteroplasmy?

Kato: It may not be better than the array. You can't usually detect heteroplasmy if it's less than a few dozen percent. In order to detect mutations less than one percent, we used to use a different method—the PNA, protein nucleic acid method. So detecting very low levels of heteroplasmy, whether by sequencing or microarray, is not an easy task. Moving to resequencing microarrays would mostly be a matter of saving time and labor. I'm looking forward to that.

FOR MORE INFORMATION

Contacts

■Tadafumi Kato, M.D., Ph.D.
Laboratory Head
Laboratory for Molecular Dyanamics of Mental Disorders
Aging and Psychiatric Research Group
RIKEN Brain Science Institute
2-1 Hirosawa
Wako
Saitama 351-0198
Japan
kato@brain.riken.jp

■Ryota Hashimoto, M.D., Ph.D.
Section Chief
Department of Mental Disorder Research
National Institute of Neuroscience
National Center of Neurology and Psychiatry, Japan
4- 1- 1, Ogawahigashicho
Kodaira- city
Tokyo 187-8502
Japan
rhashimo@ncnp.go.jp

Organizations

■RIKEN Brain Science Institute
http://www.brain.riken.go.jp/

■National Center of Neurology and Psychiatry, Japan
http://www.ncnp.go.jp/link_e.html

Further Reading

■Munakata K, Iwamoto K, Bundo M, Kato T. Mitochondrial DNA 3243A>G mutation and increased expression of LARS2 gene in the brains of patients with bipolar disorder and schizophrenia. *Biol Psychiatry* 2005; 57(5):525-32.

■Iwamoto K, Bundo M, Kato T. Altered expression of mitochondria-related genes in postmortem brains of patients with bipolar disorder or schizophrenia, as revealed by large-scale DNA microarray analysis. *Hum Mol Genet* 2005; 14(2):241-53.

■Kato T. Mitochondrial dysfunction in bipolar disorder: from 31P-magnetic resonance spectroscopic findings to their molecular mechanisms. *International Review of Neurobiology* 2005; 63: 21-40. (Review)

■Hashimoto R, Okada T, Kato T, Kosuga A, Tatsumi M, Kamijima K, Kunugi H. The breakpoint cluster region gene on chromosome 22q11 is associated with bipolar disorder. *Biol Psychiatry* 2005; 57(10):1097-102.

People

■Francis McMahon
National Institute of Mental Health
http://intramural.nimh.nih.gov/research/pi/pi_mcmahon_f.html

UPCOMING SYMPOSIA



Peter J. van der Spek, Ph.D.
Erasmus MC, The Netherlands

Dr. van der Spek will talk about the ways that genomic technologies are changing the view of molecular pathways from static snapshots to dynamic networks



David Wong, D.M.D., D.M.Sc.
University of California, Los Angeles

Dr. Wong will discuss using expression profiles of the salivary transcriptome as diagnostics for major human diseases, including oral squamous cell carcinoma



Linda Strausbaugh, Ph.D.
University of Connecticut

Dr. Strausbaugh will discuss her research investigating the suitability of high-density mitochondrial resequencing arrays in forensic applications



Anindya Dutta, M.D., Ph.D.
University of Virginia

Dr. Dutta will talk about his discovery of pan-S phase replication in cancer cells using high-resolution tiling microarrays for unbiased mapping of the human genome



Hubertus (Bert) J.M. Smeets, Ph.D.
Maastricht University

Dr. Smeets will present a "bedside to genome to bedside" approach to studying and treating mitochondrial disease

For more information on dates and times for AMB Symposia, go to "Upcoming Events" at ***www.microarraybulletin.com***

Don't miss these in-depth interviews in recent issues of AMB. Subscribe today.

Special HapMap Issue

The Broad Institute's David Altshuler, Stacey Gabriel & Mark Daly discuss the lessons learned from the HapMap Project, high-throughput genotyping and the data analysis infrastructure required to search for genes influencing common diseases.



"The possibility of making important discoveries through comprehensive genotyping is tremendously exciting. But the potential to get into trouble by overanalyzing the data is a real danger."

Baylor's Richard Gibbs talks about multiplexed genotyping assays used for the HapMap Project and performing other large-scale epilepsy and cancer genome studies.

"I see developments in microarrays in the past year or so as being analogous to the developments in sequencing we saw 10 years ago that brought us down the road of genomics so quickly and so successfully."



Special Stem Cell Issue



"Everybody is hoping to use these cells in some therapeutic protocol one day, so we need to know exactly what the cells are, what kinds of genetic changes they have, and whether those changes make any biological difference or not."

Johns Hopkins Medical Center's Aravinda Chakravarti on finding mutations in federally-approved embryonic stem cell lines.

"Our recent results are just the tip of the iceberg. There are many, many, many more molecules playing interesting roles in stem cells."



Princeton's Ihor Lemischka on using microarrays in embryonic stem cell research.

www.microarraybulletin.com

Request your complimentary copy of the Special HapMap Issue or the Special Stem Cell Issue at reprints@microarraybulletin.com

WWW.MICROARRAYBULLETIN.COM

LOG ON FOR A FREE SUBSCRIPTION

AFFYMETRIX

MICROARRAY BULLETIN

Affymetrix Microarray Bulletin (AMB) is a quarterly newsletter published by Affymetrix Inc. that highlights the work of thought leaders in molecular biology and microarray research from around the world. AMB is a scientist-to-scientist publication, written for researchers, statisticians, and software and information technology engineers. Our goal is to create a forum where the research community can discuss its work and share methods, tools, insights and information.

The online version of this newsletter can be accessed at http://www.microarraybulletin.com

Wes Conard, *Editor-in-Chief*
wes_conard@affymetrix.com

Tommy Broudy, *Managing Editor*
thomas_broudy@affymetrix.com

Rachel Shreter, *Editor*
rachel_shreter@affymetrix.com

Stacey Ryder, *Associate Editor*
Kamalia Dam, *Associate Editor*

Affymetrix Inc.
3420 Central Expressway
Santa Clara, CA 95051
408.731.5000

AFFYMETRIX®