AFFYMETRIX
MICROARRAY
BULLETIN

EXPRESSION
VOLUME 4 ■ ISSUE 2
AMB REPRINT

www.microarraybulletin.com
*Summer 2008*

# EURASNET Consortium Evaluates the Performance of Different Microarrays for Studying Alternative Splicing

## Chris Smith of Cambridge University, Tyson Clark of Affymetrix and Melissa Cline of UC Santa Cruz discuss EURASNET's approach for comparing commercial microarrays for investigating alternative splicing

By Stacey Ryder

**Cambridge, UK, August 1, 2008** — A European consortium of nearly 40 laboratories has conducted a study which identified GeneChip® Human Exon 1.0 ST Arrays as an effective tool for studying alternative splicing—a process that results in multiple isoforms of a single gene, and a major source of protein diversity.

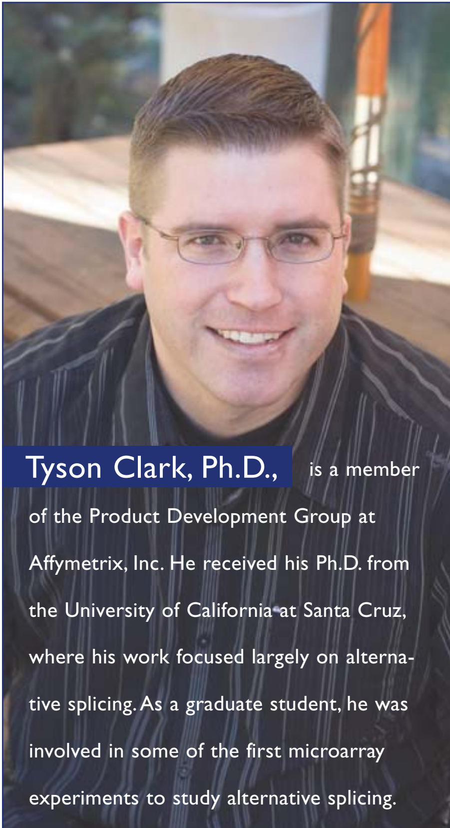
The study compared the ability of microarray platforms to investigate alternative splicing in transcripts regulated by the splicing factor polypyrimidine tract binding (PTB) protein, and its paralog, nPTB. Researchers hope that uncovering the mechanisms which regulate alternative splicing will lead to a better understanding of human disease.

The consortium, dubbed EURASNET (European Commission-funded Network of Excellence), found that GeneChip Human Exon 1.0 ST Arrays outperformed other splice-sensitive array platforms in a study which investigated alternative splicing in cancer (HeLa) cells. GeneChip Exon Arrays, which include approximately 1.4 million probe sets, allow for complementary high-resolution gene expression and alternative splicing analysis. By using these genome-wide arrays, EURASNET researchers were readily able to detect and confirm a total of 38 splicing events.

"Exon arrays are global," said Cambridge University's Dr. Chris Smith, a EURASNET member whose lab designed the study. "They feature a huge number of probe sets…ranging from the very well-annotated exons to exons for which there is only a low level of confidence in the prediction. These arrays provide the necessary power for discovering new alternative splicing events."

After the data was collected by the EURASNET team, scientists at Affymetrix grouped the resulting hits by confidence levels using a simple algorithm that normalizes exon signal to the level of gene expression. Exon-level signal intensities reflect a combination of the number of transcripts made from a gene and the frequency in which a particular exon appears in those transcripts. Included in the ordered list of hits were 32 high-confidence splicing events. Smith's group

**Chris Smith, Ph.D.,** is a professor in the Department of Biochemistry at the University of Cambridge. For the past 15 years, his group has focused on understanding the molecular mechanisms behind alternative mRNA splicing. Dr. Smith is also a member of the EURASNET consortium, a group of almost 40 laboratories across Europe that is combining resources to research alternative splicing on a larger scale. Dr. Smith and his laboratory were recently involved in a EURASNET study to determine the most appropriate commercial microarray platform for investigating alternative splicing.



**Tyson Clark, Ph.D.,** is a member of the Product Development Group at Affymetrix, Inc. He received his Ph.D. from the University of California at Santa Cruz, where his work focused largely on alternative splicing. As a graduate student, he was involved in some of the first microarray experiments to study alternative splicing.

used reverse transcription followed by real-time PCR to validate 95 percent of those events.

Dr. Smith and Tyson Clark, of Affymetrix' Product Development group, recently spoke to Melissa Cline, a Senior Informatics Postdoctoral Fellow at the University of California at Santa Cruz, to discuss the study and future directions for alternative splicing research. The three discussed:

- Advantages of the GeneChip Human Exon 1.0 ST Array
- Understanding the functional relevance of alternative splicing
- Future directions of alternative splicing research

## Study design

**Cline:** Dr. Smith, why was EURASNET interested in comparing microarrays, and how was your laboratory's experiment chosen as the model?

**Smith:** I am involved in one of the EURASNET work programs that focuses on finding global methods for analyzing alternative splicing. Within that program, there is a project aimed at identifying the best microarray platform for investigating alternative splicing. The consortium wanted to come up with a model experiment that could be used to compare different array platforms.

We had been looking at the splicing factor, PTB protein, for a number of years. We had performed experiments in which we knocked down PTB and its paralog, nPTB, in HeLa cells and then used quantitative proteomics to identify up- and down-regulated spots. Our hypothesis was that we should be able to pick up examples where alternative splicing had switched. So far we've identified 25 PTB-regulated events by this approach.

The consortium decided our experiment would be a good model experiment because we were using a completely independent technology platform to look for changes in alternative splicing, and we were investigating a splicing factor for which a large number of targets were known.

**Cline:** What array platform did you test? What were its advantages?

**Smith:** We quickly focused on the Human Exon 1.0 ST Array from Affymetrix, as during the validation exercise, it became evident to us that it was out-performing the alternatives.

In addition, exon arrays have the major advantage of being global. They feature a huge number of probe sets—approximately 1.4 million—ranging from the very well-annotated exons to exons for which there is only a low level of confidence in the prediction. These arrays provide the necessary power for discovering new alternative splicing events.

My lab carried out the experiments and supplied biological triplicate samples of both the knockdown and the control for array analysis without revealing what the experiment was, except that it was a knockdown of a splicing factor.

**Cline:** Dr. Clark, you assembled a set of predictions that were grouped by confidence level. What did you use to assess the confidence?

**Clark:** On the exon array, the signal intensity for any given exon is the combination of transcription and splicing—how many transcripts are being made from the gene and how often a particular exon is included in those transcripts.

To tease out splicing information, we use methods that normalize out those transcriptional differences to observe individual probe sets relative to an estimate of the gene expression level. Changes in the gene-level-normalized intensities suggest a change in exon inclusion rates and thus differential splicing.

The confidence levels were assigned based on a combination of the statistical significance of the change in inclusion rate and an examination of the genomic context by observing the underlying probe set intensities superimposed onto the genome.

**Smith:** The high-confidence predictions that Tyson gave us included 32 events. We validated those at about 95 percent level. The results were much better than I had expected.

## Understanding alternative splicing

**Cline:** When you observe an alternative splicing event, how can you start putting that into a functional context?

**Smith:** We are not all that interested in identifying individual alternative splicing events. We want to have a large group of co-regulated events, then get into the informatics of trying to decipher how the gene sequence makes one particular exon responsive to a particular splicing regulator.

This was one of the reasons our experiment was used as the model experiment. We know a lot about the optimal binding sequences for this protein. So when it comes to unraveling direct from indirect effects, we can look at the genomic context and whether a binding site is likely to be there.

If you conducted an experiment like ours, but didn't know what the binding site was, once you had a set of co-regulated events, you could do the informatics and try to identify motif enrichment in and around those co-regulated exons.

In fact, one of the other EURASNET members who has been collaborating with us has already made some surprising discoveries.

For instance, he has found significant enrichment of particular motifs within the exons. Before this discovery, people thought the binding sites of this protein were usually in the upstream intron or in the flanking intron. But having this large data set, we can clearly see that the exons themselves are enriched in binding sites for this protein.

Again, the rather bold, long-term aim here is to be able to use the genomic sequence, to determine the tissue specificity of a particular splicing event or whether the event is inducible in response to particular signaling pathways or affected by particular drugs.

I think we are a long way from that. But by having a very reliable array platform that can quantitatively measure changes in splicing, we can either look for tissue-specific changes or changes in response to knocking out a factor or adding a drug. Then we hope to go back and do the informatics in order to determine the code in the primary sequence that makes particular exons sensitive to these changes.

**Cline:** Some researchers say that alternative splicing is really nothing more than another mechanism to turn off protein production. Would either of you care to comment on this assessment?

**Smith:** It's true that some splicing events are on and off switches. But there are many examples of where that is not the case. Just recently there have been papers on the Drosophila Down syndrome cell adhesion molecule, DSCAM. It has more than 38,000 isoforms. These papers show that each isoform seems to have a functionally important role.

**Clark:** In terms of evolution, it's interesting to think about why a mechanism that involves splicing for turning off expression of a protein would be used rather than a more straightforward path of just turning off transcription of that gene. It's interesting to imagine how these sorts of different mechanisms evolved. Obviously there is some reason why the cell does it, but that is still a mystery.

"It's interesting to imagine how these sorts of different mechanisms evolved. Obviously there is some reason why the cell does it, but that is still a mystery."

**Smith:** My take on that would be that if you're an RNA binding protein or a gene for an RNA binding protein, the best way you can control yourself or other members of your family or neighbors is at a post-transcriptional level. You are not going to be much good at binding to DNA and affecting transcription, but it makes sense to feedback at the RNA level.

It goes beyond splicing. There are a lot of RNA binding proteins that autoregulate. Poly-A binding protein binds to a poly-A tract in its 5' UTR and it halts translation. I think feedback regulation is widely used in many biological contexts.

**Cline:** What advice would each of you give to researchers who are just starting to study alternative splicing using microarrays?

**Clark:** One microarray experiment is going to generate a lot of data. While the experiment itself takes just a couple of days, you might spend months or years following up on the data. With that in mind, carefully planning your experiment upfront will save you lots of time in the long run.

For example, adding biological replicates can improve your ability to detect alternative splicing with higher confidence. Also, a clear understanding of the limitations of your analysis method will make it easier to identify potential false positives.

### Looking to the future

**Cline:** What technological developments might help advance our understanding of alternative splicing?

**Clark:** Analysis methods for extracting alternative splicing information from microarray data are still in their infancy. Being able to run one of these experiments and use an analysis method similar to what we are able to do today for studying gene expression would be great. This would give us a list of hits that we are confident represents real alternative splicing events. Then, being able to take that information and figure out which isoforms are being expressed from the gene would be terrific. The methods for doing this are still in development.

**Smith:** Ideally, you want an array platform to generate a list of predictions of alternative splicing events in which you have high confidence but that is also as inclusive as possible. The data analysis procedure that Tyson implemented was ideal in the first instance, because it generated a "high-confidence" set of alternative splicing predictions with a very low false positive rate. However, at the moment there

**Melissa Cline, Ph.D.,** is a senior informatics postdoctoral fellow in the laboratory of Manny Ares at the University of California at Santa Cruz. She previously worked as a postdoctoral researcher on the Cytoscape Project at the Pasteur Institute and was a staff scientist at Affymetrix, Inc., where she helped develop a prototype chip set for the GeneChip® Human Exon Array.

is a trade-off. The need to minimize the false positives in a highly curated data set also leads to false negatives. We are losing a lot of information along the way. If you are looking for events where there is overlapping regulation from individual factors, you may find that you completely miss the overlap among the false negatives. So there is plenty of scope for refining the analysis procedures to reduce the extent of the trade-off.

**Cline:** What future plans does EURASNET have for this technology?

**Smith:** As a result of this pilot experiment, I expect to see more groups interested in using similar approaches to look at targets of splicing factors. In fact, one of the other groups has already published an exon array analysis of hnRNP-L-regulated alternative splicing events. We would also like to start looking at co-regulation—looking for events that are regulated by combinations of factors. It will be interesting to see how well exon arrays can investigate this in view of the trade-off we spoke about, because this type of experiment would need an analysis approach with a low level of false negatives.

We've been talking about a program to look at more global methods for analyzing alternative splicing. There are various other programs using chemical biology to try to affect splicing or to find small compounds that would free spliceosomes at particular steps or to look for small molecules that can affect particular splicing patterns and might be developed as therapeutics.

There is also a program to look at the interaction between transcription and alternative splicing. That is clearly very important. It'll be interesting to see to what extents regulation at the transcription level and elongation and chromatin modification are the direct effectors of alternative splicing. There are clear examples where that occurs.

It's very likely that array technology will be applicable to some of these other programs, too.

**Cline:** Dr. Clark, how do you anticipate alternative splicing technology being used in the next few years?

**Clark:** There are so many different potential uses for it. There is a lot of evidence that mistakes in alternative splicing can lead to human disease. I think a better understanding of that is in our future, as well as just a more general understanding of how alternative splicing is regulated. A lot of the work that Chris is doing is going to take us down that path.

**Smith:** That's a very good point. There is increasing evidence of the association of mis-splicing or changes in alternative splicing being either a direct cause of disease or associated with particular diseases. EURASNET has a program dedicated to mis-regulated alternative splicing and disease. Having a platform that is able to take a complete look at the splicing landscape between normal and disease samples would be very useful, and I am sure we will see many developments in this area in years to come.


AFFYMETRIX MICROARRAY BULLETIN

**Editorial Staff**

Rachel Steger, *Founding Editor*
rachel_steger@affymetrix.com
Bronwyn Barnett, *Managing Editor*
bronwyn_barnett@affymetrix.com
Sven Brown, *Senior Editor*
Stephen Antonopoulos, *Photo Editor*
Stacey Ryder, *Associate Editor*
Sabrina Banwait, *Associate Editor*
Marva Maida, *Graphic Designer*
Jim Butler, *Graphic Designer*


## FOR MORE INFORMATION

### Contacts


- Chris Smith, Ph.D.
Department of Biochemistry
University of Cambridge
80 Tennis Court Road
Cambridge
CB2 1GA
United Kingdom
cwjs1@mole.bio.cam.ac.uk

- Tyson Clark, Ph.D.
Senior Scientist
Affymetrix, Inc.
3420 Central Expressway
Santa Clara, CA 95051
tyson_clark@affymetrix.com

- Melissa Cline, Ph.D.
Senior Informatics Postdoctoral Fellow
Department of Molecular, Cell and Developmental Biology
Center for Molecular Biology of RNA
University of California at Santa Cruz
Santa Cruz, CA 95064
cline@biology.ucsc.edu


### Companies

- Affymetrix, Inc. – www.affymetrix.com
- Partek – www.partek.com

### Organizations

- EURASNET – www.eurasnet.info
- Cambridge University – www.cam.ac.uk

### People

- Gil Ast - www.eurasnet.info/gil_ast.shtml
- Ben Blencowe - www.utoronto.ca/intron/blencowe.html
- Juan Valcárcel - www.eurasnet.info/juan_valcarcel.shtml
- Stefan Stamm - www.eurasnet.info/stefan_stamm.shtml
- Didier Auboeuf - www.eurasnet.info/didier_auboeuf.shtml
- Albrecht Bindereif - www.eurasnet.info/albrecht_bindereif.shtml

### Further Reading


- Spellman R., Llorian M., Smith C. W. Crossregulation and functional redundancy between the splicing regulator PTB and its paralogs nPTB and ROD1. *Molecular Cell* 27(3):420-34 (2007).
- Spellman R., Smith C. W. Novel modes of splicing repression by PTB. *Trends in Biochemical Sciences* 31(2):73-6 (2006).
- Hung, L.-H., Heiner, *et al.* Diverse roles of hnRNP L in mammalian mRNA processing: a combined microarray and RNAi analysis. *RNA* (14) 284-296 (2008).
- Hattori D., *et al.* Dscam diversity is essential for neuronal wiring and self-recognition. *Nature* 13;449(7159):223-7 (2007).