

Discovery of Novel Splice Variations Improves Glial Tumor Classification

Erasmus Medical Center's Pim French and Justine Peeters talk with Noam Shomron at MIT about using exon arrays to study alternative splicing in glioma

By Megha Satyanarayana

Researchers at Erasmus Medical Center in the Netherlands have discovered expression profiles with distinct splice variants that more accurately classify two forms of glial cancers.

They made the splicing discovery using new microarrays that analyze over 1.4 million probe sets spanning all known and predicted exons; their findings may help clinicians more accurately diagnose the multiple classes and variable prognoses of brain cancer.

The team, led by Pim French, a post-

doctoral fellow in the department of neurology at the Erasmus Medical Center and the Josephine Nefkens Institute, tested 28 glioblastomas, 20 oligodendrogliomas and 6 control brain samples with the new GeneChip® Human Exon 1.0 ST Array. They discovered that both types of tumors have a distinct pattern of alternative splicing in addition to a distinct gene expression profile.

French believes that the development of a splicing-based expression profile for the different types of glial

tumors will provide a further level of certainty to histological analyses and may identify causative genetic changes.

"I would like to run a chip for every tumor that comes in, because I think you really need expression data to know what molecular subgroup a tumor is in to better aid the pathologist," said French. "Cancer is complex, and not caused by a single gene mutation. So, you really do have to have a global view of what is happening in the cells in order to get a good picture of disease



Pim French and Justine Peeters of Erasmus Medical Center

Pim French is a post-doctoral

fellow at the Josephine Nefkens Institute and department of neurology at the Erasmus Medical Center in Rotterdam, Netherlands. He received his Ph.D. at Erasmus Medical Center in cell biology, and completed a post-doctoral fellowship at the National Institute for Medical Research in London. His main focus is using microarray technology to study neurological cancers and disorders, to aid pathology and to find causative genetic mutations.



and ultimately this involves studying splicing isoforms.”

Justine Peeters led the bioinformatics effort to analyze the glioma exon array data. She is finishing her Ph.D. studies in the analysis of microarray data, with further applications to cancer, in the laboratory of Prof. Peter van der Spek at Erasmus. Her thesis project will provide an educational resource for performing expression, SNP and exon microarray experiments, from planning the experiment to analyzing the resulting data. French and Peeters’ work has recently been submitted for publication at *PNAS*.

“These arrays give us just a whole other level of analysis, really,” said Peeters. “Instead of just having your transcript measured with expression arrays, we can look at the differentially represented exons and then translate this information into changes in the transcripts. So, even though you have the same overall information, exon arrays are more of a true measure of gene expression.”

French and Peeters recently spoke to Noam Shomron, a post-doctoral fellow in the laboratory of Chris Burge at MIT about new applications for studying alternative splicing at the whole genome level. One of Noam’s research fields focuses on using novel computational

tools for studying alternative splicing events in general and for identifying unannotated alternatively spliced exons.

The three discussed:

- The advantages of exon arrays over other expression arrays in classifying gliomas
- Testing, troubleshooting and validating exon array data
- The potential application of exon arrays to clinical diagnosis and prognosis

Exon arrays vs. other expression arrays in glioma diagnosis

Shomron: Gliomas are noted for their aberrations in alternative splicing. How were you studying glioma gene expression before and could you describe your current approach?

French: Originally, we had been studying global gene expression in gliomas using GeneChip® Human Genome U133 and then using that information to supplement our pathological diagnoses. We could identify molecular subgroups of gliomas and classify them based on their expression profile.

Our current approach is to use exon arrays to look at the contribution of splicing to tumor biology. We strongly believe that exon arrays will help us find causative genetic changes in cancer. For example, exon arrays readily detect

expression of a pathological splice variant of EGFR. Exon arrays should also lead to the identification of fusion genes because the exon probes would cover the joined breakpoints. Such breakpoints are hard to find using the old U133 arrays.

Shomron: So, have you completely shifted to exon arrays or are you still using both?

French: We are still using both because we don’t want to change protocols for ongoing experiments. Furthermore, most of the gene expression information we see is in transcripts rather than in differentially expressed splice variants. So, for identifying molecular subgroups of gliomas, we use expression arrays, but I think we will shift towards using exon arrays, because in the end, you just get more data from them.

Shomron: But can you successfully retrieve expression data from the exon array as well?

French: Yes, you can and in fact, quite easily and very well.

Peeters: The probe sets used in the U133 Plus 2.0 arrays are 3-prime biased, so, they’re not really a true presentation of what is really happening with transcripts. If you have a splice variant influencing the biology of what you are looking at, this will most likely be missed with a selection of probes interrogating only the 3’ end of the gene. This is where exon arrays are more powerful in reflecting the true biology, as every exon whether it is located in the 3 prime or 5 prime end of the gene is represented in the array. You can utilize these arrays in looking at an exon by exon approach or translate this information into transcript intensities.

Shomron: So, when do you think you will completely shift to using exon arrays?

French: It mainly depends on funding for running such a large number of arrays. Also, it is slightly more difficult to handle exon arrays than the U133 Plus 2.0 arrays. So, there’s a little more training involved.

Peeters: And previously you were dealing with 54,000 probe sets. Now we have 1.4 million probe sets. It creates a

problem for computational analysis—available computer memory and available software.

Shomron: Can you elaborate on any overlapping clusters and correlations you found from using these two array platforms?

French: The two platforms correlate extremely well. The subgroups that we get with U133 Plus 2.0 arrays are identical to the subgroups that we get with the exon arrays. That was an important step in saying how well they work. Here's an example. We have two oligodendroglial tumor samples that cluster with glioblastoma samples on the exon arrays — they are outliers. On the U133 Plus 2.0 arrays, we ran the same two samples and we found that one of the two samples also clustered with glioblastomas as opposed to oligodendromas. The subgroups clustered very well and some of the same samples are outliers in both studies.

Testing and validating glioma expression studies

Shomron: In the exon array, how many positive controls of known splicing isoforms did you look at in order to gain some confidence that your arrays worked well before you proceeded for further analysis?

Peeters: We looked at EGFR variant III, which has been seen in glioblastomas containing EGFR amplification. *EGFR*, of course, is known to be differentially spliced in glioma. We also looked at other causative splice isoforms in other tissues such as breast cancer. In developing an algorithm to detect splice variants, we used this as a positive control as a guide to look at step-by-step filtering procedures. As part of this algorithm, we had to filter out exons which have a non-linear relationship with transcript as well as non-informative exons which are also represented on the chip. Signal from such exons can skew your results and lead to many false positives.

Shomron: In your studies, you note that there was a correlation in the subgroups defined through expression analysis and through histology. Was that surprising?

Peeters: Well no. Gene expression

basically directs morphology and thus, histology, but a pathologist cannot see what gene expression tells us. Therefore, the differences are quite informative for follow-up and treatment. It's possible that the underlying transcript I am seeing through expression studies can also pre-

We design two primer pairs, both of which contain one primer in the putative novel exon and one in a known exon. In designing the primers, we make sure that there is a large intronic region in between, to avoid amplifying genomic DNA. We run the assay and get bands

“Looking at the transcript, you can actually predict, even though they may look the same histologically, how the patient can respond to treatment or what type of treatment you can actually give.”

dict drug response. Looking at the transcript, you can actually predict, even though they may look the same histologically, how the patient can respond to treatment or what type of treatment you can actually give.

Shomron: Do you always perform histology in addition to arrays?

French: Not all samples are of sufficient quality to run on our arrays. Of course, for samples that are good quality, we do the comparison with histology.

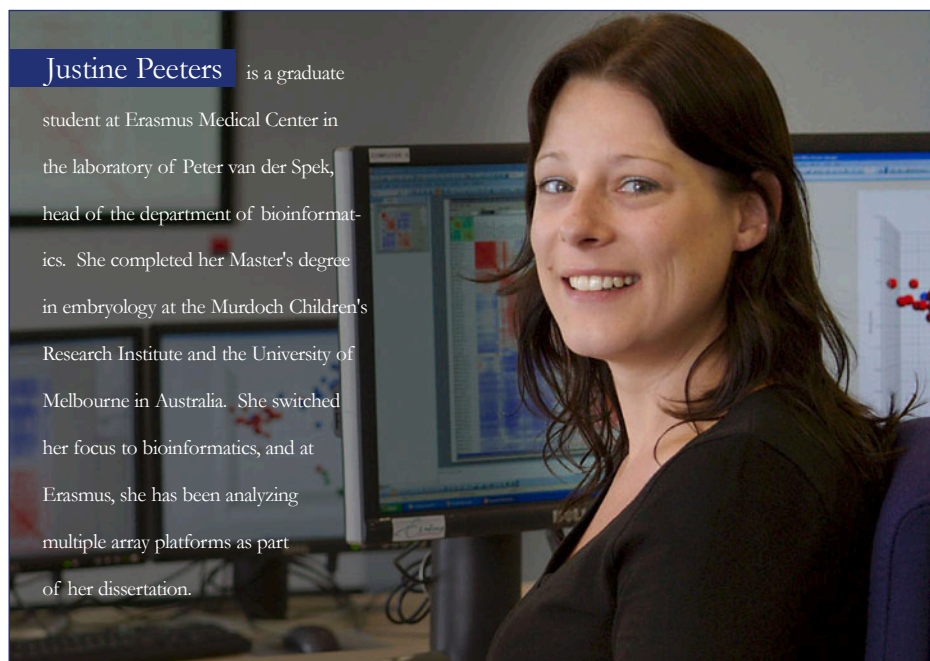
Shomron: You've identified hundreds of novel exons in gliomas that are not supported by current databases. Have you confirmed them through RT-PCR?

French: Yes. This is our protocol:

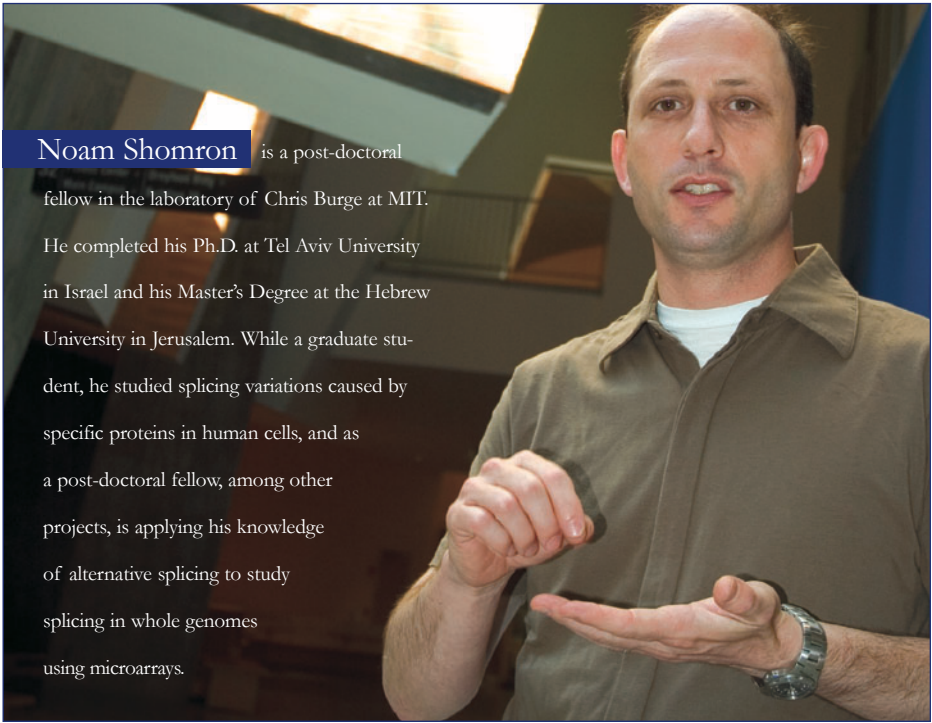
of about the expected size. We then sequence the bands and indeed, you get the product that you would expect. With the sequence, we refer to genomic databases and we often find consensus splice acceptor and donor sites.

Shomron: And what is the success rate? What is the false positive and negative of those RT-PCRs?

French: We have identified around 700, of which about 80 percent truly represent novel exons. We can confirm something like 60 to 70 percent of candidates. The 30-40 percent we cannot confirm may be due to incorrect primer design because you might have a completely differently spliced transcript. This



Justine Peeters is a graduate student at Erasmus Medical Center in the laboratory of Peter van der Spek, head of the department of bioinformatics. She completed her Master's degree in embryology at the Murdoch Children's Research Institute and the University of Melbourne in Australia. She switched her focus to bioinformatics, and at Erasmus, she has been analyzing multiple array platforms as part of her dissertation.



Noam Shomron is a post-doctoral fellow in the laboratory of Chris Burge at MIT. He completed his Ph.D. at Tel Aviv University in Israel and his Master's Degree at the Hebrew University in Jerusalem. While a graduate student, he studied splicing variations caused by specific proteins in human cells, and as a post-doctoral fellow, among other projects, is applying his knowledge of alternative splicing to study splicing in whole genomes using microarrays.

means that the number of novel exons may even be higher than we can confirm by RT-PCR.

Shomron: That's terrific. Now, let's talk technically about non-linear exons — exons that are expressed at levels outside the linear detection range of the arrays, or exons that bear no relation to its putative transcript. You mentioned that after comparing the exon array results and RT-PCR you recalculated PAC, the probability of the splicing events, based on metaprobe sets lacking “non-linear” exons. How much data was discarded as “non-linear”?

French: If you look at the 1.4 million probe sets that are on the array, only a few percent are nonlinear or nonspecific exons. But, if you try to identify regulated splice variants, then you narrow down your search to a few thousand candidate exons. So the smaller set of candidates are now biased with these nonlinear exons, and you end up with quite a high percentage — up to 80 to 90 percent in first pass, without any filtering steps.

Shomron: Then does your filtering process actually retain a large number of nonlinear exons?

French: Yes, we find it difficult to completely filter them out. But in identifying regular splice variants you ask that every probe set within the gene of inter-

est behaves in a completely linear fashion. So if the gene is upregulated two-fold in one sample or subgroup, all the individual exons of that gene should be upregulated exactly two-fold on the array. Any probe set that does not show this two-fold upregulation will be identified as a false positive candidate.

Peeters: And also, in identifying splice variants, you have to translate all of your exon probe sets into your transcript signal intensity in order to calculate differences in expression. This is a whole additional level of analysis as compared to the U133 Plus 2.0 expression arrays. What we used was an adapted correlation based PAC algorithm in our calculations of the splice variants. Including these nonlinear exons in your initial calculation of the transcripts to be utilized in the PAC algorithm disrupts the true correlation between your exon and your transcript, which complicates things.

Shomron: And, you verify them by quantitative or nonquantitative RT-PCR?

French: We use nonquantitative RT-PCR, because differences are pretty obvious. For ones that we could confirm, we have a verification success rate of about 40 percent. RT-PCR using primers that span the spliced exon gives you a semi-quantitative result anyway, because you see the ratio of inclusion and exclusion of

the spliced exon.

Shomron: I think the next stage after identifying which exons are represented is finding the difference in magnitude and the varied interplay or network between the different isoforms. That is probably something for the future after users feel confident with their exon arrays.

Future studies and using exon arrays for diagnostics

Shomron: Will your results from these arrays take you into follow up studies of individual genes or back to running more arrays and clustering?

Peeters: I think we would definitely use both approaches. Following up on these individual genes to look at their role in the development of the subtypes of brain tumors is definitely going to be interesting, but we also would be interested in running some more arrays and also reassessing the arrays that we have already run with alternative analysis techniques. We believe that we can identify causative changes using these exon arrays and hopefully identify more molecular subtypes or subgroups of brain tumors. We would like to be able to include some more histological subgroups and see whether again we can find differentially regulated splicing and even causative changes that cause the errant splicing event. So, yes there is a lot of information in these exon arrays that we can utilize.

Shomron: I agree that there is an advantage of looking at the global gene expression/isoform picture, rather than just one or two genes at a time. Do you think that splicing isoform clusters will eventually replace expression clusters in diagnosis and prognostics?

Peeters: Yes, I do think exon arrays will be the future of diagnostics and prognosis, because they do have a lot more information than the older expression arrays. As I have mentioned before, we can look specifically at different isoforms of transcripts that may be involved in patient response to drug treatment and the specific diagnosis of disease.

At the moment, it is fresh, new analysis for us. So, the most important thing for us is to see if we can use this platform

to identify differential expressed splice variants and be able to confirm them in the lab.

Shomron: Last question — if you could set up the ultimate diagnostic laboratory what would it look like?

French: You would use exon and SNP arrays to identify which molecular subgroup the tumor belongs to. These subgroups would aid in guiding therapy and will also give prognostic information for the patient. I also strongly believe that future therapies will focus on the molecular aberrations of tumors. Therefore, I would also like to perform an array experiment that would sequence all of the known tumor suppressors and oncogenes. This way, you know that in patient A you have genes X, Y, and Z mutated. Such molecular knowledge will

be of high importance to guide future therapies. I think that would be the ultimate diagnostic lab.

Shomron: Yes, I definitely agree with you. Use microarrays to analyze everything, collect the data, then look at part of it now, and then maybe reanalyze a few years later, when there are a few more revelations and a few more diagnostic tests and genes or mechanisms identified. An example would be regulation by microRNAs, a concept which has gained ground only in the past few years.

French: You never know what you are going to get from such large data sets, but there is a lot of information encoded in the expression data and most of it we do not really know at this point. It's amazing. We are only scratching the surface of what you can do with these arrays.

AFFYMETRIX MICROARRAY BULLETIN

Editorial Staff

Wes Conard, *Editor-in-Chief*
wes_conard@affymetrix.com
Tommy Broudy, *Managing Editor*
thomas_broudy@affymetrix.com
Rachel Shreter, *Editor*
rachel_shreter@affymetrix.com
Kamalia Dam, *Associate Editor*
Stacey Ryder, *Associate Editor*
Daniel Noble, *Copy Editor*
Marva Maida, *Contributing Designer*

FOR MORE INFORMATION

Contacts

■ Pim French, Ph.D.
Josephine Nefkens Institute
Department of Neurology
Erasmus MC
PO Box 1838
3000 DR Rotterdam
Netherlands
p.french@erasmusmc.nl

■ Justine Peeters
Department of Bioinformatics
ErasmusMC
Postbus 1738,
3000DR Rotterdam
Netherlands
j.peeters@erasmusmc.nl

■ Noam Shomron, Ph.D.
Department of Biology
Massachusetts Institute of Technology
77 Massachusetts Ave 68-217,
Cambridge, MA 02139
United States of America
nshomron@mit.edu
<http://web.mit.edu/nshomron/www>

Companies

■ Affymetrix, Inc. - <http://affymetrix.com>

Organizations

■ Erasmus Medical Center -
<http://www.erasmusmc.nl/>
■ Josephine Nefkens Institute -
<http://www.eur.nl/fgg/pathol/>
■ Massachusetts Institute of Technology -
<http://www.mit.edu>

People

■ Chris Burge, Ph.D., MIT -
<http://genes.mit.edu/chris/>
■ Peter van der Spek, Ph.D., Erasmus Medical
Center - [http://www.erasmusmc.nl/bioinformatics/
people/staff.shtml](http://www.erasmusmc.nl/bioinformatics/people/staff.shtml)

Further Reading

■ French PJ, Swagemakers SM, Nagel JH, Kouwenhoven MC, Brouwer E, van der Spek P, Luidert TM, Kros JM, van den Bent MJ, Sillevius Smitt PA. Gene expression profiles Associated with Treatment Response in Oligodendrogliomas. *Cancer Res.* 2005 Dec 15;65(24):11335-44.
■ Peeters JK, Van der Spek PJ. Growing applications and advancements in microarray technology and analysis tools. *Cell Biochem Biophys.* 2005;43(1):149-66.