Author's response to reviews

Title: Specific Gene Expression Profiles and Unique Chromosomal Abnormalities are Associated with Regressing Tumors Among Infants with Dissiminated Neuroblastoma.

Authors:

Cinzia Lavarino (clavarino@fsjd.org)
Nai-Kong V Cheung (cheungn@mskcc.org)
Idoia Garcia (igarcia@fsjd.org)
Gema Domenech (gema.domenech@h1.org.es)
Carmen de Torres (cdetorres@hsjdbcn.org)
Miguel Alaminos (malaminos@histolii.ugr.es)
Jose Rios (jose.rios@uab.es)
William L Gerald (geraldw@mskcc.org)
Brian Kushner (kishnerb@mskcc.org)
Mike LaQuaglia (laquaglia@mskcc.org)
Jaume Mora (jmora@hsjdbcn.org)

Version: 3 Date: 19 November 2008

Author's response to reviews: see over
Manuscript Title:

Specific Gene Expression Profiles and Unique Chromosomal abnormalities are Associated with Regressing Tumors Among Infants with Dissiminated Neuroblastoma.

Reviewer's report

Reviewer: PAOLA SCARUFFI

Version: 2

Reviewer's report:
I have reviewed the resubmission for this manuscript and I feel that the Authors have addressed my major concerns appropriately. This reviewer is delighted that Authors followed suggestions and modified both title and conclusions in order to avoid drawing definitive assertions. Actually, the manuscript is a description of chromosomal alterations and gene expression profiles, without identification of any specific gene-signature associated with infant NB outcome, as Authors themselves state in the report to my previous comments.

There are still some points that need the Authors’ attention.

1. **Authors should be careful in reporting the exact number of annotated sequences identified by gene expression analyses.** In the response to my comments, Authors correctly indicate the number of selected probes sets (Table 2A: 233 probe sets, Table 2B: 224 sets, Table 2C: 107 sets) and of annotated sequences (Table 2A: 231 sequences, Table 2B: 220 sequences, Table 2C: 107 sequences). On the contrary, in the manuscripts they still report wrong numbers of differentially expressed genes in Results (page 9, row 20; page 10, row 13; page 12, row 1), Legends of Supplementary Table 2A, Table 2B and Table 2C, Legends of Figure 1, Supplementary Figure 1A and Supplementary Figura 2B.

The reviewer is right, we apologize. The incorrect numbers reported in the manuscript have now been corrected as follows:

In the text:
- pg 9, line 19: “The analysis revealed 231 genes …”
- pg 10, line12: “… identified 220 differentially expressed genes…”
- p10, line 24: “…generated by 107 differentially expressed genes …”

In the legends:
Supplementary Data: pg 29
Table 2A. List of 231 differentially expressed genes; Raw $P < 0.01$;
Table 2B. List of 220 differentially expressed genes; Raw $P< 0.01$;
Table 2C. List of 107 differentially expressed genes; Raw $P< 0.01$.

Figure legends: pg 30
Figure 1: … profiles associated with 231 differentially expressed genes …

Supplementary Data: pg 30
Figure 1A. … months: 220 differentially expressed genes.
Figure 1B. … amplification: 107 differentially expressed genes.

2. Similarly, Authors should not use indiscriminately terms “genes” and “probe sets” in sentence “The number of genes common between…..on gene expression profiles.” at Results section at page 11, rows 7-10 and in Supplementary Figure 2.

Following the reviewers suggestions we have modified the term “genes” and “probe sets” in the text (pg 11, line 6) and in Supplementary Figure 2.

3. Since data size is limited, I would recommend to remove percentages of alterations also in the Abstract and Results (“Allelic and MYCN analysis” paragraph, pages 8-9)

As recommended by the reviewer, we have removed the percentages from the Allelic and MYCN analysis in the Results section, leaving the number of cases found with alterations out of the total number of cases analyzed (page 9, lines 2, 4, 5, 8, 9, 11 and 12).

Given the limited availability of tumour tissue for a few cases, the cohort sizes vary between the different analyses; this concept is explained in the result section (see page 8, first sentence of the second paragraph), yet, it can not be included in the abstract. For this reason we consider that the insertion of the number of cases with alterations in the Abstract could be confounding for the reader (see below). We would thus prefer leaving the percentages in the Abstract.

“Methods: Thirty-five NB tumours from patients diagnosed at < 18 months (25 stage 4 and 10 stage 4s), were evaluated by allelic and gene expression analyses.

Results: All stage 4s patients underwent spontaneous remission, only 12/25 stage 4 patients survived despite combined modality therapy. Stage 4 tumours were 19/21 near-diploid/tetraploid, 11/25 MYCN amplified, 17/22 had 1p LOH (50% 1p36), 5/22 11q and/or 14q LOH (5/22) and 8/17 had 17q gain. Stage 4s were 9/10 near-triploid, none MYCN amplified and LOH was restricted to 11q.”

4. As regarding microarray platform (Affymetrix U95), it is true that results obtained using such outdated chip have recently been reported by leader groups involved in NB investigation, but they specified that those data were partially the ones reported on earlier (i.e.: Fujita et al in JNCI 2008 and Mosse et al in Genes, Chromosomes and Cancer 2007 refer to Wang Q et al, Cancer Res 66, 2006; Schramm et al in CCR 2007 refer to Schramm A et al, Oncogene 24, 2005). Similarly, it would be important that Authors give information about the fact of having partially presented in this manuscript gene expression profiles that have
been already reported (i.e. in Alaminos M et al., Genome-wide analysis of gene expression associated with MYCN in human neuroblastoma, Cancer Res. 63(15):4538-46, 2003). This will not preclude publication of the manuscript, rather it would represent an added value because in future further information could be drawn putting together all the molecular data collected on the same subset of patients (and publicly available) by using different technologies.

The results described in this manuscript derive from a comparison analysis between subgroups of NB cases included in our microarray gene expression analysis generated using 106 NBTs. The prior study of our group reported by Alaminos M et al., (Cancer Res. 63(15):4538-46, 2003) included the first 40 samples included in the 106 MSKCC database (3 ganglioneuroma, 3 stage 4S, 5 stage 2, 8 stage 3 and 21 stage 4 cases). In that study we compared MYCN amplified versus MYCN non-amplified regardless of any other variable. Recently we have published data using the same database asking other interesting biological questions like ploidy (Cinzia Lavarino, et al. Differential expression of genes mapping to recurrently abnormal chromosomal regions provide the biological basis of ploidy in neuroblastic tumors. BMC Medical genomics 2008; 1: 36). Even though the cohorts of cases analyzed in the abovementioned studies derive from a common microarray analysis, only a very small group of cases are common to the three studies.

Public access for the microarray data has now been made available at the public data repository: caArray – Array Data Management System, National Cancer Institute.

All the data related to this study may now be viewed at:

https://array.nci.nih.gov/caarray/home.action

Experiment ID:  lava-r-00111

Reviewer: Benedikt Brors

Version: 2

Reviewer's report:
The authors have only addressed one of my previous major points sufficiently and still need to consider the following remarks:

1. (Previous major point #1) The authors need to clearly state in the manuscript that their gene expression data is a particular subset from a dataset published previously (in Alaminos et al., Cancer Res. 2003, ref. #23). It is very well justified to have this separate subtype analysis together with new aCGH analyses, but it is also important to know how this study relates to prior publications.
The results described in this manuscript derive from a comparison analysis between subgroups of NB cases included in our microarray gene expression analysis generated using 106 NBTs. The prior study of our group reported by Alaminos M et al., (Cancer Res. 63(15):4538-46, 2003) included the first 40 samples included in the 106 MSKCC database (3 ganglioneuroma, 3 stage 4S, 5 stage 2, 8 stage 3 and 21 stage 4 cases). In that study we compared MYCN amplified versus MYCN non-amplified regardless of any other variable. Recently we have published data using the same database asking other interesting biological questions like ploidy (Cinzia Lavarino, et al. Differential expression of genes mapping to recurrently abnormal chromosomal regions provide the biological basis of ploidy in neuroblastic tumors. BMC Medical genomics 2008; 1: 36). Even though the cohorts of cases analyzed in the abovementioned studies derive from a common microarray analysis, only a very small group of cases are common to the three studies.

In addition, the authors need to provide public access to their data prior to publication, e.g. by citing an accession number from a public data repository (e.g. Arrayexpress or GEO). Description of the experiments should follow the MIAME standards (Nat Genet (2001) 29:365-371).

Public access for the microarray data is now available at the public data repository: caArray – Array Data Management System, National Cancer Institute.

All the data related to this study may now be viewed at:

https://array.nci.nih.gov/caarray/home.action

Experiment ID: lavar-00111

2. (Previous major point #2) The authors need to report for their findings some sort of significance measure that has been corrected for multiple testing. It might still be acceptable to present a screening study based on findings from uncorrected ("raw") statistics provided that data are appropriately validated afterwards (as the authors did partially by qRT-PCR). However, the corresponding FDR needs to be reported even if it's high, since it gives the reader important information with regard to the estimated number of false positives within the list. It makes a large difference whether the FDR is 30%, 50%, or 90%. It should be noted that it is possible to obtain a significant number of significant "raw" findings even from random data that have no real biological structure, just because of the multiplicity of the tests. Thus, a decent FDR will protect against these effects. (See, e.g., Storey & Tibshirani PNAS (2003) 100:9440-9445; Dupuy & Simon, J Natl Cancer Inst (2007) 99:147-157.) I recommend to omit the "step-down" procedure from the manuscript as the correction for FWER is probably way too conservative for a screening study (and the results aren't reported anyway). Furthermore, as all presented lists in the supplementary files are based on a cutoff of raw P value < 0.01, the two other thresholds in the "Methods" section (namely, 0.05 and 0.1) aren't needed and should be deleted to improve clarity.
The FDR value and PPV (Positive Predictive Value) correspondent to each differentially expressed probe set has now reported in Supplementary Table 2 (last two columns on the right hand side) and the FDR and PPV value ranges have been included in the Result section at pages 9, 10 and 11.

The ranges of the FDR and PPV values found in the three analyses are:

<table>
<thead>
<tr>
<th>Analysis</th>
<th>FDR</th>
<th>PPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis Stage 4s versus Stage 4 &lt; 12 months (RAW &lt;0.01)</td>
<td>0.22 - 0.53</td>
<td>0.46 - 0.77</td>
</tr>
<tr>
<td>Analysis Stage 4s versus Stage 4 &lt; 18 months (RAW &lt;0.01)</td>
<td>0.04 - 0.54</td>
<td>0.45 - 0.95</td>
</tr>
<tr>
<td>Analysis Stage 4s versus MYCN not amplified Stage 4 &lt; 18 months (RAW &lt;0.01)</td>
<td>0.89</td>
<td>0.104</td>
</tr>
</tbody>
</table>

In the analysis performed for Stage 4s versus MYCN not amplified Stage 4 < 18 months (RAW <0.01), all differentially expressed genes showed the same FDR and PPV values. This could suggest that stage 4s NB possess an expression profile that diverges more from stage 4 NB<18m as a whole. than from stage 4 <18m MYCN non-amplified NB. This further supports that MYCN amplification influences gene expression profiles.

As requested, both the “Step-Down” procedure and the thresholds 0.05 and 0.1 in the "Methods" section have been omitted from the manuscript.

3. In their response to my previous point #3, the authors provide a valuable discussion of their findings in relation to related studies by other groups. This discussion needs to be included, however, in the manuscript, too.

As requested, it has been included in the discussion section (pg 13-14):