Reviewer's report

Title: Carboplatin-induced gene expression changes in vitro are prognostic of survival in epithelial ovarian cancer

Version: 1 Date: 2 August 2008

Reviewer: Hasan Otu

Reviewer's report:

The authors describe time-dependent transcriptional changes to carboplatin in vitro (on 36M2 ovarian cancer cells) using Affymetrix microarrays. Time points selected are 24, 30, and 36 hrs. post-treatment, with chips run in duplicate for control and treated cells. 317 genes are found to be deregulated in a time-dependent manner and 270 genes are found to exist in 40 pathways deregulated between control and treatment samples; “time-course” and “pathway” signatures as named by the authors. The goal of the study is to understand the mechanisms of chemotherapy resistance by sampling early time points following the treatment.

Both signatures are used on external data sets to successfully distinguish between control and cisplatin or carboplatin treated samples, where the latter was administered on non-ovarian cancer cell line. Both signatures are also applied in finding low and high risk groups using microarray data of patients with ovarian cancer based on disease free survival (DFS) and overall survival (OS) characteristics. Applied on two separate data sets, survival analysis with the signatures identified favorable and unfavorable groups showing significant difference on Kaplan-Meier curves.

The clinical motivation for the study is well established and the experimental design is sufficient for obtaining preliminary results regarding transcriptional changes early on following chemotherapy that could lead to resistance. The paper is well written and results regarding validation on external datasets and biological/clinical relevance of deregulated genes/pathways are satisfactory. However, I believe the analysis methods used may not readily apply to the data set and some steps in the analysis need to be further clarified.

Major Compulsory Revisions:

1) The number of data points (3) is too small for application of time-series regression analysis. Although no reference has been given to details of time series analysis (I could not locate this on BRB’s user manual where a link is given to a manual plugin.doc, which was not reachable; would “Significance analysis of time course microarray experiments” by Storey et al. be an option?) assuming quadratic functions are used to fit the data, three time points is the theoretical bare minimum for the formulation in regression analysis to work. This lower bound should not be considered sufficient to obtain statistically sound
results. There would be two ways to solve this problem: i) increase number of time points ii) obtain deregulated genes among the treatment samples (different time points), obtain deregulated genes between treatment and control (each time point), combine the results and discover patterns of expression through clustering or some similar method on this pooled list.

2) It is not clear what type of gene expression patterns the performed time-series analyses outputs; this should be further explained. The authors state “This approach [...] identifies genes whose variation of expression over time was different between carboplatin and control treated cells.”. For example, does this approach identify genes who gradually increase in both control and treated cells over time with expression levels higher in treatment (e.g. (10 15 20) in control (40 50 60) in treatment)? One could argue that the given example has similar “variation of expression over time between carboplatin and control treated cells” (hence may be missed by the performed time-series analysis) but could be important for carboplatin treatment.

3) Although the authors state identification of 317 time-dependent deregulated genes, they also mention “genes upregulated in carboplatin treated cells”. It should be explained which samples and statistical methods have been used to obtain these genes. In Figure 1 caption F-test is mentioned but I am assuming that only the treated cells are considered there. However, FC values were shown in comparison the baseline-0 hours. Is the baseline involved in finding upregulated genes? It should also be clarified how the genes in Table 2 obtained? Is every treatment time point compared to baseline or its paired control time point; what does the FC in this table refer to?

4) In pathway analysis, all genes are input to the algorithm and a deregulation score for the pathway is calculated. However, it is not clear which samples are used. If time points are combined (i.e. all 12 chips are used) then it could hurt the individual p-values calculated for the genes as time-dependent changes are not accounted for (e.g. a gene like (40 44 41) in control but (80 45 42) in treatment could get a low p-value due to its variation in treatment but could be an important gene). If the pathway analyses are done at each time point then the number of samples for the LS/KS method is jeopardized, where it is generally suggested to have at least 5 samples (again bare minimum, the more the better) for this pathway analysis (compared to availability of four samples (chips) per time point). I suggest the authors follow a strategy similar to the one suggested in 1) and use “time-dependent carboplatin deregulate genes” for GO enrichment analysis or Ingenuity pathway analysis without regard to the support of samples used to obtain the genes. In other words, with the sample size/time points limitation, it is more sensible to obtain a gene list and investigate pathways/functional categories that are most abundantly comprised of these genes.

5) In Table 3, the authors list genes found in significantly deregulated pathways and call the ensemble of these genes (270 in total), the “pathway signature”. Is the “Genes” column listed in Table 3 show the total number of genes in the pathway? For example, “cytokine network” is shown to have 5 genes but in biocarta database the genes in this pathway are listed as ****interferon, alpha 1
Minor Essential Revisions:

1) In explaining the design of the experiment it is stated that 100uM resulted in changes at 36 hrs. but the results (Sup. Fig1) show 48 hrs. changes. The authors either could say “results not shown” for 36 hrs. changes or include it in the graph. In Supp. Fig. 1, units of y-axis should be noted.

2) In explaining survival analysis, the authors state “Kaplan-Meier DFS and OS curves were plotted for two risk groups, with higher or lower than median risk of death or recurrence.” I am wondering if all the samples were included in the two risk groups or were there samples left out from the high (or similarly low) risk groups because they resembled more of the survival in the opposite group. The reason for this is that since L1OXV is used, if the expression data were not informative of survival, we could still see a split but one where high risk could have favorable survival values. I would also suggest increase in the number of permutations (from 100) for log-rank test.

3) In Figure1 legend “denote expression” should read “denote FC”. In the figure itself, time 0 could be omitted since FC values are calculated based on time zero.

4) Would it be possible to support the validation of carboplatin related genes at individual gene level maybe by doing RT-PCR and/or comparing with other microarray data sets to see the overlap? Some studies that could be utilized are: [Reference 24 cited by the authors]; [Peters D, Freund J, Ochs RL. Genome-wide transcriptional analysis of carboplatin response in chemosensitive and chemoresistant ovarian cancer cells. Mol Cancer Ther 2005 Oct;4(10):1605-16]; [Boyer J, Allen WL, McLean EG, Wilson PM, McCulla A, Moore S, Longley DB, Caldas C, Johnston PG. Pharmacogenomic identification of novel determinants of response to chemotherapy in colon cancer. Cancer Res. 2006 Mar 1;66(5):2765-77. **although oxaliplatin is used on colon cancer, could be interesting to compare]; [Huang K, Mok SC, Ng S. Identification of drug associated genes in ovarian cancer by RNA profiling. GEO Accession # GSE2058]
Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:
I declare that I have no competing interests