Reviewer's report

**Title:** An integrative approach to identify cancer chemoresistance-associated pathways

**Version:** 2  **Date:** 6 October 2010

**Reviewer:** Meng Li

**Reviewer's report:**

The manuscript by Chao et al., entitled “An integrative approach to identify cancer chemoresistance-associated pathways” uses in silico methods to systematically evaluate the mechanism of platinum resistance onset. The authors reconstructed an interaction network focusing on reported platinum resistance mechanism by integrating selected pathways from three major pathway/protein function databases, PID, KEGG and TRANSFAC. Starting with the reconstructed network, the authors used k-shortest path algorithm to identify the most relevant pathways to chemoresistance. The identified pathways were then scored using differential expression values of their members from two previous microarray studies in chemo-sensitive/-resistant ovarian and lung cancers. In addition, the authors also used an additional pathway database Reactome to evaluate the significance of pathway members based on pathway topology characteristics.

The highlight of the paper is that the authors used a simple graphic approach to pinpoint the key regulators contributing to chemoresistance in a customized landscape of molecular networks. However, some major questions need to be addressed to justify the approach and the significance of findings. In particular,

**Major:**

1. Only one microarray result was used for each of the ovarian and lung cancers. Many microarray analyses of chemosensitive samples and their resistant counterparts have been conducted in different labs using tissues or cell lines models. Despite some major findings, there are significant differences between these independent studies. In this manuscript, the microarray results were used as key data to evaluate the relevance of the discovered pathways and chemoresistance mechanism. To have the findings better justified, I suggest the authors analyze more microarray results, and taking into consideration the discrepancies between microarray experiments.

2. In the results and discussion, the authors argued that “most results are supported by known biological evidences, which indicate that this work has the latent capacity to predict candidate chemoresistance-associated genes”. However, it seems that both the integrated network and the seed nodes selections were already pre-filtered based on prior knowledge on chemoresistance mechanisms. Thus, the identified shortest pathways may represent highly connected paths between the selected seeds, regardless of their relevance to chemoresistance.
3. Please provide more details on network construction.

a) It seems that gene symbols were used as basic annotation to build the interaction network. However, gene symbol is one of the most ambiguous naming systems (with many synonymous). If possible, I would suggest the authors use a more defined naming system for network construction, such as RefSeq, Entrez, Genbank, etc.

b) Please provide version information and statistics on each of the three databases used to construct the network.

c) Could the authors provide a summary of results in main text or as additional files? For example, what are the pathways and what are their scores (or a general table with number of pathway members and scores).

Minor:

1. It is unclear how the seed notes were selected, especially the ones listed as user interested gene symbols. Please provide more references.

2. The selection of terms from the three databases was not clarified. On page 8, “Arising form the focus of this study is to identify the differential expression pathways during platinum-based chemotherapy drugs resistance, we determine diversified pathways correlated with cancer diseases, DNA repair, and metabolism to parse and integrate”. Please provide more reference or basis for the selection of terms.

3. DNA appears multiple times in a single pathway. When preparing the integrated network, the authored mentioned in the methods that pathways were assembled without duplicate links and elements. Was DNA excluded from this rule?

4. Please remove return marks from all figures.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Needs some language corrections before being published

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests