Reviewer's report

Title: CDCOCA: A statistical method to define complexity dependence of co-occurring chromosomal aberrations

Version: 1 Date: 22 September 2010

Reviewer: Nuria Lopez-Bigas

Reviewer's report:

The authors describe a statistical approach to identify co-occurring copy number alterations (CNA) and they apply it to two datasets, one of bladder cancer and one of mantle cell lymphoma.

The method, named CDCOCA, is simple yet powerful. The highest advantage in contrast to previous methods is that it takes into account sample CNA complexity, this is if a sample has high number of CNA the simulated datasets will maintain this complexity.

The article is well written and the methods are clearly explained. I just have some comments and suggestions:

1. Page 3 line 23, 24. The sentence “A gain and loss association on same chromosome (e.g. -1p and +1q) will be referred as bidirectional change”. What’s the need of defining this term? It is not used anywhere in the paper, so it is not necessary to define it.

2. How many simulations are done in the two cases presented? What is the value of S? This should be clear in the methods.

3. Figure legends are not clear enough and contain errors.
   3.1. Indicate that g is for gain and l is for loss in Figure 4, 5, 6 and 7.
   3.2. Fig5 -> Figure 4 represents associations as in figure 3 – it should say Figure 5 represents associations as in figure 4.
   3.3. Fig5 -> Most frequent CNA associations are (one) involving genetically linked CNA.
   3.4. Fig7 -> figure legend unclear

4. The R package “CDCOCA” could be submitted to bioconductor.

5. Authors use a different p-values cutoff for the two datasets. Why is that? How did they decide the p-value cutoff?

- Typos:
6. Abstract line 9: then coming from samples, should be than coming from samples.
7. Page 3 line 18: for very chromosomal arm, should be for every…
arbitrary cutoff on the p-value, why this is different for the two datasets.
8. Page 3 first line in Model section: dimentions should be dimensions.
9. Page 4 line 1: “..weights given to samples” -> incorrect sentence.

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.