Reviewer’s report

Title: Genome instability model of metastatic neuroblastoma tumorigenesis by a dictionary learning algorithm

Version: 3  Date: 16 June 2015

Reviewer: Nicola Ancona

Reviewer’s report:

In this paper, the authors address the study of metastatic neuroblastoma (NB) in pediatric patients and try of characterizing the genomic events underlying the onset and progression of the disease by comparing data relative to stage 4S and stage 4. To this end, they make use of array comparative chromosome hybridization (aCGH) data of 190 metastatic NB patients (63 stage 4S and 127 stage 4) collected merging six publicly available datasets from GEO repository. They propose Enhanced Fused Lasso Latent Feature Model (E-FLLat) a new model for aCGH data analysis developed in the general framework of Regularization Theory which is the state-of-the-art framework for solving ill-posed problems characterized by a small number of samples and a huge number of variables as the one addressed in this paper. They show that their regularized approach is able to highlight chromosomal aberrations as triggering events for malignant transformation.

Remarks

My overall evaluation of the paper is positive. The authors exhibit a deep knowledge both of the methodological aspects described in the manuscript, both of the biological aspects related to this complex and multifactorial disease. The authors provide novel insights relative to the progression of this pathology making use of publicly available data and a novel dictionary learning method in which each sample is represented as a linear combination of atoms of a dictionary. Interestingly, these atoms are learned from the available data and for this reason this approach could really provide new hypotheses on the pathogenesis of NB.

Minor Essential Revisions

Despite the manuscript provides novel insights on this form of pediatric cancer, there are some aspects that the authors should address before publication.

1) In recent years, there has been much work done in developing statistical methods for analyzing aCGH data, but a description of the state-of-the-art is missing. The authors should describe the drawbacks of the current methods and how the method they propose overcomes these critical issues.

2) A representation of a typical aCGH sample by using a picture could increase the comprehension of the methods proposed in the paper. In particular, the
authors should show how a single sample could be represented through a linear combination of atoms with suitable coefficients. On this aspect, I found the paper of Novak et al. [18] very interesting.

3) Finally, it could be interesting to provide a description of the genes harbored in the chromosomal regions highlighted by the statistical approach, describing the biological mechanisms underlying the implications of these genes to the onset and progression of NB.

**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.