Reviewer's report

Title: Prognostic and predictive value of cathepsin X in serum from colorectal cancer patients

Version: 1
Date: 11 February 2014
Reviewer: Paul Span

Reviewer's report:

In this manuscript, the authors set out to validate earlier results on the prognostic value of serum Cat-x in a limited number of colorectal patients. No difference in serum Cat-x is found between controls and CRC patients, and no difference in survival between patients with high or low Cat-x serum levels. However, further analyses revealed that Cat-x was prognostic within the group with resectable, stage I-III, disease, and that this prognostic value was only present in patients that did not receive chemotherapy. The authors conclude that although the validation was negative, the present results should be confirmed in prospective studies.

Major remarks

As often with this kind of studies, the statistical methods are important. In this manuscript, the authors report values as mean +/- SD, suggesting normal distribution of values, but also employ Spearman rank correlation tests, a non-parametric test. Similarly, Cat-x is entered as a continuous variable in Cox regression, but only after log2 transformation. Why was this done? Was this to normalize Cat-x values? If so, values should not have been reported as mean +/-SD. On the other hand, from Table 2 is seems that Cat-x levels are normally distributed. The authors should more clearly describe which test was done and why. E.g., on page 9 they report multivariate analyses including interaction terms, but in the statistics section of the M&M section this is not explained, or at least not in sufficient detail.

In line with the previous comment, on page 10 the authors report how they used median-dichotomized patients, with or without chemo, and then between tertiles, and stratified for stage, i.e. I-III or III alone etc. They do these analyses in both univariate and multivariate analyses. Overall, this large number of analyses would suggest that this study might suffer from a large chance of false-positive results, especially as the authors do not correct for multiple testing. I would strongly suggest to limit the number of statistical tests, and the number of variables in multivariate analyses.

Can the authors hypothesize on the origin of Cat-x in controls? It would seem strange that serum levels are similar in patients and controls, while on the other hand Cat-x is informative on survival of patients.

Minor remarks
On page 8, the authors state that they used Cox regression analyses to assess the “association of total Cat-x with overall survival and other clinicopathological parameters.” The authors probably mean “association of total Cat-x and other clinicopathological parameters with overall survival.”

Page 13: what is a “breast cancer expressing mice model”?  

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

**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