Reviewer’s report

Title: Presence of activating KRAS mutations correlates significantly with expression of tumour suppressor genes DCN and TPM1 in colorectal cancer

Version: 1 Date: 19 March 2009

Reviewer: Donia Macartney-Coxson

Reviewer’s report:

General Comments

The authors carried out a study to identify novel genes involved in the development of colorectal cancer (CRC) initially using cancer and apoptosis specific micro-arrays. The clinical resource of 16 matched colon normal/tumour samples, though not large in size is significant. Having identified 30 differentially expressed genes they went on to validate their micro-array data for 4 genes, and to investigate the presence of KRAS activating mutations and potential correlations between KRAS mutations and mRNA expression. The most pertinent observation may be the correlation between changes in mRNA expression for TPM1 and lack of KRAS activating mutations. The correlation with DCN requires further analysis (see major criticism 2a).

The authors state that their aim is to identify novel genes involved in CRC development. However, they do not go on to investigate the 3 differentially expressed genes which had not been previously investigated/identified in the existing micro-array analyses they cite. Instead they concentrate on 4 genes which “have been shown to play a significant role in CRC development.” It is certainly valuable to study known CRC genes but the manuscript should be amended to make the aims, experimental approach and arguments consistent.

The paper would be stronger if other good candidates within the 30 differentially expressed genes had been investigated. Although in the absence of independent samples to validate the observation (see major criticism 1) this would be of limited use.

Major criticisms
1. Abstract/Discussion/Conclusion

The authors should be careful not to overstate the significance of their gene expression data. For instance in the abstract “We are the first to report under-expression of three important tumour suppressor genes....” Under expression of TPM1, DCN, and SLC26A3 mRNA has been previously reported in CRC in independent micro-array analyses (as per authors table 2). It would be more accurate to say that their observations provide further weight/evidence for decreased mRNA expression of TPM1, DCN, SLC26A3 and CALM3 in colorectal cancer.

With regard to the real-time PCR data the authors have carried out this analysis on the same samples which they used for the micro-array and as such have validated their micro-array data rather than independently validating the observation of differential expression. Independent validation would need to be carried out in additional samples not used for the micro-array. Real-time PCR is indeed a robust and more economic way to do this in a larger number of samples.

2. Real-time PCR.

a). The ABI assay-on-demand which the authors have selected for DCN spans a SNP (single nucleotide polymorphism) in the primer or probe sequence (see ABI assay information). If any of the 16 samples contain this SNP it is likely to significantly affect the efficiency of the PCR reaction and therefore the gene expression value observed. Therefore, the genotype of the 16 samples needs to be verified. If any individual has the SNP the analysis will have to be repeated with another assay. From the variability in expression observed in Figure 2 I would guess that this might well be the case.

b). The real-time PCR methods section states that VIPR1 was measured in tumour samples and their adjacent normal tissue. No primer/probe sequence or Assay-on-Demand is referenced. Was VIPR1 used as a control? The authors used GAPDH as an endogenous control. They need to clearly state whether GAPDH showed significant variation in expression between tissues and samples.

3. Graphs should clearly state that mean data are shown on the Y-axis (Figures 1 and 2). Error bars should show Standard error of the mean, rather than standard deviation, as statistical inference is being made. (Altman DG & Bland, J.M. Standard deviations and standard errors. British Medical Journal 2005, 331: 903).

4. Table 2:

This is an important table which summarises the micro-array results, and places them in the context of other publically available micro-array data. As such the columns showing “Down”, “Up” and “No Change” would be more informative if they contained a ratio of change in expression equivalent to the authors observed value (“Log2 ratio”). The appropriate data could be referenced in superscript e.g. -1.1616. The table could also be simplified by removing Location
and RefSeq GI (as this will be available in their publically submitted data). The “Name” and “Function” columns would more accurately be labelled ‘Gene Symbol’ and “Gene Name”. For ease of comprehension the “other methods” column could be renamed along the lines of “independent validation”.

5. Figure 1.
The strength of this study is the 16 matched normal/tumour samples. By presenting the normal and tumour gene expression data as independent groups this reduces the power of the study. The dCt data for each set of paired samples could be presented, or alternatively the fold change. Real-time PCR analyses fold change could be calculated for each of the 16 pairs, and the mean fold change compared to the null hypothesis of no fold change. Using the power of the matched pairs will increase the statistical power of the analyses. Admittedly, this would make the variation between individuals more obvious, but if anything this adds strengths to the authors argument that there are a number of pathways (including ones yet to be identified) to CRC development. This would also neatly summarise the real-time PCR results section into a figure rather than text.

6. With regard to the KRAS activating mutations did the authors investigate the normal DNA for those with a mutation in order to differentiate between germline or somatic mutation?

Minor revisions

1. Abstract:
a) “despite the identification of the major genes” delete “the”.
b) “using real-time PCR, we also searched for chromosomal abnormalities....” Edit sentence to make it clear that real-time was used for mRNA expression, and copy number variation analyses, and that potential correlations between gene expression and KRAS mutation were investigated.

2. In the methods section define dCt. For instance, test gene expression was normalised to GAPDH (dCt).

3. The results section states that 31 differentially expressed genes were identified. Table 2 shows 30, and the discussion mentions 30.

4. The authors should clearly state in the results why they looked for KRAS activations in codons 12 and 13 alone. In addition, it would be useful to highlight why they initially used SSCA, and its utility for codon 12 and 13 mutations, rather than directly sequencing the samples.

5. Discussion
The background on each gene of interest is quite long and could be edited. It could more briefly highlight agreement/disagreement with the literature and discuss why down-regulation might be pertinent to cancer development, and possible alternate pathways (as per the initial aim).
Discretionary Comments

1. The Table 2 legend says that “fold change” is given as Log2 (tumour/normal). The authors may wish to change the values to an actual transposed fold change as this is intuitively easier for the reader to understand. For instance, Log2 ratio of -0.50 would be -1.4 fold change. The table legend also refers to NSCLC rather than CRC.

2. It would be useful to know why the authors selected the particular previous micro-array data to compare their results to. Do they feel these are representative of the publically available data?

Overview:
In its current form the manuscript is of limited interest. Verification in additional samples (if available, accepting that this is a limited resource) and/or of additional genes and possible protein for a couple of good candidates would greatly increase its interest to those in the specific field of interest.

Level of interest: An article of limited interest

Quality of written English: Acceptable

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