Reviewer's report

Title: Genome-wide scan for colorectal cancer susceptibility genes confirms linkage to chromosome 3q

Version: 1 Date: 14 December 2007

Reviewer: Jan Lubinski

Reviewer's report:

SUMMARY

The manuscript by Picelli et al. presents data from a genome-wide scan for linkage analysis on 30 Swedish families with a familial colorectal cancer aggregation. The study is an extension of a former one, based on 18 families. No susceptibility region could be detected under the assumption of locus homogeneity, however considering locus heterogeneity and non-parametric analysis authors were able to highlight several suggestive loci and confirm them by fine-mapping. Strongest evidence for linkage was detected in 3q. A set of 20 functionally interesting genes selected from that region was additionally sequenced to determine the presence of specific mutations potentially associated with the disease. 10 SNPs among those genes found in the family with the largest contribution to the LOD score values in 3q, were independently analyzed for association using 190 familial colorectal cancer cases and paired matched controls, but no evidence for association was found.

1. Is the question posed by the authors new and well defined?

Yes. In order to extend a former preliminary analysis on 18 families, 12 additional families were tested, the interesting regions fine-mapped, and an association study performed on an independent set of patients an controls to determine the presence of specific disease-associated genes in the region with the greatest linkage.

2. Are the methods appropriate and well described, and are sufficient details provided to replicate the work?

The methods are appropriate and well described, however there are some ambiguities (see I.5.). Details provided are sufficient to replicate the work.

3. Are the data sound and well controlled?

The data are well controlled, and the main result overlaps with a recent finding in an independent genome-wide study. From that point of view the data provide an important support for the region 3q as a risk locus.

4. Does the manuscript adhere to the relevant standards for reporting and data deposition?
Yes, the manuscript matches the standards for data reporting and deposition.

5. Are the discussion and conclusions well balanced and adequately supported by the data?

Discussion and Conclusion are mainly well balanced and supported by the results. However the discussion includes too much details of the results section and seems therefore a bit redundant.

6. Do the title and abstract accurately convey what has been found?

Title and abstract accurately convey the main information, however the authors claim to have â##confirmedâ## the linkage of 3q to colorectal cancer, while the results are not significant under assumption of locus homogeneity, just suggestive, and thus rather â##supportâ## than â##confirmâ## any finding (see II.1.).

7. Is the writing acceptable?

Yes. The manuscript reads easily and the style leaves a good impression.

I. Minor essential revisions

1) Abbreviations: The authors make an inconsistent use of abbreviations. Following the instructions for authors, abbreviations should be explained where they first appear in the text. That includes the abstract. So, HNPCC, FAP, HLOD, NPL, MLPA and SNP should be explained already in the abstract. Actually, NPL and HLOD are first explained in page 10, after they have been used countless times. That is not acceptable. Moreover CRC, MSI, SNP and PCR are not explained nowhere. Whereas PCR and SNP are may be already established in the jargon, CRC and MSI are surely not of common use.

2) Abstract: page 2 [â#ï exclued a common susceptibility region â#ï] may better read [.. excluded any â#ï], otherwise one asks himself â##which one?â##.

3) Methods: page 5 [â#ï on chromosomes 6 â#ï] should read [â#ï on chromosome 6 â#ï].

4) Methods: page 7 [â#ï 190 familiar colorectal â#ï] should read [â#ï 190 familial colorectal â#ï] to be consistent with the rest of the text.

5) Results and Discussion: There are some minor inconsistencies with the given data.

a) Methods (page 6): [â#ï 65cM interval â#ï markers D3S1558 and D3S3592.] and Results (page 9) [â#ï between D3S1558 and D3S592 â#ï 65cM.,] but later in Discussion (page 11) the interesting region is [â#ï 69cM â#ï markers D3S1278 and D3S1262.]

b) Results (page 10): [â#ï with the nonparametric (NPL) score of 2.0 â#ï at
marker D6S1656], but later in Discussion we read [â#i NPL score of 2.06 for D6S1656 â#i]. 2.0 is not an acceptable rounding for 2.06, so we guess there is an error.

c) Methods (page 6): [Database analysis revealed 340 genes â#i] but in Discussion we read [â#i contains more than 340 genes â#i].

6) Methods: page 5 additional sets of markers for chromosome 6 and 3 were used for fine mapping. For chromosome 3 the number is given, but not for chromosome 6.

II. Major Compulsory Revisions

1) In the title authors claim to have â##confirmedâ## the linkage of 3q to colorectal cancer. Results, however, are not significant under assumption of locus homogeneity, just suggestive. More relevant results are found for locus heterogeneity and nonparametric analysis, but altogether the title may be a bit overstated. â##Supportâ## may be a better option than â##confirmâ##. The same argument applies throughout the text.

2) The authors claim in Methods that FAP was excluded by colonoscopic examination. However, sometimes APC and MYH positive patients show attenuated symptoms, where presence of polyps is difficult to determine by colonoscopy. So APC and MYH are usually analyzed too, to exclude mutations. That approach was for instance followed in Kemp et al. 2006, the article where a linkage in 3q was found, and the present article supports. Did the authors perform also a APC and MYH analysis? If not, please explain why.

3) Relationship patterns are checked using Pedcheck, and authors claim that â##any markers violating Mendelian rulesâ## were deleted. If Mendelian rules are not fulfilled, one may also exclude the whole person/genotype from the study. It depends on how many markers do not fulfill the rule. We guess the authors did this carefully, but it would be desirable to know the range of markers violating the rule per person (in number of markers).

4) 20 genes are chosen - if we follow Methods â## by their (potential) function, however in Discussion also â##expressionâ## seems to play a role. Did the authors also perform expression analysis, are the data from literature? Moreover it would be interesting to know the exact criteria followed concerning gene function.

5) Familial colorectal cancer patients are chosen for the cases group in the SNP association analysis. Usually consecutive cases are taken for that purposes. Why did the authors choose familial cases?

6) Authors consider missense and promoter mutation as harmless. That is inexact, missense mutations have often been reported as risk increasing factors (e.g. for MLH1 exon 18). Authors claim they were not able to detect a significant association between cases and matched controls. However no further information is given to the reader. So he cannot distinguish between the case
where a missense mutation is quite frequent in both groups (thus pointing to a harmless polymorphism) and the case where the missense mutation is rare but more frequent in the patients group (a low risk marker), and nevertheless not significant due to small sample size. To interpret correctly that part of the results (beyond the search for high-risk markers) it would be important to have information on the statistical power for the compared SNP frequencies. The reader should know if the statistical power was high enough (remember we had just 190 individuals in each group).

7) The discussion is highly redundant with the result section. Many details are unnecessarily repeated. Some data are shown even for the first time in Discussion, as if it was the Results section: (page 11) [After fine mapping, an NPL score of 2.3 was obtained â#i], where is that important information in the Results section?

Discussion should in general be revisited to avoid a mere repetition of the results section.

What next?: Accept after discretionary revisions

Level of interest: An article of importance in its field

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