Reviewer's report

Title: Prognostic relevance of Wnt-inhibitory factor-1 (WIF1) and Dickkopf-3 (DKK3) promoter methylation in human breast cancer

Version: 1 Date: 24 November 2008

Reviewer: Mary Jo Fackler

Reviewer's report:

This manuscript reports a fairly standard prognostic study of the methylation of two genes previously defined to be of interest in breast cancer. Among the results, of greatest interest is the finding of a moderately high hazard ratio of 17.2 for DKK3 methylation as a predictor of poor overall survival. The number of cancers studied (150) is potentially high enough to permit multivariate regression analysis that includes many of the most common clinical variables. The tables are very useful. Table 3, for example, addresses some of the requirements of the REMARK criteria.

Unfortunately, it appears that the main reason for good performance of the DKK3 marker is that it 1) is associated with other conventional variables and 2) was not handled appropriately, as were some of the conventional variables, in the lifetable models.

Major points

1) The statistical conclusions are in question due to a failure to correct for having performed multiple statistical comparisons. Despite the use of dozens of statistical tests, the authors claim in the Methods section that p values less than 0.05 “were considered significant” in the absence of appropriate correction. This is inappropriate. Because of this shortcoming, many of the “findings” of the study cannot actually be distinguished from random patterns. For example, the hazard ratio for DKK3 methylation in predicting poor overall survival is significant at the uncorrected p value of 0.007. But once a correction is applied, this is no longer statistically significant, for the corrected p value would be much higher than 0.05. Likewise, the association of DKK3 methylation and short disease-free survival was not statistically significant.

2) The independence of DKK3 methylation as a prognostic factor is not established. See below.

3) To call DKK3 a “strong prognostic factor” and a “potent prognostic biomarker” in the Abstract seems misleading (even if we were to ignore the lack of statistical significance when appropriate corrections are made for multiple testing). In the univariate association study for predicting poor survival, the sensitivity and specificity were 46% (1-54%) and 97%, respectively. Thus, the “strong” test misses most of the patients who might benefit from an altered care pattern, and would not change the care of those receiving standard care. In the multivariate
result, the confidence interval covers a broad range, which at its lowest is given by the authors to be a hazard ratio of 2.2. As discussed by Pepe et al (Am J Epi 159:882), hazard ratios in the range of 2 to 20 are not likely to separate the patients into clinically distinct treatment categories. That is, they are not clinically “potent”. A more convincing argument would need to be offered for the clinical advantages of the new tests.

4) The statistical test applied was insufficient to conclude that DKK3 methylation was an independent variable associated with prognosis. The Abstract does not report that a prognostic model containing DKK3 methylation status can out-perform a prognostic model lacking this new variable. One method by which to do this compares the loglikelihood values of a model with and without the variable. Unless the two models differ significantly, one would assume that DKK3 was merely sharing the information already available from older, established clinicopathological parameters. The beta value provided in a given Cox model does provide a measure of the strength of the prognostic vector assigned by the model to the variable, but one needs to compare two models in order to determine whether the model has attributed some of the information that was already available from other variables. In order to do this, one needs to compare the new model to one specifically lacking only the variable of interest, in which, obviously, the model could not be attributing the already-available information to the new and herein-absent variable.

5) The statistical treatments of histological type in the Cox analysis appears incorrectly performed. According to Table 5, histological type was treated as an ordinal variable. Histological type is ungrouped, presumably leaving three data categories to be assigned numerical values by the software. Cox analysis classically fails with anything other than binary data types, and thus it is usually advisable to “dummy up” these complex variables to convert them into binary variables. Whatever category is assigned to the highest ordinal value is assumed to perform exponentially worse or exponentially better than the middle ordinal value, when each is compared to the lowest ordinal value. This is a silly assumption. For example, a clumsy treatment of “histological type” could yield a poor performance in the Cox model – which seems plausible due to the low HR of 1.05 for this variable.

6) The association of DKK3 methylation and overall survival is rendered largely irrelevant by the lack of statistical association between DKK3 methylation and disease-free survival (see Figure 5). Thus, it seems to be a random coincidence.

Minor points

7) The statistical treatments of age in the Cox analysis is unclear, but is potentially important because of the discovered association of age and methylation status. According to Table 5, the continuous variable age was treated as “ordinal”. This is ambiguous. If age was “dummied up” into a binary variable (into two categories) by dividing the data at the median, this would not be an “ordinal” dataset. To treat age as an ordinal variable implies more than two categories, yet just how the Cox proportional hazards assumptions would handle
ordinal data having three or more categories is not explained. The authors should clearly explain why age was not handled in binary form as were the other continuous variables (such as tumor size).

8) The grouping of all three types of primary breast carcinoma (ductal, lobular, and other) is problematic. These are very different tumor type. It might not be reasonable to expect a prognostic behavior to be manifest similarly among the three categories. The Cox model and the univariate models DO expect similarity (“proportionality”) of behavior from a variable among the subclassifications of the patient set. Stratifying the data and performing a separate analysis for each distinct category is one way to handle such issues.

9) The Methods section does not state that all diagnoses were reviewed by coding the diagnostic samples and submitting them to a pathologist expert in the disease.

10) The statistical methods state that “only patients for whom the status of all variables was known were included in the proportional hazard models”. How many remained suitable is unclear.

11) The specificity of the finding was not seriously examined. It is possible, for example, that the tumors having worse prognosis are methylated with some specificity at many genes, and that DKK3 is merely one of them. Without having examined the methylation of a large panel of gene (i.e., many more than two genes), one cannot state that the choice of gene, nor the function of the gene, has any special relation to the prognostic associations sought.

12) The conclusion in the Abstract, that WIF1 and DKK3 have different prognostic associations (“only DKK3 methylation proves …”) , is not established with a statistical direct comparison of the two markers’ associations. The conclusion thus exceeds the data. The more likely explanation for the behavior of these two markers is that due to lack of adequate statistical power, this is an apparent, rather than an actual, difference.

13) To “conclude that DKK3 may exert important tumor suppressive functions” exceeds the data. The data concern only an association with prognosis. Associations cannot be used to conclude causative relationships unless all other alternative explanations can be excluded.

14) The statistical treatments of the nodes and histological grade are unclear. Grades 1 and 2 were grouped, but it is unclear that this would be appropriate. Similarly, N1-3 are grouped, without comment.

Miscellaneous

15) Footnote #2 of Table 5 seems to state that only the HR values of the multivariate analysis were obtained from the Cox hazard estimates. How were the HR values of the “univariate” analyses obtained?

16) The Abstract’s paragraph on Methods should state that the methylation was interpreted in a binary, or qualitative, fashion. An alternative would be to use a
quantitative scale.

17) In the Abstract, providing the brand of software is presumably irrelevant, for all competent statistical packages should give the same answers. If the authors feel that the choice of software is a relevant detail deserving of mention in the Abstract, they should convey in the Abstract why they chose one package over another. If, instead, they feel that the choice was irrelevant to the results obtained, the brand need not be mentioned except in the formal Methods section. In the Abstract, mentioning the types of statistical tests is indeed informative, but is most informative for the reader when placed adjacent to the p values in the Results paragraph rather than in the Methods.

18) The Abstract mentions a statistical test finding that “WIF1 methylation was significantly correlated with methylation of DKK3”, yet the statistical methods listed (Fisher test) is a test of association, not a test of correlation. In the formal Methods section, the correct term, “associations”, is used by the authors.

19) In the Abstract, the Kaplan-Meier equations are seemingly depicted as a statistical evaluation, which they are not. Instead, it was the logrank test that the authors used to statistically evaluate the lifetable data in a univariate manner.

20) The term “clinicopathological patient parameters” is redundant, for “clinicopathological parameters” is adequate.

21) The word “Univariate” in the Results of the Abstract should be revised to read, “In univariate analysis,“.

22) Proper hyphenation is needed and aids readability. “Patients with DKK3 methylated tumors”, for example, should read, “patients with DKK3-methylated tumors”, since the “patients specifically with DKK3” had not exclusively “methylated the

**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:**

We declare that we have no competing interests.