Reviewer’s report

Title: Genomic alterations identified by array comparative genomic hybridization as prognostic markers in tamoxifen-treated estrogen receptor-positive breast cancer

Version: 1 Date: 9 December 2005

Reviewer: Michelle Nessling

Reviewer’s report:

General
The authors present an array-CGH study with genomic DNA of 28 primary invasive estrogen receptor (ER)-positive breast cancer tissues. All patients had received tamoxifen treatment for one year, and within 5 years of diagnosis, roughly one third of the patients (9/28) displayed distant metastasis (Recurrence group), whereas 19 patients were without pathological findings (Non-recurrence group). The focus of the study was to identify prognostic markers for clinical outcome. Various genomic alterations could be detected. The most common altered chromosomal regions among the 28 investigated tumor samples match previously published data from the literature. Applying different tools of statistical analysis, distinct genomic regions were identified of which DNA copy number changes discriminated patients in the Recurrence group from those in the Non-recurrence group, namely loss of 11p15 and 1p36 as well as gain of 8q21. These genomic regions might harbor genes with metastasis-suppressing or -enhancing properties. The present study by Han et al. delivers interesting data on genomic aberrations in genomes of ER-positive breast cancer tissues.

---------------------------------------------------------------

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

Major criticism on the study concerns the poor description of statistical analysis throughout the text. For a reader unspecialized in statistics the different steps in analysis are not comprehensible.
1. Methods (Statistical analysis, page 9): Of the various statistical approaches the principles should be outlined and their assets and drawbacks commented. In particular, the authors should give reasons why they chose which kind of analysis.
The parameters of analysis must be explained and the abbreviations notified.

---------------------------------------------------------------

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)
1. Methods (page 6, 1st paragraph): The authors should notify the commercial provenience of the array of 1,440 human BAC clones they have used for the present study (MacArrayTM Karyo1400 from Macrogen, Inc.).
2. Methods (page 6, 2nd paragraph): Even though the sequences of the BAC clones on the array were blasted and mapped according to their position in the human genome, genomic aberrations identified are only given in the minor resolution typically reached with conventional, low-resolution CGH onto metaphase chromosomes. No information about genes, located in the respective regions which are potentially affected with loss or gain, are indicated. I wonder why no genes on BAC clones are identified. Since data do not represent high-resolution, they should not be presented as such.
3. Methods (page 8, 1st paragraph): "Clones that had more than 50% missing values were excluded for further analysis." The different BAC clones on the array are spotted in triplicate, thus "more than 50%" is equivalent with 2/3 or all 3 spots missing. Why not say that for calculating the mean, at least 2 values of the spotted triplicate of a specific BAC clone are essential. With the next sentence, the authors say, that they evaluated only triplicates: "This threshold was surpassed by 3,993 of 4,320 spots (triplicate of 1,331 different BAC clones)". It would be clearer to say that 109 different BAC
clones have been excluded from further analysis because values of at least 2 spots of the triplicate have been missing.

4. Entire document: The use of percent indication with numbers in the decimal place to state the amount of patients or clones is an overkill in accuracy (one out of 28 patients makes roughly 4%!). It would be better to display the exact number or when preferring percent indication, to round to whole digits.

5. Results (page 10, 3rd paragraph): "...previous conventional or array CGH breast cancer studies [13, 28, 29]." Reference 14 in which many of the detected genomic aberrations had been described may also be cited in this place.

6. Table 1: Headings are not concise: Since the authors indicate the clinicopathological characteristics numerical and, in brackets, in percent, they should clarify both in the headline (e.g. n (%)), with percent indications rounded to significant digits. The first parameter in the table states the mean age of the patients and in brackets the range of the age. In the first column, the indication "Age (range)" would make this more clear.

7. Table 2: What do the ranges indicated in the [%]-column stand for?

8. Table 3 and Table 4: Since the term "amplification" generally is used to describe high-level gain, the authors should change the title in Table 3 accordingly, e.g. "...differences in frequency of gain...., Gain of 5q34 is missing in Table 3, but is mentioned in results (page 11, 2nd paragraph), whereas gains of 19p and 19q are listed in the table but not mentioned in results. Furthermore there is inconsistency concerning the use of p value and adjusted p value in Tables and results, e.g. p values in Tables 3 and 4 and adjusted p values in results (page 11, 2nd paragraph).

9. Table 5: The content of Table 5 is unclear. What does 'd' mean? 'Fold change' with respect to what? The authors should discuss why 'q' adopts two discrete values only.

10. Table 6: In the last row of the p value the point in front of the p value is missing. The authors should elucidate the information content among 'Hazard ratio', '95% Confidence interval' and 'p value' in this context.

11. Figure 4B: "+ censored" is given twice.

12. Discussion (pages 12-13): The discussion would better be written more straightforwardly. In the first paragraph, e.g., the authors discuss their approach referring to chromosomal imbalances which become specified in a subsequent paragraph (2nd paragraph on page 13).

13. Discussion (pages 12-13): The statement: "We also demonstrated an excellent correlation using real-time PCR data for validation of array data" (page 12, 1st paragraph) disagrees with the statement "Although 11p15-5 loss was the strongest prognostic factor found in our array CGH study, we cannot validate this with our real-time PCR method..." (page 13, end of 2nd paragraph).

14. The results of the present study should additionally be discussed in the context of one very recent publication by Arpino et al. (J Natl Cancer Inst 2005;97:1254-61), a large and comprehensive evaluation of the biologic and clinical characteristics of invasive breast cancers (more than 44 000 patients), comparing ER+/PR- with those that are ER+PR+.

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: No

Declaration of competing interests:
I declare that I have no competing interests