Reviewer's report

Title: Low expression of a few genes indicates good prognosis in estrogen receptor positive breast cancer

Version: 2 Date: 12 April 2009

Reviewer: Balazs Győrffy

Reviewer's report:

The question of adjuvant chemotherapy in node-negative ER+ breast cancer patients is a subject of extensive current literature. In this well written in silico study, Buechler identified the AP4 test which is capable to discriminate poor and good prognosis group in ER+ breast cancer patients thus providing information about the potential benefits of adjuvant chemotherapy. The AP4 test targets the same patient group as two currently available gene expression based assays - the Oncotype DX and the Mammaprint tests. However, the AP4 set comprises a highly condensed gene set, containing only 4 genes: 2x KI67, CDC6, and SPAG5. Since the genes can be measured by simple PCR technique, the study delivers significant improvement and effective alternative of the two above assays.

Major Compulsory Revisions:
• This is a hypothesis-generating in silico study. Given the large number of genes on the Affymetrix arrays, it is highly challenging to avoid false positive results in the development of a condensed predictor. Therefore, the specificity and sensitivity of the Affymetrix probes measuring the four genes must be validated before a diagnostic test is developed. (Alternatively, IHC or RT-PCR of the three genes could be measured in an independent sample set.)

Minor Essential Revisions:
• Why was GCRMA used for normalization? (GCRMA normalizes the complete set of samples, and thus makes extension of the dataset by single arrays not possible.) Are results using datasets generated with MAS5 normalization identical?
• BLAST should be performed to exclude potential cross-hybridization of the four probe sets to other genes.
• Many of the samples in GSE6532 are identical to samples in GSE4922 (these two GEO datasets were used among others to perform the statistical computations). Were these samples excluded? (see Gyorffy et al, Breast Cancer Res Treat, 2009).
• What is the predictive power, if we use only ONE of the four genes? E.g. Kaplan-Maier for a single gene?
• A more detailed comparison to Oncotype DX should be added (e.g. KI67 is also included among their 16 genes - both tests overlap in focusing on proliferation
signatures).

• Why is only CDC6 on the first figure? A similar graph for the other 3 genes should be added.

• Figure 4: MKI67 expression delivers the same result as the Ki-67 labeling index. What is the advantage of the AP4 test over the measurement of MKI67 expression alone?

**Level of interest:** An article of outstanding merit and interest 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.