

EDITORIAL

Microarray data analysis in neoadjuvant biomarker studies in estrogen receptor-positive breast cancer

Jingqin Luo^{1,3} and Matthew J Ellis^{*2,3}

See related research by Taylor *et al.*, <http://breast-cancer-research.com/content/12/3/R39>

Abstract

Microarray data have been widely utilized to discover biomarkers predictive of response to endocrine therapy in estrogen receptor-positive breast cancer. Typically, these data have focused on analyses conducted on the diagnostic specimen. However, dynamic temporal changes in gene expression associated with treatment may deliver significant improvements to the current generation of predictive models. We present and discuss some statistical issues relevant to the paper by Taylor and colleagues, who conducted studies to model the prognostic potential of gene expression changes that occur after endocrine treatment.

Taylor and colleagues [1] examined the time course of gene expression profile changes in estrogen (E2)-treated and E2 and tamoxifen-treated mouse xenografts. The authors presented three distinct categories of gene expression temporal profiles, each characterized by two sets of genes. Differentially expressed genes at some early time points following treatment were found to be prognostic of survival in clinical data sets, but not those identified at other time points. This implies that the timing of the post-treatment sample for gene expression analysis will be critical for the development of prognostic and predictive biomarkers.

Adjuvant endocrine treatment in estrogen receptor-positive (ER+) breast cancer patients reduces the risk of relapse and death from breast cancer [2], but large numbers of patients still die of endocrine therapy-resistant disease [3]. Researchers have therefore devoted intensive efforts to identify molecular biomarkers to

predict response to endocrine treatment and, in spite of the inherent heterogeneity among ER+ breast tumors, gene expression signatures have been successfully developed [4-6]. However, the existing signatures are based on gene expression information in a single baseline tumor sample that may not capture all the biological information necessary for predictive accuracy. Clinically, patients fall into three broad categories, continuously responding, continuously resistant, and a substantial group of patients with an initial response followed by a transition at varying rates to an acquired resistance phenotype. Late recurrence in resistant patients might be avoided if these tumors could be identified early, before the onset of clinical resistance, and subjected to an effective salvage intervention. Therefore, the discovery of gene signatures differentiating the three response groups logically requires the identification of temporal changes in gene expression along the treatment course. The paper by Taylor and colleagues [1] illustrates this principle.

Microarray gene expression data were used by Taylor and colleagues [1] for discovery and validation of gene expression signatures. Overall, the paper is a good example of the practice of microarray data analysis. Raw data were deposited in CaArray [7] to be available to the public, which encourages research reproducibility. After the gene discovery process, validation in multiple independent public datasets was carried out. An important caveat of this aspect of the paper is that these datasets are not particularly suitable to assess the primary hypothesis because all these studies report only baseline array gene expression levels - not treatment-induced changes. The true test of the approach would be to compare the prognostic information in the baseline sample with post-treatment samples taken at different time points from the same patient. Moreover, significance of potential therapy-response gene signatures in treated versus untreated patient cohorts should be interpreted with caution.

The paper also identifies areas for methodological improvement. The data analysis is limited to two-class comparison at each individual time point, which neglects time dependency in gene expression profiling. The definition of the six sets of genes is subjectively determined

*Correspondence: mellis@dom.wustl.edu

²Department of Medicine, Washington University School of Medicine, St Louis, MO 63110, USA

Full list of author information is available at the end of the article

by known pathways. Ideally, the continuous longitudinal gene expression profiling would be better analyzed as a whole by functional data analysis techniques [8,9]. Rather than traditional cluster analysis [10], cluster tools designed specifically for time course gene expression data, such as CAGED (Cluster Analysis of Gene Expression Dynamics) [11], would probably serve better. Furthermore, the class comparison in the paper depends on fold change alone, a common error in the analysis of microarray data [12]. Fold change is easy to calculate and understand; however, it is a single ratio without consideration of variability. Use of fold change usually leads to high false positives since small changes in genes with low expression levels can lead to large fold change. The hierarchical algorithm is applied to public microarray data in the paper to divide samples into low and high expression groups. The use of unsupervised clustering for class prediction is very subjective [12]. The two-color microarray design was used while the authors commented on the possible benefits of using a one-color design. The two-color with common reference design has been the most widely used in microarray experiments for its ease of implementation and analysis. The one-color design has recently emerged to be a favorite because of its simplicity and flexibility after confirmation of comparative data quality to its two-color counterpart. However, the two-color design is still reported to exhibit a small advantage in detecting differential genes, especially for genes of small fold changes [13].

In their paper, Taylor and colleagues speculated that the 'early/transient' expression changes are the causative events for tumor inhibition. This might be true but needs to be investigated more carefully in future studies. Meanwhile, it is important to acknowledge the fact that some patients who respond initially and exhibit the early/transient expression change may acquire resistance gradually. It will be challenging to pick these patients out for individualized treatment planning as the critical changes may take place only after months or years of endocrine therapy exposure.

In conclusion, we fully agree on the importance of investigating temporal gene expression profiling for prediction of treatment response. More well-planned studies will be required for insights into these complicated data sets and variability in response to treatment will be an important consideration. The task of obtaining consecutive gene expression profiling at multiple time points remains a challenging prospect but might be feasible in well planned neoadjuvant endocrine therapy studies where patients might be triaged to alternative therapy if an unresponsive gene expression profile emerged, even when the patient was in response clinically.

Abbreviations

ER = estrogen receptor.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Division of Biostatistics, Washington University School of Medicine, St Louis, MO 63110, USA. ²Department of Medicine, Washington University School of Medicine, St Louis, MO 63110, USA. ³Siteman Cancer Center, Washington University School of Medicine, St Louis, MO 63110, USA.

Published: 20 August 2010

References

1. Taylor KJ, Sims AH, Liang L, Faratian D, Muir M, Walker G, Kuske B, Dixon JM, Cameron DA, Harrison DJ, Langdon SP: **Dynamic changes in gene expression *in vivo* predict prognosis of tamoxifen-treated patients with breast cancer.** *Breast Cancer Res* 2010, **12**:R39.
2. Ali S, Coombes RC: **Endocrine-responsive breast cancer and strategies for combating resistance.** *Nat Rev Cancer* 2002, **2**:101-112.
3. Loi S, Haibe-Kains B, Desmedt C, Wirapati P, Lallemand F, Tutt AM, Gillet C, Ellis P, Ryder K, Reid JF, Daidone MG, Pierotti MA, Berns EM, Jansen MP, Foekens JA, Delorenzi M, Bontempi G, Piccart MJ, Sotiriou C: **Predicting prognosis using molecular profiling in estrogen receptor-positive breast cancer treated with tamoxifen.** *BMC Genomics* 2008, **9**:239.
4. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, Hiller W, Fisher ER, Wickerham DL, Bryant J, Wolmark N: **A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer.** *N Engl J Med* 2004, **351**:2817-2826.
5. Ma XJ, Wang Z, Ryan PD, Isakoff SJ, Barmettler A, Fuller A, Muir B, Mohapatra G, Salunga R, Tuggle JT, Tran Y, Tran D, Tassin A, Amon P, Wang W, Enright E, Stecker K, Estepa-Sabal E, Smith B, Younger J, Balis U, Michaelson J, Bhan A, Habin K, Baer TM, Brugge J, Haber DA, Erlander MG, Sgroi DC: **A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen.** *Cancer Cell* 2004, **5**:607-616.
6. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, Quackenbush JF, Stijleman IJ, Palazzo J, Marron JS, Nobel AB, Mardis E, Nielsen TO, Ellis MJ, Perou CM, Bernard PS: **Supervised risk predictor of breast cancer based on intrinsic subtypes.** *J Clin Oncol* 2009, **27**:1160-1167.
7. **CaArray** [https://cabig.nci.nih.gov/tools/caArray]
8. Ramsay JO, Silverman BW: *Functional Data Analysis*. New York: Springer; 2005.
9. Ferreira L, Hitchcock DB: **A comparison of hierarchical methods for clustering functional data.** *Comm Stat Simulation Computation* 2009, **38**:1925-1949.
10. Eisen MB, Spellman PT, Brown PO, Bolstein D: **Cluster analysis and display of genome-wide expression patterns.** *Proc Natl Acad Sci USA* 1998, **95**:14863-14868.
11. Ramoni MF, Sebastiani P, Kohane IS: **Cluster analysis of gene expression dynamics.** *Proc Natl Acad Sci U S A* 2002, **99**:9121-9126.
12. Simon R, Radmacher MD, Dobbin K, McShane LM: **Pitfalls in the use of DNA microarray Data for Diagnostic and prognostic classification.** *J Natl Cancer Inst* 2003, **95**:14-18.
13. Patterson TA, Lobenhofer EK, Fulmer-Smentek SB, Collins PJ, Chu TM, Bao W, Fang H, Kawasaki ES, Hager J, Tikhonova IR, Walker SJ, Zhang L, Hurban P, de Longueville F, Fuscoe JC, Tong W, Shi L, Wolfinger RD: **Performance comparison of one-color and two-color platforms within the Microarray Quality Control (MAQC) project.** *Nat Biotechnol* 2006, **24**:1140-1150.

doi:10.1186/bcr2616

Cite this article as: Luo J, Ellis MJ: Microarray data analysis in neoadjuvant biomarker studies in estrogen-receptor positive breast cancer. *Breast Cancer Research* 2010, **12**:112.