Expression profiling to predict outcome in breast cancer: the influence of sample selection

Gruvberger, Sofia; Ringnér, Markus; Edén, Patrik; Borg, Åke; Fernö, Mårten; Peterson, Carsten; Meltzer, Paul S

Published in:
Breast Cancer Research

DOI:
10.1186/bcr548

2003

Link to publication

Citation for published version (APA):
Gruvberger, S., Ringnér, M., Edén, P., Borg, Å., Fernö, M., Peterson, C., & Meltzer, P. S. (2003). Expression profiling to predict outcome in breast cancer: the influence of sample selection. Breast Cancer Research, 5(1), 23-26. https://doi.org/10.1186/bcr548

Total number of authors:
7

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Commentary

Expression profiling to predict outcome in breast cancer: the influence of sample selection

Sofia K Gruvberger1, Markus Ringnér2,3, Patrik Edén2, Åke Borg1, Mårten Fernö1, Carsten Peterson2 and Paul S Meltzer3

1Department of Oncology, The Jubileum Institute, Lund University Hospital, Lund, Sweden
2Complex Systems Division, Department of Theoretical Physics, Lund University, Lund, Sweden,
3Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA

Corresponding author: Paul S Meltzer (e-mail: pmeltzer@mail.nih.gov)

Received: 7 August 2002 Accepted: 20 September 2002 Published: 11 October 2002

Breast Cancer Res 2003, 5:23-26 (DOI 10.1186/bcr548)
(Print ISSN 1465-5411; Online ISSN 1465-542X)

The original version of this commentary contained an error in Figure 1. This has been corrected.
See reply, http://breast-cancer-research.com/content/5/1/57

Abstract

Gene expression profiling of tumors using DNA microarrays is a promising method for predicting prognosis and treatment response in cancer patients. It was recently reported that expression profiles of sporadic breast cancers could be used to predict disease recurrence better than currently available clinical and histopathological prognostic factors. Having observed an overlap in those data between the genes that predict outcome and those that predict estrogen receptor-α status, we examined their predictive power in an independent data set. We conclude that it may be important to define prognostic expression profiles separately for estrogen receptor-α-positive and estrogen receptor-α-negative tumors.

Keywords: breast cancer, estrogen receptor, gene expression, microarray, prognosis

Introduction

Recent studies have shown that microarray analysis of tumors generate molecular phenotypes that can be used to classify tumors into subgroups that are not evident by traditional histopathological methods, to improve diagnosis, and to predict outcomes of disease [1–6]. In a recent study, van ’t Veer et al. [7] reported, among other important conclusions, that expression profiles of sporadic primary breast tumors could be used to predict disease recurrence better than currently available clinical and histopathological prognostic factors (i.e. tumor grade, tumor size, angioinvasion, age, and estrogen receptor [ER]-α). Those investigators identified a set of 231 genes, from which they extracted 70 that were used to predict recurrences within 5 years with an accuracy of 89% for a test set of 20 tumors. However, our experience suggests that, in breast cancer patient populations, predicting clinical parameters studied. ER-α status has been shown to leave a remarkable imprint on the expression profile of a tumor, and may have confounding effects on the clinical factor studied [7–9]. Therefore, we believe that it is necessary to be cautious when interpreting the results from the study of van ’t Veer et al. and similar studies.

In the set of tumors analyzed by Van’t Veer et al. [7] there is an evident correlation between clinical outcome for the patients and ER-α status of the tumors. For that reason, we investigated whether their results can be applied to a data set from a different patient cohort in which there is no correlation between ER-α status and prognosis.

Estrogen receptor-α and gene expression in breast cancer

ER-α status is by far the most obvious and noticeable subdivision of breast tumors based on gene expression profiles. This has become evident in several studies.
[4,7–9]. In a previous study [8] we used the gene expression profiles of 47 tumors and artificial neural networks to generate a prediction model based on the 100 highest ranked discriminator genes that could accurately predict the ER-α status of 11 breast tumors from an independent test set. Furthermore, we found that sets of genes much further down our ranked list still carried substantial information about ER-α status. Using a similar approach, West et al. [9] also succeeded in predicting the ER-α status of a set of breast tumors. In their study, van ’t Veer et al. [7] also found that the main subdivision of tumors by unsupervised clustering was into two groups based on ER-α status. Indeed, there is considerable overlap between the genes that discriminate ER-α status in their data and those that were reported by us and by West et al. In addition, van’t Veer et al. also found that as many as 2460 genes (out of 24,479 included in their arrays) carried information on ER-α status, confirming our finding that a large number of genes expressed in breast tumors are associated with ER-α status.

Prediction of prognosis from gene expression profiles
In the study by van’t Veer et al. [7], disease recurrence was better predicted using the expression signature of 70 identified genes as compared with currently available clinical and histopathological prognostic factors (tumor grade, tumor size, angioinvasion, age, and ER-α status). We investigated the expression of approximately 7000 genes in a patient cohort of 44 primary breast tumors (size 20–50 mm, from lymph node-negative patients all treated with adjuvant tamoxifen). In contrast to the findings of van’t Veer et al., we could not predict relapse with statistical confidence. The tumors we studied were selected to comprise four nearly equal sized groups: ER-α-positive or ER-α-negative, each subdivided into good (no distant metastases within a follow-up period of 5–11 years, median 8.0 years) and poor (distant metastases within 6 years) prognosis groups (Fig. 1). Using either artificial neural networks [5,8] or the signal-to-noise statistic [1] and random permutation tests, we were unable to develop a statistically reliable outcome classifier in these data. This may suggest that, because ER-α-positive and ER-α-negative tumors have such distinct expression profiles, the molecular pathways that are characteristic of the aggressive tumors within these two groups may not overlap to a substantial degree. Hence, it would probably be beneficial for those two groups to be considered separately when predicting prognosis on the basis of gene expression profiles. Unfortunately, such an analysis could not be performed satisfactorily in this data set because the number of tumors in the ER-α subgroups was too low. Additionally, it should be noted that the effect of ER-α status on outcome might be even more pronounced in this data set because the patients received endocrine treatment after primary surgical resection.

Furthermore, we investigated whether the outcome predictor gene set generated by van’t Veer et al. [7] could be applied to our patient cohort. We found that none of the outcome predictor genes identified by those investigators, which were also represented in our array data (58 out of 231), was a significant outcome discriminator (OR = 1.9, 95% CI 0.5–6; Fisher’s exact test), whereas no discrimination was seen in our data (OR = 1.9, 95% CI 0.5–6; P = 0.4). We used multidimensional scaling analysis to illustrate these results (Fig. 2).

Intriguingly, for the training set of 58 sporadic tumor samples that van’t Veer et al. [7] analyzed, their supervised ER-α predictor gene set also had predictive power for outcome (OR = 3.7, 95% CI 1.3–11; P = 0.02), perhaps owing to the predominance (80%) of ER-α-positive tumors in the good prognosis group (Fig. 3). Likewise, in their independent test set, all of the tumors with a good prognosis were ER-α-positive except one (86%; Fig. 4). In addition, 71% of the 231 prognostic genes identified by van’t Veer et al. were also listed by them as ER-α status reporter genes, confirming an overlap between the predictors of prognosis and ER-α status. It is important to note, however, that they were able to achieve better prognostic predictions using the 231 selected genes (OR = 15, 95% CI 4–56; P = 0.000004) than with the ER-α predictor gene set. Nevertheless, we believe that the correlation between prognosis and ER-α status in samples reported by van’t Veer et al. (Figs 3 and 4) might have led to the selection of a prognostic gene set that may not be broadly applicable to other breast tumor cohorts in which no correlation between ER-α status and prognosis exists.
Conclusion

Most breast cancer gene expression studies to date have examined limited sample numbers and tend not to have sufficient power to allow analysis within standard clinico-pathologically defined subsets. Moreover, the majority of breast cancer patients today receive some form of postoperative treatment (radiation, endocrine and/or chemotherapy) that influences the clinical course and significance of prognostic factors. In general, there are risks inherent in the development of a prognostic classifier from a sample set comprised of patients who are heterogeneous for clinico-pathological variables of known prognostic significance. Although the resulting predictive classifier may accurately describe the sample set used for its development, that result may merely reflect the influence of known prognostic variables on gene expression, and it may fail when applied to an independent test set containing patients who are indistinguishable on the basis of all standard clinical and pathological markers, but who still differ in terms of outcome. Although it is still of great importance to gain an understanding of which patterns of gene expression are linked to known variables such as ER-α status, studies should be designed to reveal rather than obscure these links, and to uncover any potential gene expression patterns that predict outcome within uniform groups. If microarray based prognostic tools are to become widely used in the clinical setting, then additional studies must be conducted that seek and independently validate prognostic information in larger sample sets, which are carefully selected and matched with respect to known prognostic variables, such as ER-α status and therapy.
References

1. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999, 286:531-537.

2. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J Jr, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Staudt LM, et al.: Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000, 403:503-511.

3. Bittner M, Meltzer P, Chen Y, Jiang Y, Seltor E, Hendrix M, Radmacher M, Simon R, Yakhini Z, Ben-Dor A, Sampas N, Dougherty E, Wang E, Marincola F, Gooden C, Lueders J, Glatfelter A, Pollock P, Carpten J, Gillanders E, Leja D, Dietrich K, Beaudry C, Berens M, Alberts D, Sondak V: Molecular classification of cutaneous malignant melanoma by gene expression profiling. Nature 2000, 406:536-540.

4. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Aas T, Sae A, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Bittner M, Szent-Györgyi A, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D: Molecular portraits of human breast tumours. Nature 2000, 406:747-752.

5. Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu CR, Peterson C, Meltzer PS: Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nat Med 2001, 7:673-679.

6. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hiest T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Bittner M, Szent-Györgyi A, Zhu SX, Lonning PE, Borresen-Dale AL: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci USA 2001, 98:10869-10874.

7. van ’t Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH: Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002, 415:530-536.

8. Gruvberger S, Ringner M, Chen Y, Panavally S, Saal LH, Borg A, Pettersson M, Peterson C, Meltzer PS: Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. Cancer Res 2001, 61:5979-5984.

9. West M, Blanchette C, Dressman H, Huang E, Ishida S, Spang R, Zuzan H, Olson JA Jr, Marks JR, Nevins JR: Predicting the clinical status of human breast cancer by using gene expression profiles. Proc Natl Acad Sci USA 2001, 98:11462-11467.

10. Allander SV, Nupponen NN, Ringner M, Hostetter G, Maher GW, Goldberger N, Chen Y, Carpten J, Elkahloun AG, Meltzer PS: Gastrointestinal stromal tumors with KIT mutations exhibit a remarkably homogeneous gene expression profile. Cancer Res 2001, 61:8624-8628.

11. Khan J, Simon R, Bittner M, Chen Y, Leighton SB, Pohida T, Smith PD, Jiang Y, Gooden GC, Trent JM, Meltzer PS: Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. Cancer Res 1999, 59:5009-5013.

Correspondence
Paul S Meltzer, Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MA 20892-8000, USA. Tel: +1 301 594 5283; fax: +1 301 480 3281; e-mail: pmeltzer@mail.nih.gov