

## SHORT COMMUNICATION

# Molecular profiling contributes more than routine histology and immunohistochemistry to breast cancer diagnostics

Christine Shiang and Lajos Pusztai\*

### Introduction

Initial enthusiasm to explore gene expression profiling and other high-throughput molecular methods as molecular diagnostic tools for breast cancer has given way to increasing skepticism. Several investigators suggested that these novel analytical methods may not have advanced diagnostic medicine beyond what optimally performed histology and immunohistochemistry (IHC) could deliver. There is some truth in this criticism, particularly when it comes to clinically useful assays that gene expression profiling methods have yielded. However, this overly simplistic assessment of molecular profiling overlooks three important contributions that high-throughput gene expression analysis has brought to breast cancer research and treatment. First, results from gene expression profiling studies have fundamentally changed our conceptual approach to breast cancer. Second, these methods have yielded several commercially available new diagnostic tests that fill a previously unmet diagnostic niche and have started to impact routine care, at least in the United States. Third, the impact of the large volume of molecular data that these studies have generated will have a lasting impact on breast cancer research. The in-depth analysis of the many tantalizing observations made from comprehensive genomic characterization of breast cancers has barely begun.

### The conceptual impact of molecular profiling on breast cancer research

Gene expression profiling provided the first glimpse of the true complexity of the molecular machinery of breast cancer. The earliest of these studies already revealed large-scale molecular differences between estrogen receptor (ER)-positive and ER-negative cancers and also

revealed two robust subsets within the ER-positive cancers [1]. These molecular differences between breast cancer subsets together with the important clinical differences that also distinguish these groups have prompted a conceptual shift in the classification of breast cancer. Breast cancer is no longer considered a single disease with variable ER expression, histology and grade but a collection of genuinely different neoplastic diseases that arise from the breast epithelium. The long recognized heterogeneity in ER expression and grade have not led to similar shifts in classification in the past because the scale of molecular differences that exist between these disease types remained hidden until high-throughput molecular analytical methods become available. The implications of the new classification schema reach far beyond a simple ER-based stratification of breast cancer. Different molecular types of breast cancers will require separate clinical trials, different prognostic and predictive markers and different therapeutic strategies. Continued conduct of studies that include all types of breast cancers is akin to conducting a trial for leukemia where patients with acute myeloid leukemia, acute lymphocytic leukemia, chronic lymphocytic leukemia, chronic myeloid leukemia, and so on, are all eligible for participation. Results from such studies would be unstable and have limited practical value considering the vast clinical and molecular differences between these different types of leukemias. Indeed, the next generation of therapeutic and biomarker studies in breast cancer are increasingly being targeted to molecularly defined subsets such as triple-negative/basal-like or ER-positive high risk (Luminal A or MammaPrint or Oncotype Dx high risk groups) breast cancers.

### The impact of molecular profiling on molecular diagnostics

Molecular profiling simultaneously measures a large number of variables (that is, gene expression values, DNA copy numbers or single nucleotide polymorphisms) and the simplest goal of the analysis is to find individual variables that are associated with a disease subset or

\*Correspondence: [lpusztai@mdanderson.org](mailto:lpusztai@mdanderson.org)  
Department of Breast Medical Oncology, University of Texas MD Anderson Cancer Center, Houston, TX 77230-1439, USA

clinical outcome group. This analysis strategy brought into focus two very important statistical concepts, long neglected in traditional biomarker research: the importance of combining individual, independent markers into multivariate prediction models; and the need to guard against false discovery due to multiple comparisons. Invariably, more than one marker is associated with any particular outcome or disease subset. Historically, markers were used as stratification tools and classification schemas were either restricted to a single marker (that is, groups were defined as marker high versus marker low) or multiple markers were used as sequential stratification tools. However, subsetting of patients through multiple layers of dichotomous markers is not practical and leads to unstable results due to the rapidly diminishing numbers in the subsets. The statistically optimal use of independent variables is to construct a multivariate prediction model. Despite close to four decades of IHC literature, very few papers describe correctly trained and independently validated multivariate prediction models. This has started to change lately, due to the impact of molecular profiling studies, and will undoubtedly increase the value of IHC-based tests through combining multiple different IHC markers into more powerful prediction models.

Molecular profiling also brought into the forefront the importance of guarding against false positive discoveries due to multiple comparisons. When only a single marker is assessed and a 5% significance level is applied to the statistical test, there is only a 5% chance of incorrectly rejecting the null hypothesis (that is, lack of association between a marker and an outcome) if the null hypothesis is correct. However, if one performs 100 independent tests where all null hypotheses are correct, the expected number of false positive findings is 5. The probability of finding at least one positive association among the 100 tests is virtually 100% even if none of the markers is associated with the endpoint. Historically, IHC studies often evaluated multiple different markers and the same marker may have been correlated with several different endpoints, yet adjustments for the multiple hypothesis testing were rarely performed. More recently, investigators and journal editors started to require such adjustments even in 'low throughput' multiple comparison studies, which will raise the level of evidence that these analyses produce.

Perhaps the most important practical contribution of genomics to breast cancer management was the development of multi-gene assays (Oncotype Dx, MammaPrint, Genomic Grade Index) that can distinguish low and high risk prognostic groups among ER-positive, early stage breast cancers [2,3]. In the past, selection of adjuvant chemotherapy for ER-positive cancers was based on tumor size, nodal status, histologic grade, patient

preference, and comorbid illnesses. However, none of these variables, with the exception of grade, has a consistent association with sensitivity to chemotherapy or endocrine therapy. Combination of these variables into a summary recommendation about therapy is subjective and frequently leads to variable recommendations by different physicians. Multivariate genomic assays that take input from ER and human epidermal growth factor receptor 2 (HER2) expression as well as from a number of proliferation-related genes can stratify ER-positive cancers into low and high risk prognostic groups objectively and this information is additive or complementary to prognostic risk based on tumor size and nodal status. These assay results can also inform about general chemotherapy sensitivity [4,5]. It has also become apparent that the most important prognostic and predictive component of these first generation assays comes from their ability to measure proliferation reproducibly and quantitatively [6]. Hence, simpler proliferation measurements may accomplish the same. However, an important feature of these commercially available tests is that they are standardized and validated in multiple independent studies. Multi-IHC tests, including ER, progesterone receptor, HER2, Ki67 (and other genes) may accomplish similar risk stratification in the future but, currently, despite over 25 years of research, no standardized and externally validated IHC-based risk stratification assay exists for breast cancer.

### **Future potential**

Molecular profiling is uniquely suited for multiplex assays. A single assay from one needle biopsy specimen can generate a large number of data points. A variable assortment of different genes (or other molecular variable) can be used to issue simultaneously several different prognostic or predictive results. The cost of gene expression analysis has dropped substantially over the past few years and the analytical validity of the different platforms is well established [7,8]. Gene expression results and other molecular readouts are quantitative over a relatively broad dynamic range and can easily be fed into standardized computer prediction algorithms. On the other hand, multiplex IHC assays are cumbersome to perform and the quantification of the signal is still not standardized across pathology laboratories. It is hard to imagine that one could perform more than a few multi-IHC tests on the same case if each tests relies on measuring four to six different antigens that require individual sections and separate scoring. Combining the individual IHC results into several different multi-IHC scores is not well suited for automation and could be error prone if performed by humans. The future of molecular profiling as a diagnostic tool will depend to a large extent on the content that can

be generated by these platforms. If new and clinically useful, validated predictive signatures can be developed, molecular profiling has a bright future as diagnostic technology. The more such signatures exist, the greater the utility of a high-throughput, standardized, easily automated platform.

Finally, the impact of the large volumes of public data that molecular profiling studies have generated cannot be compared with the impact of IHC studies that measure the expression of one or a handful of genes [9]. The in-depth analysis of the many tantalizing observations made from comprehensive genomic characterization of breast cancers has barely begun and may ultimately represent the most important future contribution of molecular profiling to cancer research.

#### Abbreviations

ER = estrogen receptor; HER2 = human epidermal growth factor receptor 2; IHC = immunohistochemistry.

#### Competing interests

The authors declare that they have no competing interests.

#### Acknowledgements

This article has been published as part of *Breast Cancer Research* Volume 12 Supplement 4, 2010: Controversies in Breast Cancer 2010. The full contents of the supplement are available online at <http://breast-cancer-research.com/supplements/12/S4>

Published: 20 December 2010

#### References

1. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lønning PE, Børresen-Dale AL, Brown PO, Botstein D: **Molecular portraits of human breast tumours.** *Nature* 2000, **406**:747-752.

2. Sotiriou C, Pusztai L: **Gene-expression signatures in breast cancer.** *N Engl J Med* 2009, **360**:790-800.
3. Marchionni L, Wilson RF, Wolff AC, Marinopoulos S, Parmigiani G, Bass EB, Goodman SN: **Systematic review: gene expression profiling assays in early-stage breast cancer.** *Ann Int Med* 2008, **148**:358-369.
4. Liedtke C, Hatzis C, Symmans WF, Desmedt C, Haibe-Kains B, Valero V, Kuerer H, Hortobagyi GN, Piccart-Gebhart M, Sotiriou C, Pusztai L: **Genomic grade index is associated with response to chemotherapy in patients with breast cancer.** *J Clin Oncol* 2009, **27**:3185-3191.
5. Gianni L, Zambetti M, Clark K, Baker J, Cronin M, Wu J, Mariani G, Rodriguez J, Carcangiu M, Watson D, Valagussa P, Rouzier R, Symmans WF, Ross JS, Hortobagyi GN, Pusztai L, Shak S: **Gene expression profiles in paraffin-embedded core biopsy tissue predict response to chemotherapy in women with locally advanced breast cancer.** *J Clin Oncol* 2005, **23**:7265-77.
6. Wirapati P, Sotiriou C, Kunkel S, Farmer P, Pradervand S, Haibe-Kains B, Desmedt C, Ignatiadis M, Sengstag T, Schütz F, Goldstein DR, Piccart M, Delorenzi M: **Meta-analysis of gene-expression profiles in breast cancer: toward a unified understanding of breast cancer sub-typing and prognosis signatures.** *Breast Cancer Res* 2008, **10**:R65.
7. MAQC Consortium, Shi L, Reid LH, Jones WD, Shippy R, Warrington JA, Baker SC, Collins PJ, de Longueville F, Kawasaki ES, Lee KY, Luo Y, Sun YA, Willey JC, Setterquist RA, Fischer GM, Tong W, Dragan YP, Dix DJ, Frueh FW, Goodsaid FM, Herman D, Jensen RV, Johnson CD, Lobenhofer EK, Puri RK, Schrf U, Thierry-Mieg J, Wang C, Wilson M, et al.: **The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements.** *Nat Biotechnol* 2006, **24**:1151-1161.
8. Shi L, Campbell G, Jones WD: **The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models.** *Nat Biotechnol* 2010, **28**:827-838.
9. Pusztai L, Leyland-Jones B: **Promises and caveats of in-silico biomarker discovery.** *Br J Cancer* 2008, **99**:385-386.

doi:10.1186/bcr2735

**Cite this article as:** Shiang C, Pusztai L: **Molecular profiling contributes more than routine histology and immunohistochemistry to breast cancer diagnostics.** *Breast Cancer Research* 2010, **12**(Suppl 4):S6.