

Review

High-throughput genomic technology in research and clinical management of breast cancer

Molecular signatures of progression from benign epithelium to metastatic breast cancer

Karin Rennstam and Ingrid Hedenfalk

Department of Oncology, Clinical Sciences, Lund University, Lund, Sweden

Corresponding author: Ingrid Hedenfalk, Ingrid.Hedenfalk@med.lu.se

Published: 2 August 2006

This article is online at <http://breast-cancer-research.com/content/8/4/213>

© 2006 BioMed Central Ltd

Breast Cancer Research 2006, **8**:213 (doi:10.1186/bcr1528)

Abstract

It is generally accepted that early detection of breast cancer has great impact on patient survival, emphasizing the importance of early diagnosis. In a widely recognized model of breast cancer development, tumor cells progress through chronological and well defined stages. However, the molecular basis of disease progression in breast cancer remains poorly understood. High-throughput molecular profiling techniques are excellent tools for the study of complex molecular alterations. By accurately mapping changes in the genome and subsequent biological/molecular pathways, the chances of finding potential novel treatment targets as well as intervention strategies are enhanced, and ultimately lives can be saved. This review provides a brief summary of recent progress in identifying molecular markers for invasiveness in early breast lesions.

Introduction

The commonly accepted model for cancer development is that cancer cells, over a long period of time, acquire the hallmarks of malignancy (e.g. oncogene activation and loss of tumor suppressor gene function) [1]. The vast majority of breast neoplasia arises in the ductal epithelial cells, and is generally believed to be a chronological progression through defined clinical and pathological stages. These stages start with premalignant atypical ductal hyperplasia (ADH), which may progress to preinvasive ductal carcinoma *in situ* (DCIS), followed by a possible progression to invasive ductal carcinoma (IDC) and culmination in metastatic disease [2]. Atypical lobular hyperplasia and lobular carcinoma *in situ*, comprising a small proportion of all breast neoplasia, are breast cancer risk factors and constitute nonmandatory precursors for the successive development of invasive carcinoma in either breast, of either ductal or lobular type [3].

Alternative pathways to development of invasive breast cancer have been suggested (for review, see the report by Simpson and coworkers [4]), emphasizing the importance of studying benign proliferative and preinvasive breast lesions in relation to invasive disease. Finding molecular markers of invasive or metastatic potential in early stage lesions would therefore have considerable impact on breast cancer diagnosis, treatment and prognosis.

Although comparative genomic hybridization (CGH) and loss of heterozygosity have provided compelling evidence that ADH and DCIS are precursors to IDC, the molecular basis of progression in early stages of breast cancer remains poorly understood [2]. This is partly due to tumor heterogeneity, with intra-tumor as well as inter-tumor variation based on, for instance, varying grades of mitotic activity, cellular differentiation and presence of normal and inflammatory cells.

There is evidence to support the notion that breast cancer arises from mutated mammary stem/progenitor cells, which have been termed 'breast cancer stem cells' because of their exclusive ability to maintain tumor formation and growth, as reviewed by Behbod and coworkers [5]. Al-Hajj and colleagues [6] were the first group to identify this population of highly tumorigenic cells in human breast tumor isolates. When transplanted into NOD/SCID mice, as few as 100 of these cells were able to form tumors.

High throughput genome-wide array based techniques such as array-CGH and transcriptional profiling provide an opportunity to discover genes and/or pathways that are specifically activated or inactivated during tumor progression. This review

ADH = atypical ductal hyperplasia; BAC = bacterial artificial chromosome; CGH = comparative genomic hybridization; DCIS = ductal carcinoma *in situ*; FISH = fluorescence *in situ* hybridization; IDC = invasive ductal carcinoma; LOH = loss of heterozygosity; SAGE = serial analysis of gene expression..

Table 1**Glossary of terms used in the present review**

Term	Description
Microarray	Microscope slide or chip with immobilized probes, usually cDNA, BAC, or oligo probes
cDNA	Complementary DNA: a DNA that is complementary to a given RNA, which serves as a template for synthesis of the DNA in the presence of reverse transcriptase. When printed on microarray slides, they are usually 0.5-2 kilobases (kb)
BAC	Bacterial artificial chromosome: a DNA construct, based on a fertility plasmid, used for transforming and cloning in bacteria, usually <i>Escherichia coli</i> . Its usual insert size is 150 kb, with a range from 100 to 300 kb
Oligo	Oligonucleotide: a relatively short single-stranded nucleic acid chain, usually 20-80 base pairs
Gene expression profiling	Global analysis of transcribed RNA in a tissue of interest. Usually based on the cohybridization of differentially labeled test and reference cDNAs to probes on microarray slides
SAGE	Serial analysis of gene expression: generates a library of expressed genes by taking a raw count of sequence tags, each representing a transcript in the RNA population
CGH	Comparative genomic hybridization: comparative hybridization of differentially labeled DNA of interest (e.g. tumor) and a reference (any genomic DNA) to normal metaphase chromosomes
Array-CGH	Array-based CGH: comparative hybridization of differentially labeled DNA of interest (e.g. tumor) and a reference (any genomic DNA) to probes on a microarray slide

focuses on the efforts that have been made to find molecular markers for invasiveness in early breast lesions and to find metastasis-associated gene signatures that are present in early tumorigenesis. A glossary of terms used is provided in Table 1.

Tissue heterogeneity

Tumors as well as precancerous lesions are heterogeneous cell populations that harbor normal stromal and inflammatory cells in addition to cancer cells. The presence of these nonmalignant cells could mask the detection of genetic and transcriptional alterations in cancer cells. However, recent advances in cell isolation techniques, reviewed by Eltoum and coworkers [7], permit precise isolation of separate cell populations for individual analysis.

Precancerous or early stage breast cancer lesions are diminutive, and the amount of material obtained from these small specimens is often far less than is required for genome-wide analyses. This problem can be addressed by the use of amplification techniques such as degenerate oligonucleotide primed polymerase chain reaction, which permits relatively uniform amplification of the entire genome. Similarly, RNA can be linearly amplified, for instance with T7 polymerase, for global gene expression studies. However, these methods of amplification can result in misrepresentation of certain genes or genomic regions.

During the past few years the role of the cellular micro-environment in tumorigenesis has become an intense area of research. This is partly due to studies demonstrating that genetic abnormalities occur not only in cancer cells but also in stromal cells [8]. Moreover, Kurose and coworkers [9]

demonstrated high frequencies of somatic mutations in *TP53* (encoding tumor protein p53) and *PTEN* (encoding phosphate and tensin homolog) in both breast neoplastic epithelium and stroma. Ding and colleagues [10] recently evaluated the level of EZH2 protein (a transcriptional repressor that is involved in controlling cellular memory) in breast tissue samples (normal, ADH and DCIS lesions). EZH2 expression was elevated in ADH compared with normal epithelium, and was even higher in DCIS. Of interest, histologically normal lobules adjacent to ADH and DCIS exhibited a significantly increased number of cells expressing EZH2 when compared with distant normal lobules, indicating that elevated levels of EZH2 protein expression can detect a precancerous state in morphologically normal breast epithelium. In an elegant study, Allinen and coworkers [11] described comprehensive transcriptional profiles of each individual cell type composing normal breast tissue and *in situ* and invasive breast carcinoma (including epithelial cells, leukocytes, myofibroblasts and myoepithelial cells, and endothelial and stromal cells) using a sequential isolation protocol for each cell population combined with serial analysis of gene expression (SAGE). They determined that extensive transcriptional changes occur in all cell types during cancer progression, but genomic alterations were detected only in epithelial cancer cells. Molecular characterization of each constituent cell type will contribute to our understanding of the role played by these cells in breast tumorigenesis, and may also provide new molecular targets for breast cancer intervention and treatment.

Genome-wide molecular profiling applications

The analysis of gene expression profiles can give insights into changes in biochemical pathways that occur during malignant

transformation and cancer progression. Transcriptional profiles corresponding to thousands of genes can be established simultaneously by global profiling techniques (i.e. microarrays and SAGE). Microarray techniques are based on the hybridization of cDNA samples to immobilized probes on microarray slides, whereas the SAGE method generates a library of expressed genes by taking a raw count of sequence tags, each representing a transcript in the RNA population. Absolute quantification of the expressed genes permits the creation of gene expression profiles that can be compared with libraries from other cell or tissue types [12]. A major advantage of SAGE over microarrays is that it does not require any prior knowledge of the sequences to be analyzed. However, microarrays are more amenable to the analysis of large sample sets. Of interest, a study comparing SAGE and microarray data [13] revealed a good correlation between the two techniques.

Array-based CGH can be used to identify high-resolution global genomic changes acquired during cancer progression. In array-CGH differentially labelled test DNA (e.g. tumor) and normal control DNA is co-hybridized onto a representation of the genome, consisting of a multitude of printed spots of target DNA. Arrays made from cDNA have been used most often for this purpose [14], but the use of cDNA clones as targets for genomic DNA is hampered by the suboptimal hybridization of genetic material present in introns in the genomic DNA but absent in cDNAs. Bacterial artificial chromosome (BAC) arrays, on the other hand, utilize segments of human genomic DNA as hybridization targets; 32k tiling BAC arrays provide an average resolution of about 80 kilobases [15]. High-density oligonucleotide arrays have a higher resolution of regions of interest than do BAC arrays, but they are usually nontiling [16]. Custom-made arrays are commercially available from several vendors, and these enable individual probe design of single exon resolution.

Gene expression and breast cancer classification

Gene expression profiling has proven to be a useful and reliable tool for classifying breast cancers into subgroups that reflect different histopathological characteristics as well as differential prognostic outcome. It has been suggested that estrogen receptor negative and positive breast cancers can be subdivided into Her-2 positive basal-epithelial like, normal breast-like and luminal-like [17]. The potentially different origins of the tumor cells may signify distinct pathways of tumorigenesis and differences in the clinical course of the disease.

Germ-line mutations in the *BRCA1* and *BRCA2* genes together account for a significant portion of hereditary breast cancers. They have been shown to leave a characteristic imprint on the panel of genes expressed by the tumors [18], with *BRCA1*-dependent tumors exhibiting a transcriptional profile similar to the basal subtype of tumors [19]. These

findings suggest that the cellular origin of *BRCA1* and *BRCA2* mutation positive tumors may differ, or that these tumors traverse down separate pathways in their progression toward malignancy [18]. Furthermore, the molecular subclassification of non-*BRCA1/2* familial breast cancers into homogeneous subgroups underscores the potential differences in cellular origin and/or disease progression due to the presence of multiple diverse underlying genetic alterations, which is reflected in the phenotype of the tumors [20].

Transcriptional profiling of premalignant and early stage breast cancer

Using SAGE analysis on a small set of normal breast tissues, DCIS and IDC tumors, Abba and coworkers [21] detected significant changes that occur during the course of breast cancer progression. They were also able to identify genes and gene families commonly deregulated across samples within each specific stage in the transition from benign breast tissue to IDC. By comparing differential gene expression profiles established by cDNA microarrays between normal cells, primary invasive carcinoma and metastatic cells, Mimori and coworkers [22] were able to detect genes directly associated with each tumor stage in tumor development and gave clues to the comprehensive identification of metastasis-related genes in clinical breast cancer biopsies. In contrast, using the combination of laser capture microdissection and DNA microarrays to generate gene expression profiles of premalignant, preinvasive and invasive stages of human breast cancer, Ma and colleagues [23] discovered extensive similarities across the distinct stages of progression, suggesting that transcriptional alterations granting the potential for invasive growth are already present in the preinvasive stages. Interestingly, they found that different histological grades were associated with distinct gene expression signatures, and that a subset of genes associated with high histological grade was correlated with the transition from preinvasive to invasive growth. In accordance with this, Weigelt and coworkers [24] showed that distant metastases exhibit both the same breast cancer subtype and transcriptional signature as their primary tumors, which was interpreted by some as the capacity to metastasize being an inherent feature of most breast cancers.

Several studies attempting to classify breast tumors into good or poor prognosis categories have been reported. Strikingly, very few genes are found in common among these independent gene signatures. Although this may in part be explained by the use of different microarray platforms, among other differences, it has become increasingly evident that further data from well designed trials are needed to identify key determinants before these diagnostic techniques may be introduced into the clinical setting [25]. Nonetheless, these studies have shown us that stratification of breast tumors by clinicopathological and transcriptional profiles before determination of prognostic and treatment predictive genetic signatures may be the most effective approach to achieve

improved and tailored clinical management. Importantly, histological grade, largely coinciding with hormone receptor status, strongly reflects the magnitude and type of genetic aberrations in invasive breast cancers (for review, see the report by Simpson and coworkers [4]), emphasizing the correlation between genotype and phenotype during disease progression. These findings stress the need to combine histopathological parameters with molecular profiling techniques for translation into clinical practice.

Genetic aberrations in premalignant and early stage breast lesions

A multitude of molecular studies have been performed in DCIS and IDC tumors with the common goal of identifying genes involved in the initiation of sporadic disease, and investigation of the link between *in situ* and invasive carcinoma. Lukas and coworkers [26] found that the frequency of *TP53* mutations in DCIS was similar to that found in invasive tumors. Moreover, the *in situ* and invasive components exhibited identical mutations, reinforcing the clonal relationship between *in situ* and invasive lesions. Upon investigation of *HER-2* in a cohort of women diagnosed with benign breast disease, Stark and colleagues [27] concluded that women with benign breast biopsies exhibiting both *HER-2* amplification and a proliferative histopathological lesion may be at substantially increased risk for developing subsequent breast cancer. Overexpression of the *HER-2/neu* protein in otherwise benign biopsies may indicate a further increase in risk. Moreover, several studies have analyzed the identity and distribution of chromosomal alterations in ductal hyperplasias and *in situ* and invasive carcinomas. In general, more advanced tumors exhibit more genetic changes, although many of the changes are already present in *in situ* carcinomas or even in ductal hyperplasia, suggesting a progressive accumulation of genomic aberrations.

Combining molecular approaches

The combination of array-CGH and gene expression profiling is probably one of the most reliable and comprehensive ways to find new marker genes for breast cancer progression and metastasis. In a recent study conducted by Yao and coworkers [28], including DCIS, IDC and lymph node metastases, the authors identified 49 minimal commonly amplified regions, including known (1q, 8q24, 11q13, 17q21-q23 and 20q13) and previously uncharacterized regions (12p13 and 16p13). They confirmed that the overall frequency of copy number aberrations was higher in invasive tumors than in DCIS, with several aberrations occurring only in invasive cancer. By combining array-CGH and SAGE data they were able to distinguish a number of putative breast cancer oncogenes.

Ultimately, the genome-wide search for genes and biochemical pathways or networks causing phenotypic changes during breast tumorigenesis will require the integration of both genomic, transcriptional, and proteomic approaches.

This article is part of a review series on *High-throughput genomic technology in research and clinical management of breast cancer*, edited by Yudi Pawitan and Per Hall.

Other articles in the series can be found online at http://breast-cancer-research.com/articles/review-series.asp?series=bcr_Genomic

Finding pathways and networks that are involved in cancer progression when interpreting data from genome-wide analyses can be tremendously complex, and therefore gene ontology tools can be invaluable. Validation experiments of the results from genome-wide screens must be performed using molecular techniques such as immunohistochemistry, fluorescent *in situ* hybridization, or chromogenic *in situ* hybridization. For such purposes the use of tissue microarray technology has proven useful. This technique allows for simultaneous analysis of several hundreds of samples in a single staining experiment [29]. Also, it has become increasingly evident that epigenetic changes must be taken into consideration in the investigation of breast cancer aetiology. Yang and coworkers [30] showed that methylation changes occur not only in tumor cells but also in normal breast tissue as far as 4 cm away from primary tumor sites. Functional studies using cell line or animal models to investigate the role of individual genes or gene products may shed further light on the events that underlie malignant transformation and disease progression.

Conclusion

We conclude that high throughput genomic and gene expression analyses have proven to be valuable tools for identifying putative molecular markers for tumor development and metastatic potential. It is important to verify these findings with other molecular techniques as well as in large clinical trials. Moreover, functional validation of causal relationships between genetic alterations and disease aetiology would increase our biological understanding of breast tumorigenesis, in addition to providing molecular targets for intervention, diagnosis and treatment.

Competing interests

The authors declare that they have no competing interests.

References

1. Hanahan D, Weinberg RA: **The hallmarks of cancer.** *Cell* 2000, **100**:57-70.
2. Allred DC, Mohsin SK, Fuqua SA: **Histological and biological evolution of human premalignant breast disease.** *Endocr Relat Cancer* 2001, **8**:47-61.
3. Ellis IO, Schnitt SJ, Sastre-Garau X, Bussolati G, Tavassoli FA, Eusebi V, Peterse JL, Mukai K, Tabár L, Jacquemier J, et al.: **Tumours of the breast.** In *Pathology and Genetics of Tumours of the Breast and Female Genital Organs*. Edited by Jaffe ES, Stein H, Vardiman JW. Lyon, France: IARC Press; 2001:9-110.
4. Simpson PT, Reis-Filho JS, Gale T, Lakhani SR: **Molecular evolution of breast cancer.** *J Pathol* 2005, **205**:248-254.

5. Behbod F, Rosen JM: **Will cancer stem cells provide new therapeutic targets?** *Carcinogenesis* 2005, **26**:703-711.
6. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF: **Prospective identification of tumorigenic breast cancer cells.** *Proc Natl Acad Sci USA* 2003, **100**(7):3983-3988.
7. Eltoum IA, Siegal GP, Frost AR: **Microdissection of histologic sections: past, present, and future.** *Adv Anat Pathol* 2002, **9**: 316-322.
8. Kurose K, Hoshaw-Woodard S, Adeyinka A, Lemeshow S, Watson PH, Eng C: **Genetic model of multi-step breast carcinogenesis involving the epithelium and stroma: clues to tumour-microenvironment interactions.** *Hum Mol Genet* 2001, **10**:1907-1913.
9. Kurose K, Gilley K, Matsumoto S, Watson PH, Zhou XP, Eng C: **Frequent somatic mutations in PTEN and TP53 are mutually exclusive in the stroma of breast carcinomas.** *Nat Genet* 2002, **32**:355-357.
10. Ding L, Erdmann C, Chinnaiyan AM, Merajver SD, Kleer CG: **Identification of EZH2 as a molecular marker for a precancerous state in morphologically normal breast tissues.** *Cancer Res* 2006, **66**:4095-4099.
11. Allinen M, Beroukhi R, Cai L, Brennan C, Lahti-Domenici J, Huang H, Porter D, Hu M, Chin L, Richardson A, et al.: **Molecular characterization of the tumor microenvironment in breast cancer.** *Cancer Cell* 2004, **6**:17-32.
12. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW: **Serial analysis of gene expression.** *Science* 1995, **270**:484-487.
13. Ishii M, Hashimoto S, Tsutsumi S, Wada Y, Matsushima K, Kodama T, Aburatani H: **Direct comparison of GeneChip and SAGE on the quantitative accuracy in transcript profiling analysis.** *Genomics* 2000, **68**:136-143.
14. Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D, Brown PO: **Genome-wide analysis of DNA copy-number changes using cDNA microarrays.** *Nat Genet* 1999, **23**:41-46.
15. Ishkanian AS, Malloff CA, Watson SK, DeLeeuw RJ, Chi B, Coe BP, Snijders A, Albertson DG, Pinkel D, Marra MA, et al.: **A tiling resolution DNA microarray with complete coverage of the human genome.** *Nat Genet* 2004, **36**:299-303.
16. Lucito R, Healy J, Alexander J, Reiner A, Esposito D, Chi M, Rodgers L, Brady A, Sebat J, Troge J et al.: **Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation.** *Genome Res* 2003, **13**:2291-2305.
17. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, et al.: **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.** *Proc Natl Acad Sci USA* 2001, **98**:10869-10874.
18. Hedenfalk I, Duggan D, Chen Y, Radmacher M, Bittner M, Simon R, Meltzer P, Gusterson B, Esteller M, Kallioniemi OP, et al.: **Gene-expression profiles in hereditary breast cancer.** *N Engl J Med* 2001, **344**:539-548.
19. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, et al.: **Repeated observation of breast tumor subtypes in independent gene expression data sets.** *Proc Natl Acad Sci USA* 2003, **100**: 8418-8423.
20. Hedenfalk I, Ringner M, Ben-Dor A, Yakhini Z, Chen Y, Chebil G, Ach R, Loman N, Olsson H, Meltzer P, et al.: **Molecular classification of familial non-BRCA1/BRCA2 breast cancer.** *Proc Natl Acad Sci USA* 2003, **100**:2532-2537.
21. Abba MC, Drake JA, Hawkins KA, Hu Y, Sun H, Notcovich C, Gaddis S, Sahin A, Baggerly K, Aldaz CM: **Transcriptomic changes in human breast cancer progression as determined by serial analysis of gene expression.** *Breast Cancer Res* 2004, **6**:R499-R513.
22. Mimori K, Kataoka A, Yoshinaga K, Ohta M, Sagara Y, Yoshikawa Y, Ohno S, Barnard GF, Mori M: **Identification of molecular markers for metastasis-related genes in primary breast cancer cells.** *Clin Exp Metastasis* 2005, **22**:59-67.
23. Ma XJ, Salunga R, Tuggle JT, Gaudet J, Enright E, McQuary P, Payette T, Pistone M, Stecker K, Zhang BM, et al.: **Gene expression profiles of human breast cancer progression.** *Proc Natl Acad Sci USA* 2003, **100**:5974-5979.
24. Weigelt B, Hu Z, He X, Livasy C, Carey LA, Ewend MG, Glas AM, Perou CM, Van't Veer LJ: **Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer.** *Cancer Res* 2005, **65**:9155-9158.
25. Brenton JD, Carey LA, Ahmed AA, Caldas C: **Molecular classification and molecular forecasting of breast cancer: ready for clinical application?** *J Clin Oncol* 2005, **23**:7350-7360.
26. Lukas J, Niu N, Press MF: **p53 mutations and expression in breast carcinoma in situ.** *Am J Pathol* 2000, **156**:183-191.
27. Stark A, Hulka BS, Joens S, Novotny D, Thor AD, Wold LE, Schell MJ, Melton LJ, 3rd, Liu ET, Conway K: **HER-2/neu amplification in benign breast disease and the risk of subsequent breast cancer.** *J Clin Oncol* 2000, **18**:267-274.
28. Yao J, Weremowicz S, Feng B, Gentleman RC, Marks JR, Gelman R, Brennan C, Polyak K: **Combined cDNA array comparative genomic hybridization and serial analysis of gene expression analysis of breast tumor progression.** *Cancer Res* 2006, **66**: 4065-4078.
29. Kononen J, Bubendorf L, Kallioniemi A, Barlund M, Schraml P, Leighton S, Torhorst J, Mihatsch MJ, Sauter G, Kallioniemi OP: **Tissue microarrays for high-throughput molecular profiling of tumor specimens.** *Nat Med* 1998, **4**:844-847.
30. Yan PS, Chinnambally V, Ibrahim A, Liu JC, Diaz N, Centeno B, Weber F, Leu YH, Eng C, Yeatman TJ, et al.: **Mapping geographic zones of cancer risk with epigenetic biomarkers in normal breast tissue [abstract].** In *Proceedings of AACR 97th Annual Meeting; 1-5 April 2006; Washington, DC.* Philadelphia: American Association for Cancer Research; 2006.