Impact of DNA Amplification on Gene Expression Patterns in Breast Cancer

Elizabeth Hyman,1 Päiviikki Kauraniemi,2 Sampsa Hautaniemi, Maija Wolf, Spyro Mousses, Ester Rozenblum, Markus Ringnér, Guido Sauter, Outi Monni, Abdel Elkahloun, Olli-P. Kallioniemi, and Anne Kallioniemi

Howard Hughes Medical Institute-NIH Research Scholar, Bethesda, Maryland 20892 [E. H.]; Cancer Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland 20892 [E. H., P. K., S. H., M. W., S. M., E. R., M. R., A. E., O. K., A. K.]; Laboratory of Cancer Genetics, Institute of Medical Technology, University of Tampere and Tampere University Hospital, FIN-33200 Tampere, Finland [P. K., A. K.]; Signal Processing Laboratory, Tampere University of Technology, FIN-33101 Tampere, Finland [S. H.]; Institute of Pathology, University of Basel, CH-4003 Basel, Switzerland [G. S.]; and Biomedicum Biochip Center, Helsinki University Hospital, Biomedicum Helsinki, FIN-00014 Helsinki, Finland [O. M.]

ABSTRACT

Gene expression patterns revealed by cDNA microarrays have facilitated classification of cancers into biologically distinct categories, some of which may explain the clinical behavior of the tumors (1–6). Despite this progress in diagnostic classification, the molecular mechanisms underlying gene expression patterns in cancer have remained elusive, and the utility of gene expression profiling in the identification of specific therapeutic targets remains limited.

INTRODUCTION

Gene expression patterns revealed by cDNA microarrays have facilitated classification of cancers into biologically distinct categories, some of which may explain the clinical behavior of the tumors (1–6). Despite this progress in diagnostic classification, the molecular mechanisms underlying gene expression patterns in cancer have remained elusive, and the utility of gene expression profiling in the identification of specific therapeutic targets remains limited.

Accumulation of genetic defects is thought to underlie the clonal evolution of cancer. Identification of the genes that mediate the effects of genetic changes may be important by highlighting transcripts that are actively involved in tumor progression. Such transcripts and their encoded proteins would be ideal targets for anticancer therapies, as demonstrated by the clinical success of new therapies against amplified oncogenes, such as ERBB2 and EGFR (7, 8), in breast cancer and many other solid tumors. Besides amplifications of known oncogenes, over 1100 publications have described the use of comparative genomic hybridization (CGH) to analyze the pattern of copy number alterations in cancer, but very few of the genes affected are known. Here, we performed high-resolution CGH analysis on cDNA microarrays in breast cancer and directly compared copy number and mRNA expression levels of 13,824 genes to quantitate the impact of genomic changes on gene expression. We identified and mapped the boundaries of 24 independent amplicons, ranging in size from 0.2 to 12 Mb. Throughout the genome, both high- and low-level copy number changes had a substantial impact on gene expression, with 44% of the highly amplified genes showing overexpression and 10.5% of the highly overexpressed genes being amplified. Statistical analysis with random permutation tests identified 270 genes whose expression levels across 14 samples were systematically attributable to gene amplification. These included most previously described amplified genes in breast cancer and many novel targets for genomic alterations, including the HOXB7 gene, the presence of which in a novel amplicon at 17q21.3 was validated in 10.2% of primary breast cancers and associated with poor patient prognosis. In conclusion, CGH on cDNA microarrays revealed hundreds of novel genes whose overexpression is attributable to gene amplification. These genes may provide insights to the clonal evolution and progression of breast cancer and highlight promising therapeutic targets.

20 recurrent regions of DNA amplification have been mapped in breast cancer by CGH5 (9, 10). However, these amplicons are often large and poorly defined, and their impact on gene expression remains unknown.

We hypothesized that genome-wide identification of those gene expression changes that are attributable to underlying gene copy number alterations would highlight transcripts that are actively involved in the causation or maintenance of the malignant phenotype. To identify such transcripts, we applied a combination of cDNA and CGH microarrays to: (a) determine the global impact that gene copy number variation plays in breast cancer development and progression; and (b) identify and characterize those genes whose mRNA expres-

---

1 Received 5/29/02; accepted 8/28/02.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

2 Supported in part by the Academy of Finland, Emil Aaltonen Foundation, the Finnish Cancer Society, the Pirkanmaa Cancer Society, the Pirkanmaa Cultural Foundation, the Finnish Breast Cancer Group, the Foundation for the Development of Laboratory Medicine at the University Research Fund of the Tampere University Hospital, the Foundation for Commercial and Technical Sciences, and the Swedish Research Council.

3 Contributed equally to this work.

4 To whom requests for reprints should be addressed, at Laboratory of Cancer Genetics, Institute of Medical Technology, Lenkelinkatu 6, FIN-33200 Tampere, Finland. Phone: 358-3247-4125; Fax: 358-3247-4168; E-mail: anne.kallioniemi@uta.fi.

5 The abbreviations used are: CGH, comparative genomic hybridization; FISH, fluorescence in situ hybridization; RT-PCR, reverse transcription-PCR.
MATERIALS AND METHODS

Breast Cancer Cell Lines. Fourteen breast cancer cell lines (BT-20, BT-474, HCC1143, Hs578t, MCF7, MDA-361, MDA-436, MDA-453, MDA-468, SKBR-3, T-47D, UACC812, ZR-75-1, and ZR-75-30) were obtained from the American Type Culture Collection (Manassas, VA). Cells were grown under recommended culture conditions. Genomic DNA and mRNA were isolated using standard protocols.

Copy Number and Expression Analyses by cDNA Microarrays. The preparation and printing of the 13,824 cDNA clones on glass slides were performed as described (11–13). Of these clones, 244 represented uncharacterized expressed sequence tags, and the remainder corresponded to known genes. CGH experiments on cDNA microarrays were done as described (14, 15). Briefly, 20 μg of genomic DNA from breast cancer cell lines and normal human WBCs were digested for 14–18 h with AluI and RsaI (Life Technologies, Inc., Rockville, MD) and purified by phenol/chloroform extraction. Six μg of digested cell line DNAs were labeled with Cy3-dUTP (Amersham Pharmacia) and normal DNA with Cy5-dUTP (Amersham Pharmacia) using the Bioprime Labeling kit (Life Technologies, Inc.). Hybridization (14, 15) and posthybridization washes (13) were done as described. For the expression analyses, a standard reference (Universal Human Reference RNA; Stratagene, La Jolla, CA) was used in all experiments. Forty μg of reference RNA were labeled with Cy3-dUTP and 3.5 μg of test mRNA with Cy5-dUTP, and the labeled cDNAs were hybridized on microarrays as described (13, 15). For both microarray analyses, a laser confocal scanner (Agilent Technologies, Palo Alto, CA) was used to measure the fluorescence intensities at the target locations using the DEARRAY software (16). After background subtraction, average intensities at each clone in the test hybridization were divided by the average intensity of the corresponding clone in the control hybridization. For the copy number analysis, the ratios were normalized on the basis of the distribution of ratios of all targets on the array and for the expression analysis on the basis of 88 housekeeping genes, which were spotted four times onto the array. Low quality measurements (i.e., copy number data with mean reference intensity <100 fluorescent units, and expression data with both test and reference intensity <100 fluorescent units and/or with spot size <50 units) were excluded from the analysis and were treated as missing values. The distributions of fluorescence ratios were used to define cutpoints for increased/decreased copy number. Genes with CGH ratio >1.43 (representing the upper 5% of the CGH ratios across all experiments) were considered to be amplified, and genes with ratio <0.73 (representing the lower 5%) were considered to be deleted.

Statistical Analysis of CGH and cDNA Microarray Data. To evaluate the influence of copy number alterations on gene expression, we applied the following statistical approach. CGH and cDNA calibrated intensity ratios were log-transformed and normalized using median centering of the values in each cell line. Furthermore, cDNA ratios for each gene across all 14 cell lines were median centered. For each gene, the CGH data were represented by a vector that was labeled 1 for amplification (ratio, >1.43) and 0 for no amplification. Amplification was correlated with gene expression using the signal-to-noise statistics (1). We calculated a weight, \( w_e \), for each gene as follows:

\[
\frac{m_{e1} - m_{e0}}{\sigma_{e1} + \sigma_{e0}}
\]

where \( m_{e1} \), \( \sigma_{e1} \), and \( m_{e0}, \sigma_{e0} \) denote the means and SDs for the expression levels for amplified and nonamplified cell lines, respectively. To assess the statistical significance of each weight, we performed 10,000 random permutations of the label vector. The probability that a gene had a larger or equal weight by random permutation than the original weight was denoted by \( \alpha \). A low \( \alpha \) (<0.05) indicates a strong association between gene expression and amplification.

Genomic Localization of cDNA Clones and Amplicon Mapping. Each cDNA clone on the microarray was assigned to a Unigene cluster using the Unigene Build 141.6 A database of genomic sequence alignment information for mRNA sequences was created from the August 2001 freeze of the University of California Santa Cruz’s GoldenPath database.7 The chromosome and bp positions for each cDNA clone were then retrieved by relating these data sets. Amplicons were defined as a CGH copy number ratio >2.0 in at least two adjacent clones in two or more cell lines or a CGH ratio >2.0 in at least three adjacent clones in a single cell line. The amplicon start and end positions were

---

6 Internet address: http://research.nhgri.nih.gov/microarray/downloadable_cDNA.html.
7 Internet address: www.genome.ucsc.edu.
extended to include neighboring nonamplified clones (ratio, <1.5). The amplion size determination was partially dependent on local clone density.

**FISH.** Dual-color interphase FISH to breast cancer cell lines was done as described (17). Bacterial artificial chromosome clone RP11-361K8 was labeled with SpectrumOrange (Vysis, Downers Grove, IL), and SpectrumGreen-labeled chromosome 7 and 17 centromere probes (Vysis) were used as a reference. A tissue microarray containing 612 formalin-fixed, paraffin-embedded primary breast cancers (17) was applied in FISH analyses as described (18). The use of these specimens was approved by the Ethics Committee of the University of Basel and by the NIH. Specimens containing a 2-fold or higher increase in the number of test probe signals, as compared with corresponding centromere signals, in at least 10% of the tumor cells were considered to be amplified. Survival analysis was performed using the Kaplan-Meier method and the log-rank test.

**RT-PCR.** The HOXB7 expression level was determined relative to GAPDH. Reverse transcription and PCR amplification were performed using Access RT-PCR System (Promega Corp., Madison, WI) with 10 ng of mRNA as a template. HOXB7 primers were 5′-GAGCAGAGGGACTCGGACTT-3′ and 5′-GGTTCAGGTAGCGATTGTAG-3′.

### RESULTS

**Global Effect of Copy Number on Gene Expression.** 13,824 arrayed cDNA clones were applied for analysis of gene expression and gene copy number (CGH microarrays) in 14 breast cancer cell lines. The results illustrate a considerable influence of copy number on gene expression patterns. Up to 44% of the highly amplified transcripts (CGH ratio, >2.5) were overexpressed (i.e., belonged to the global upper 7% of expression ratios), compared with only 6% for genes with normal copy number levels (Fig. 1A). Conversely, 10.5% of the transcripts with high-level expression (cDNA ratio, >10) showed increased copy number (Fig. 1B). Low-level copy number increases and decreases were also associated with similar, although less dramatic, outcomes on gene expression (Fig. 1).

**Identification of Distinct Breast Cancer Amplions.** Base-pair locations obtained for 11,994 CDNs (86.8%) were used to plot copy number changes as a function of genomic position (Fig. 2, Supplement Fig. A). The average spacing of clones throughout the genome was 267 kb. This high-resolution mapping identified 24 independent breast cancer amplicons, spanning from 0.2 to 12 Mb of DNA (Table 1). Several amplification sites detected previously by chromosomal CGH were validated, with 1q21, 17q12–q21.2, 17q22–q23, 20q13.1, and 20q13.2 regions being most commonly amplified. Furthermore, the boundaries of these amplicons were precisely delineated. In addition, novel amplicons were identified at 9p13 (38.65–39.25 Mb), and 17q21.3 (52.47–55.80 Mb).

**Direct Identification of Putative Amplification Target Genes.** The cDNA/CGH microarray technique enables the direct correlation of copy number and expression data on a gene-by-gene basis throughout the genome. We directly annotated high-resolution CGH plots with gene expression data using color coding. Fig. 2C shows that most of the amplified genes in the MCF-7 breast cancer cell line at 1p13, 17q22–q23, and 20q13 were highly overexpressed. A view of chromosome 7 in the MDA-468 cell line implicates EGFR as the most highly overexpressed and amplified gene at 7p11–p12 (Fig. 3A). In BT-474, the two known amplicons at 17q12 and 17q22–q23 contained numerous highly overexpressed genes (Fig. 3B). In addition, several genes, including the homeobox genes HOXB2 and HOXB7, were highly amplified in a previously undescribed independent amplicon at 17q21.3. HOXB7 was systematically amplified (as validated by FISH, Fig. 3B, inset) as well as overexpressed (as verified by RT-PCR, data not shown) in BT-474, UACC812, and ZR-75-30 cells. Furthermore, this novel

### Table 1 Summary of independent amplicons in 14 breast cancer cell lines by CGH microarray

<table>
<thead>
<tr>
<th>Location</th>
<th>Start (Mb)</th>
<th>End (Mb)</th>
<th>Size (Mb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1p13</td>
<td>132.79</td>
<td>132.94</td>
<td>0.2</td>
</tr>
<tr>
<td>1q21</td>
<td>173.92</td>
<td>177.25</td>
<td>3.3</td>
</tr>
<tr>
<td>1q22</td>
<td>179.28</td>
<td>179.57</td>
<td>0.3</td>
</tr>
<tr>
<td>3p14</td>
<td>71.94</td>
<td>74.66</td>
<td>2.7</td>
</tr>
<tr>
<td>7p11.1–7p11.2</td>
<td>55.62</td>
<td>60.95</td>
<td>5.3</td>
</tr>
<tr>
<td>7q31</td>
<td>125.73</td>
<td>130.96</td>
<td>5.2</td>
</tr>
<tr>
<td>7q32</td>
<td>140.01</td>
<td>140.68</td>
<td>0.7</td>
</tr>
<tr>
<td>8q21.11–8q21.13</td>
<td>86.45</td>
<td>92.46</td>
<td>6.0</td>
</tr>
<tr>
<td>8q21.3</td>
<td>98.45</td>
<td>103.05</td>
<td>4.6</td>
</tr>
<tr>
<td>8q23.3–8q24.14</td>
<td>129.88</td>
<td>142.15</td>
<td>12.3</td>
</tr>
<tr>
<td>8q24.22</td>
<td>151.21</td>
<td>152.16</td>
<td>1.0</td>
</tr>
<tr>
<td>9p13</td>
<td>38.65</td>
<td>39.25</td>
<td>0.6</td>
</tr>
<tr>
<td>13q22–q31</td>
<td>77.15</td>
<td>81.38</td>
<td>4.2</td>
</tr>
<tr>
<td>16q22</td>
<td>86.70</td>
<td>87.62</td>
<td>0.9</td>
</tr>
<tr>
<td>17q11</td>
<td>29.30</td>
<td>30.85</td>
<td>1.6</td>
</tr>
<tr>
<td>17q12–q21.2</td>
<td>39.79</td>
<td>42.80</td>
<td>3.0</td>
</tr>
<tr>
<td>17q21.32–q21.33</td>
<td>52.47</td>
<td>55.80</td>
<td>3.3</td>
</tr>
<tr>
<td>17q22–q23</td>
<td>63.81</td>
<td>69.70</td>
<td>5.9</td>
</tr>
<tr>
<td>17q23.3–q24.3</td>
<td>69.93</td>
<td>74.99</td>
<td>5.1</td>
</tr>
<tr>
<td>19q13</td>
<td>40.63</td>
<td>41.40</td>
<td>0.8</td>
</tr>
<tr>
<td>20q11.22</td>
<td>34.59</td>
<td>35.85</td>
<td>1.3</td>
</tr>
<tr>
<td>20q13.12</td>
<td>44.00</td>
<td>45.62</td>
<td>1.6</td>
</tr>
<tr>
<td>20q13.12–q13.13</td>
<td>46.45</td>
<td>49.43</td>
<td>3.0</td>
</tr>
<tr>
<td>20q13.2–q13.32</td>
<td>51.32</td>
<td>59.12</td>
<td>7.8</td>
</tr>
</tbody>
</table>
amplification was validated to be present in 10.2% of 363 primary breast cancers by FISH to a tissue microarray and was associated with poor prognosis of the patients (P < 0.001).

Statistical Identification and Characterization of 270 Highly Expressed Genes in Amplicons. Statistical comparison of expression levels of all genes as a function of gene amplification identified 270 genes whose expression was significantly influenced by copy number across all 14 cell lines (Fig. 4, Supplemental Fig. B). According to the gene ontology data, 891 of the 270 genes represented hypothetical proteins or genes with no functional annotation, whereas 179 had associated functional information available. Of these, 151 (84%) are implicated in apoptosis, cell proliferation, signal transduction, and transcription, whereas 28 (16%) had functional annotations that could not be directly linked with cancer.

DISCUSSION

The importance of recurrent gene and chromosome copy number changes in the development and progression of solid tumors has been characterized in >1000 publications applying CGH (9, 10), as well as in a large number of other molecular cytogenetic, cytogenetic, and molecular genetic studies. The effects of these somatic genetic changes on gene expression levels have remained largely unknown, although a few studies have explored gene expression changes occurring in specific amplicons (15, 19–21). Here, we applied genome-wide cDNA microarrays to identify transcripts whose expression changes were attributable to underlying gene copy number alterations in breast cancer.

The overall impact of copy number on gene expression patterns was substantial with the most dramatic effects seen in the case of high-

---

8 Internet address: http://www.geneontology.org/.
level copy number increase. Low-level copy number gains and losses also had a significant influence on expression levels of genes in the regions affected, but these effects were more subtle on a gene-by-gene basis than those of high-level amplifications. However, the impact of low-level gains on the dysregulation of gene expression patterns in cancer may be equally important if not more important than that of high-level amplifications. Aneuploidy and low-level gains and losses of chromosomal arms represent the most common types of genetic alterations in breast and other cancers and, therefore, have an influence on many genes. Our results in breast cancer extend the recent studies on the impact of aneuploidy on global gene expression patterns in yeast cells, acute myeloid leukemia, and a prostate cancer model system (22–24).

The CGH microarray analysis identified 24 independent breast cancer amplics. We defined the precise boundaries for many amplics detected previously by chromosomal CGH (9, 10, 25, 26) and also discovered novel amplicons that had not been detected previously, presumably because of their small size (only 1–2 Mb) or close proximity to other larger amplics. One of these novel amplicons involved the homeobox gene region at 17q21.3 and led to the overexpression of the HOXB7 and HOXB2 genes. The homoeodomain transcription factors are known to be key regulators of embryonic development and have been occasionally reported to undergo aberrant expression in cancer (27, 28). HOXB7 transfection induced cell proliferation in melanoma, breast, and ovarian cancer cells and increased tumorigenicity and angiogenesis in breast cancer (29–32). The present results imply that gene amplification may be a prominent mechanism for overexpressing HOXB7 in breast cancer and suggest that HOXB7 contributes to tumor progression and confers an aggressive disease phenotype in breast cancer. This view is supported by our finding of amplification of HOXB7 in 10% of 363 primary breast cancers, as well as an association of amplification with poor prognosis of the patients.

We carried out a systematic search to identify genes whose expression levels across all 14 cell lines were attributable to amplification status. Statistical analysis revealed 270 such genes (representing ~2% of all genes on the array), including not only previously described amplified genes, such as HER-2, MYC, EGF, ribosomal protein S6 kinase, and ALB, but also numerous novel genes such as NRAS-related gene (1p13), syndecan-2 (8q22), and bone morphogenetic protein (20q13.1), whose activation by amplification may similarly promote breast cancer progression. Most of the 270 genes have not been implicated previously in breast cancer development and suggest novel pathogenic mechanisms. Although we would not expect all of them to be causally involved, it is intriguing that 84% of the genes with associated functional information were implicated in apoptosis, cell proliferation, signal transduction, transcription, or other cellular processes that could directly imply a possible role in cancer progression. Therefore, a detailed characterization of these genes may provide biological insights to breast cancer progression and might lead to the development of novel therapeutic strategies.

In summary, we demonstrate application of cDNA microarrays to the analysis of both copy number and expression levels of over 12,000 transcripts throughout the breast cancer genome, roughly on January 13, 2018. © 2002 American Association for Cancer Research.

**REFERENCES**


Impact of DNA Amplification on Gene Expression Patterns in Breast Cancer

Elizabeth Hyman, Päivikki Kauraniemi, Sampsa Hautaniemi, et al.


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/62/21/6240

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2005/01/25/62.21.6240.DC1

Cited articles
This article cites 31 articles, 12 of which you can access for free at:
http://cancerres.aacrjournals.org/content/62/21/6240.full#ref-list-1

Citing articles
This article has been cited by 53 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/62/21/6240.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link http://cancerres.aacrjournals.org/content/62/21/6240. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.