Distinct Amplification of an Untranslated Regulatory Sequence in the *egfr* Gene Contributes to Early Steps in Breast Cancer Development

Nicola Tidow, Almuth Boecker, Hartmut Schmidt, Konstantin Ageopoulos, Werner Boecker, Horst Buerger, and Burkhard Brandt

Institutes of Clinical Chemistry and Laboratory Medicine [W. T. A. B., H. S., K. A., B. B.] and Pathology [W. B., H. B.], University of Münster, 48149 Münster, Germany

Abstract

Overexpression of the epidermal growth factor receptor (*egfr*) gene is a common feature in breast cancer. We demonstrated recently that the expression of EGFR in breast cancer strongly correlates with the length of a CA simple sequence repeat within the first 2000 bases in intron 1 of the *egfr* gene [CA simple sequence repeat (CA-SSR) I; H. Buerger et al., Cancer Res., 60: 854–857, 2000].

Using a standardized semiautomated method of microsatellite analysis for loss of heterozygosity detection, we identified an allelic imbalance (AI) at the *egfr* locus in 55 of 163 primary breast cancer cases. Fine mapping of the chromosomal region at 7p12–15 around the *egfr* gene using 10 CA-SSR markers showed that mutations of *egfr* in breast cancer are frequently restricted to the first intron of *egfr*. Thereby, the simple sequence repeat CA-SSR I in intron 1 was affected in 84% of the patients with AI. Reverse transcription-PCR analysis of 23 breast cancer tissues with AI excluded the presence of in-frame deletions between exon 2 and exon 7. For additional characterization of the underlying phenomenon leading to the detection of an AI in microsatellite analysis, a quantitative 5′-nuclease assay for the first CA-SSR I in intron 1 was established. In breast cancer cases with AI the presence of amplifications of this sequence was shown. Kaplan-Meier analysis revealed a statistically significant worse prognosis for patients with AI in the cancer tissue at the *egfr* locus compared with patients without AI. Interestingly, 75% of the patients bearing AI of CA-SSR I in the tumor also showed AI at normal, nontumorous breast tissue. Our data strongly support the assumption that distinct amplifications in intronic sequences of the *egfr* gene, which enhance the basic transcription activity of the gene, represent one of the first steps in breast carcinogenesis. Furthermore, they point to the presence of prognosis-associated markers for breast cancer already in morphological normal breast tissue.

Introduction

The *egfr* is a prominent member of the *erbB* gene family, which encodes a receptor tyrosine kinase with a pivotal role in the regulation of cell growth and differentiation in response to specific ligands such as epidermal growth factor and transforming growth factor α. Overexpression of EGFR, which has been shown in a variety of tumors, resulted in cellular transformation as shown in NIH-3T3 cells (1, 2).

In breast cancer patients, EGFR overexpression was associated with a poor clinical outcome and no benefit from adjuvant endocrine therapy (3, 4). Amplifications of *egfr* in breast cancer, as described for c-erbB2 (3), account for <5% of all breast cancer cases. Therefore, it was broadly accepted that overexpression of EGFR in breast cancer is commonly regulated on the transcriptional level. Furthermore, in a few breast cancer cases an in-frame deletion comprising exon 2 and exon 7 has been described, encoding a constitutively activated EGFR (egfr-vIII), which was found frequently in glioblastomas (5–7).

Premature termination of transcription between exons 1 and 2 of the *egfr* gene occurring proximal to an enhancer region in intron 1 has been described (8). The transcription activity of the *egfr* gene depends strongly on the cooperation of this enhancer in intron 1 with an enhancer sequence upstream of the transcriptional start side (9). The enhancer in intron 1 is located in close proximity to a polymorphic CA-SSR containing 14–21 CA dinucleotides (10). For this CA-SSR we have shown that *egfr* transcription is modulated in *vitro* and in *vivo* (11). Decreasing numbers of CA dinucleotides in CA-SSR I correlated with an increased *egfr* transcription and EGFR protein expression (12).

Our data indicate that CA-SSR I is the frequent target for mutations within the *egfr* gene in breast cancer. For the first time we show that the mutations are amplifications. Furthermore, these amplifications are likely to represent early steps in breast carcinogenesis, moreover thereby inducing a breast cancer subtype associated with poor clinical outcome. The discovery of one of the putative, first genetic hits in a subtype of breast cancer might serve as a starting point in the definition of new genetic pathways in breast cancer.

Materials and Methods

Patients. In total, 163 Caucasian patients from the Department of Gynecology of the University Hospital of Muenster primarily diagnosed with breast cancer who had undergone mastectomy with full dissection of axillary lymph nodes were included. The average age at time of primary diagnosis was 57.3 years (range, 34–82 years). For 82 patients the time of follow-up exceeded 5 years. The patient stages were as follows: I, 29.5%; IIa, 25.6%; IIb, 15.4%; IIIa, 2.6%; IIIb, 6.4%, and IV, 20.5%. Lymph node involvement was observed in 52% of the cases. Histologically proven tumor tissues were frozen immediately after surgery.

Microsatellite PCR. Genomic DNA was isolated from peripheral blood and fresh-frozen tumor tissue of patients with breast cancer. DNA isolation from peripheral blood was performed using the QIAamp Blood kit (Qia-gen), and tumor tissue DNA was extracted using the QIAamp DNA Mini kit (Qiagen) according to the manufacturer’s instructions. As a control for PCR fragment length DNA from the tumor cell line MDA-MB-468 was analyzed.

PCR amplification was performed using AmpliTag DNA Polymerase (Applera) in a 25-μl reaction volume containing 200 nM of each primer, 1× GeneAmp buffer II, 2 mM MgCl₂, 100 μM of each GeneAmp deoxynucleotide triphosphate (Applera), and 20 ng of sample DNA. The primer sequences specific for microsatellite markers near the chromosomal locus of *egfr* on chromosome 7p are given in Table 1. Downstream primers were labeled with a fluorescent dye (FAM). Separation was done with a four-color laser-induced
fluorescence capillary electrophoresis system (ABI PRISM 310 DNA Analyser and ABI PRISM 3700 DNA Analyser). One to 2 μl of the amplified PCR products were diluted in 20 μl of water (high-performance liquid chromatography grade) containing 0.5 μl GENESCAN 500 TAMRA or 400 HD (Rox) fluorescent size standard (Applera). Denatured PCR fragments were separated on the ABI PRISM 310 and ABI PRISM 3700 DNA Analyser (Applera). Evaluation of the collected data was accomplished with the GeneScan Analysis Software (Applera). All of the analyses were performed at least in duplicates of independent PCRs.

For standardization of the analysis, the AI score (which is similar to the loss of heterozygosity score defined by Canzian et al.; Ref. 13) was calculated according to the equation published by Canzian et al. (13), which represents the ratio of the intensities of the two alleles in the tumor tissue corrected by that in nontumor. The AI score estimates the imbalance of the two allele intensities in tumor. The reliability of peak area and lengths for AI score determination was tested by 10 runs of independent PCR amplification. Denatured PCR fragments were separated using size standards and a high-performance liquid chromatography grade marker (GeneScan 500 TAMRA or 400 HD). The AI score was calculated by 1-dc for the loss of the longer allele and 1/(1−dc) for the loss of the shorter allele resulting in cutoff values of 0.79 (loss of longer allele) and 1.27 (loss of shorter allele). Because of the low SD of the method, AI could be detected sensitively if the peak area for one allele in relation to the other decreased in the tumor DNA to 79% of the normal.

**RT-PCR Analysis.** Quick-frozen tumor tissues from 23 breast cancer patients with AI at the egfr region were submitted to extraction of total RNA using the single-step isolation described by Chomczynski and Sacchi (15) using RNAzol (Diagnostic International). RT-PCR was performed with the Titan One Step RT-PCR kit purchased from Roche Molecular Diagnostics according to the manufacturer’s instructions. In brief, total RNA was reverse-transcribed using avian myeloblastosis virus reverse transcriptase. Amplification was performed using Taq DNA polymerase and Pwo DNA polymerase, and the following oligonucleotides: K1:5′-CA-SSR VI AG TTC CTG ACT GGG AAT TCG AT TTG GCC AAA TTA CAC ACC TTT G-3′, K2:5′-CA-SSR III AGG GAG AGC CG-3′, and V3B:5′-GTG GAG ATC GCC ACT GAT G-3′, and V3B:5′-GTG GAG ATC GCC ACT GAT G-3′. Cycling conditions for primer pairs V3A/V3B and V3B/K1 were as follows: reverse transcription at 50°C for 30 min, initial denaturation at 94°C for 2 min, 35 cycles of denaturation (94°C for 1 min), annealing (59°C for 1 min), and extension (68°C for 1 min), and a final extension at 68°C for 7 min. Amplified fragments were separated on an agarose gel and visualized by ethidium bromide staining. Gel-excised fragments were purified using the QIAEX II system (Qiagen) and submitted to direct DNA sequencing using Big Dye Terminator Cycle Sequencing reagents (Applera) and the ABI PRISM 3700 DNA Analyser. The primer pairs and amplification conditions were optimized using both normal lymphocyte DNA and plasmid DH5αE801P containing an exon2/exon7 deletion fragment of the egfr. The primer pair V3A/V3B results in a regular fragment of 1153 bp spanning exon 1 to exon 8, and a 352-bp fragment is amplified when exons 2-7 are deleted. Primer pair V3B/K1 generates a 307-bp fragment when the mRNA is intact and a 251-bp fragment when exons 2-7 are deleted.

**Quantitative Real-Time PCR (5′ Nuclease Assay).** Primers specific for sequences flanking the first CA repeat in the first intron of the egfr gene (CAFor: 5′-tgaagaacctgacgaccagca-3′) were designed using Primer Express software (Applera) and a universal probe consisting of 15 CA repeats, and a 5′ fluorescent label (CA-Fam) was designed. The primers represented specific sequences for egfr, which were checked by BLAST search. Primers and probes were also designed for two different single-copy genes, SOD2 (chromosome 6q25 GenBank accession no. 65965), forward primer: 5′-GGAGAAACGTGACCGCTGC-3′, reverse primer: 5′-CCCTTATTGGAACCAACGACCAC-3′, VIC-labeled probe: 5′-CAACCTTGGACCTTGGGAAAG-3′, reverse primer: 5′-CAGCTCCTACGTGGCAGAAG-3′, VIC-labeled probe: ATGGGCGTGTCACCTGGCAACC-3′. For all of the genes, the amplification efficiency was maximized (69–97%) to allow for the most efficient PCR amplification. PCR analysis was performed using TaqMan Universal Mix (Applera) and an ABI Prism 7900HT Sequence Detection System (Applied Biosystems). A PCR cycle consisted of denaturation at 95°C for 15 s and primer annealing and extension at 60°C for 1 min, and was repeated 40 times. PCR reactions were performed at least in triplicates, and serial dilutions of DNA ensured accuracy of gene dosage quantification.

**Amplification of the egfr gene** was measured in the breast cancer-derived cell line MDA-MB-468 in comparison to normal leukocytes. MDA-MB-468 DNA reportedly displays a 30–50-fold amplification of the egfr gene (16) and was used as a positive control. DNA concentrations were normalized to both SOD2 and HBB, two different single-copy genes (17).

**Induction and Detection of Fragile Sites in B Lymphocytes.** Heparinized whole peripheral blood (1 ml) was cultured in 10 ml Ham’s F-10 medium supplemented with 20% FCS, 100 μmol 1-glutamine, 144 μg phenytoin, and 10,000 units of streptomycin/penicillin at 37°C for 72 h. At 24, 24, and 6 h before harvest, aphidicolin (0.5 μmol final concentration), ethanol (0.5% final concentration), and caffeine (2.2 mm final concentration), respectively, were added to the lymphocyte culture for fragile site induction.

Dividing cells were blocked in metaphase by addition of colcemide 1.5 h before harvest. Standard cytogenetic techniques were used for harvesting and slide preparation. The probe for egfr detection was derived from homo sapiens PAC clone containing the whole egfr gene (GenBank accession no. AC006977). DNA was labeled with digoxigenin-11-dUTP by nick translation following standard protocols (18). The probe was denatured for 5 min at 70°C in 50% formamid-0.6× SSC. Hybridization to the metaphase spreads was carried out overnight at 37°C in a 50% formamid-1× SSC-10% dextran sulfate solution in the presence of Cot-1 DNA (Life Technologies, Inc.) and HPL-DNA (Sigma). Posthybridization washes were performed at 45°C in 50% formamide-2× SSC followed by blocking with 3% BSA in 4× SSC at 37°C. Probe detection was performed using mouse-anti-digoxigenin (Sigma) and Cy3-labeled goat-antimouse antibodies. The chromosomes were counterstained with 4′,6-diamidino-2-phenylindole, and images were captured with an Olympus BX61 microscope connected with a digital camera system DF50 (Hamburg, Germany). Giemsa-tropsin banding was performed according to the standard protocol of Seabright (19). Briefly, slides were incubated 10–40 s in trypsin (1 mg/ml in NaCl) at 37°C, quickly rinsed in water, stained 4 min in Giemsa, and then washed in water. Complete metaphase spreads (n > 100) were

**Table 1** Primer sequences of microsatellite markers and CA-SSR markers in the egfr region

<table>
<thead>
<tr>
<th>Primer</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>D7S3429</td>
<td>CAG TCC TGG AGT TGG TTA AGT</td>
<td>CTG GGA GTC AAG CAG TTT GG</td>
</tr>
<tr>
<td>D7S494</td>
<td>CTC GGG TGA CAG AGA TTT TTG</td>
<td>AAT CCT GTT ATT AAC TGG TTT T</td>
</tr>
<tr>
<td>CA-SSR I</td>
<td>GTT AGA ATT TGG GAC AAC C</td>
<td>TTC TTC TGC ACA CTT GGC AC</td>
</tr>
<tr>
<td>CA-SSR II</td>
<td>CTG AGT GCT ATC TTC TTT CC</td>
<td>GCA GAG GTG CAC AAA GGA GTA A</td>
</tr>
<tr>
<td>D7S2350</td>
<td>AAG TCT CAT TGG TCT CGT TTG</td>
<td>AGT CTC CTC GTC TCA CAC CT</td>
</tr>
<tr>
<td>CA-SSR III</td>
<td>AG GCC CAC AGA GGA GAT AAC AG</td>
<td>CAG GTG TAG ATO CCA AAG A</td>
</tr>
<tr>
<td>CA-SSR IV</td>
<td>GC AAT TCA TCC AAA CCC TGA CC</td>
<td>AGA GTG GAC TAG GAA ATG CTA GGA G</td>
</tr>
<tr>
<td>CA-SSR VI</td>
<td>AG TGC CTG ACT GGT TCT AGT</td>
<td>TGT GCC AAA TTA CAC ACC TTT G</td>
</tr>
<tr>
<td>D7S599</td>
<td>CCA ACT ATA TGG ACA GGA ACC</td>
<td>TCG CCA ACA TTT ATA CA C</td>
</tr>
<tr>
<td>D7S2467</td>
<td>TGCC ATA GTG CGT ATG TTG CC</td>
<td>AAC GGT CAT CTT GTC TGCG</td>
</tr>
</tbody>
</table>
assessed at the 350–400-band level by conventional light microscopy (Olympus BX61).

Microdissection of Tissue from Normal, Premalignant Lesions and Tumor Tissues. Tissues from 16 patients, all known to harbor AI at CA-SSR I in their invasive carcinoma, were laser microdissected (Palmar, Oberlenheim, Germany) from serial 10-μm sections of paraffin-embedded tissue after hematoxylin staining. The sections were evaluated by two pathologists (W. B., H. B.). DNA was isolated from both nontumorous lymphocytes and invasive carcinoma from paraffin-embedded tissue. In these 16 cases, nontumorous, morphologically normal breast lobules (n = 69) could be isolated. Breast lobules were pooled to maximize the DNA yield. Only breast lobules with a distinct myoepithelial layer and a single layer of luminal epithelial cells with small, uniform nuclei with evenly distribute chromatin were selected for AI analysis. Twenty-three samples consisted of benign, metaplastic, or hyperplastic proliferative breast lesions such as adenosis, ductal hyperplasia, or apocrine adenosis. In 3 cases ductal hyperplasias with nuclear atypias were present. Also, 3 samples of DCIS and 9 samples of associated lymph node metastasis could be analyzed. Benign and malignant breast lesions were defined according to Elston and Ellis (20).

For isolation of DNA from deparaffinized, microdissected tissue the material was suspended in 10–200 μl TE buffer [10 mM Tris/ HCl and 1 mM EDTA (pH 7.5)] and incubated with 1–20 μl proteinase K (600 mAu/ml) at 56°C for 10 h. The tube was boiled (10 min), chilled (10 min at 4°C), and enclosed centrifuged for 1 min at 14,000 × g. The supernatant was transferred into a fresh tube before DNA was pelleted in 70% ethanol and centrifuged for 20 min. After two washing steps (70% ethanol; 20 min) 14,000 × g) the pellet was air dried and rehydrated by LTE or distilled water and stored frozen at −20°C. Resulting DNA amounts were determined before CA-SSR I microsatellite analysis by quantitative real-time PCR. Both previously mentioned single-copy genes, SOD2 and HBB (primer and probe sequences as shown above), were amplified and quantified via the 5’ nuclease assay. PCR analysis was performed using TaqMan Universal Mix (Applera), 400 nmol of forward and reverse primer, 200 nmol probe, 2.5 μl DNA solution, and an ABI Prism 7900HT Sequence Detection System (Applera). PCR conditions were the same as described above. As a control, DNA isolated from formalin-fixed, paraffin-embedded lymph node tissue was photometrically quantified. A standard curve was obtained by serial dilution of control DNA ranging from 120 pg to 10 ng per (each) PCR assay and single-copy gene. The test DNA was amplified in the same PCR run, and the concentration was determined using the standard curve. Each DNA was tested once for each single-copy gene, because the total DNA amount after microdissection is limited. The DNA was used for CA-SSR I microsatellite analysis if the obtained DNA concentrations for SOD2 and HBB were equal.

To ensure the specificity of microsatellite analysis in microdissected tissues, we determined the minimum DNA template concentration by performing serial dilutions of DNA starting at 30 ng. Reproducibility of results was approved by duplicate experiments using a minimum of 1 ng of DNA. Therefore, microsatellite analysis of microdissected tissues was performed with 1.2 ng of DNA.

The PCR protocol for the determination of AI at microdissected tissues was adopted for the use of low cell numbers as follows. PCR assays of egfr using primers for CA-SSR I (Table 1) were performed in 10-μl reactions containing 1 × PCR Buffer II (Applera), 2 μM MgCl₂, 50 μM of each GeneAmp deoxynucleotide triphosphate (Applera), 1 μM of forward and reverse primer, 1.2 ng of template DNA, and 0.5 units of Ampli Taq Gold (Applera). The PCR reactions were overlaid with mineral oil and carried out in a 96-well thermocycler (GeneAmp PCR System 9700; Applera). A denaturation step at 95°C for 10 min was followed by 40 cycles of denaturation at 95°C for 30 s, annealing at 55°C for 30 s, primer extension at 72°C for 30 s, and one final extension at 72°C for 7 min. Separation of PCR fragments was performed with a four-color laser-induced fluorescence capillary electrophoresis system as described above.

Statistical Methods. Actual survival probabilities were calculated by Kaplan-Meier method, and the differences between the groups were estimated by the log-rank or Wilcoxon test. It should be noted that the number of patients in the tables and figures does not add up to 163 because of missing data, e.g., not performed experiments, and patients who were not at least 5 years in or lost in follow-up. Multivariate analysis was performed using Cox proportional hazards regression model.

Results

AI at CA-SSR I as Determined by PCR-based Microsatellite Analysis Occurred in High Frequencies in Invasive Breast Cancer. The egfr locus on chromosome 7p was tested for AI using a standardized semiautomated method for microsatellite analysis. DNA was isolated from normal lymphocytes and tumor tissue of 163 breast cancer patients, and assessed for AI using primers specific for CA-SSR I in the first intron of the egfr gene. We identified AI in 55 of 163 (34%) primary cases. CA-SSR I centers the known regulatory sequences of the egfr gene, which is located within the 5’-flanking region and intron 1. In the cancerous tissue AI occurred preferentially with an increase of the peak representing the shorter allele of the CA repeat.

To estimate the size of the mutated sequence we used 17 microsatellite markers on chromosome 7p to test 64 breast cancer patients (Fig. 1). The more distant microsatellite loci, D7S519, D7S665, and D7S510, showed no AI, and D7S674, D7S670, D7S478, and CA-SSR V were not informative (homozygous). Ten markers around and in the egfr gene revealed AI on 7p in 23 cases. No AI could be detected in 41 cases. Interestingly, 84% (16 of 19) of informative cases with AI showed AI at CA-SSR I located between enhancers 1 and 2 in the first intron of the egfr (Table 2). In addition to the AI at CA-SSR I, AI could be shown at different microsatellite markers on chromosome 7p. The affected sequence reaches up to 8 megabases with the shortest affected region being up to 2 megabases around CA-SSR I (e.g., see patients 78/414 and 7/425). However, all of the AIs occurred in the egfr gene, and/or the close vicinity upstream and/or downstream of the gene and at no other regions of chromosome 7p.

AI at CA-SSR I Was Not Associated with the vIII Deletion Mutant of EGFR. To investigate whether the AI at CA-SSR I is caused by the well-defined exon2/exon7 deletion resulting in EGFR isoform vIII (5), we performed RT-PCR analysis using specific primer pairs as described in “Material and Methods.” RT-PCR analysis of tumor tissues from 23 patients with egfr AI displayed only regular transcripts (Fig. 2). These results were confirmed by Western blotting where only full-length EGFR proteins could be detected (data not shown). Thus, exon2/exon7 deletions could not be detected in these tissues and, therefore, are not related to the occurrence of AI in the egfr gene.

![Fig. 1. Mapping of AIs on chromosome 7p. In a PCR-based microsatellite analysis we used 17 different markers to map the region around the egfr gene on chromosome 7p. Ten microsatellite markers including six CA-SSR markers in the egfr gene showed AI. No AI could be detected using D7S510, D7S665, and D7S519, which are located more distal on chromosome 7p and markers D7S670, D7S674, D7S478, and CA-SSR V were not informative. As an example results of CA-SSR I microsatellite analysis are given in the right panel.](https://example.com/fig1.png)
AI at CA-SSR I Represents an Amplification as Determined by Quantitative Real-Time PCR. To additionally study whether the AI represented a deletion or an amplification of one allele, we developed a quantitative S' nuclease assay. Egfr-specific primers that flank the CA-SSR I in the first intron and a FAM-labeled (CA)15 probe were used in this assay. PCR amplification was performed in serial dilutions of leukocyte and tumor DNA normalized to two different housekeeping genes as described in "Materials and Methods." The sensitivity of the method allowed the detection of an amplification if the AI score exceeds 0.56 or 1.3 (Fig. 3). As a control, DNA isolated from the MDA-MB-468 cell line was tested and displayed 30–60-fold amplification of egfr CA-SSR I. Up to 16-fold amplification of the egfr CA-SSR I region could be detected in patients with an AI score of 0.23. In cases with an AI score between 0.56 and 1.27 or 1.29, no significant differences in the gene dosage between tumor DNA and normal DNA could be assessed, but no case with a decreased gene dosage was found for the egfr CA SSR I by this method.

The Induction of a Common Fragile Site in the egfr Region Gives First Hinds for a Higher Probability for Amplifications at That Locus. Because of the fact that the sequence downstream from the CA-SSR I contains sequences that suggest a higher probability for DNA rearrangements (MER 22 and Alu I), we treated lymphocytes for the induction of common fragile sites by the method of Yunis and Soreng (21). Single-strand breaks at common fragile sites are well-known early events to trigger amplification by breakage-fusion-bridge cycles (22). As shown in Fig. 4, a single-strand break could be observed by Giemsa stain and FISH analysis, which occurred near the egfr gene region as shown by the FISH analysis.

AI at CA-SSR I Occurred in Normal Mammary Gland Tissue and Benign Lesions, as well as in Invasive and Metastatic Cancer Tissues. In total, 127 microdissected samples of breast tissue from 16 patients with AI at the egfr locus were studied. Between 100 and 500 cells were microdissected from formalin-fixed, paraffin-embedded normal, noncancerous, and cancer tissues. In addition, invasive breast cancer tissue and metastatic tissue was evaluated as a control. AI analysis was performed using primers specific for the CA-SSR I of the egfr gene in microsatellite PCR. All of the 30 samples from metastatic tissue, invasive breast cancer tissue, and DCIS displayed AI at CA-SSR I (Table 3). In 5 of 23 and 4 of 5

### Table 2: Frequency of AI in breast cancer samples

<table>
<thead>
<tr>
<th>Patient No.</th>
<th>D7S2429</th>
<th>D7S494</th>
<th>CA-SSR I</th>
<th>CA-SSR II</th>
<th>D7S2550</th>
<th>CA-SSR III</th>
<th>CA-SSR IV</th>
<th>CA-SSR VI</th>
<th>D7S2467</th>
<th>D7S5499</th>
</tr>
</thead>
<tbody>
<tr>
<td>74/425</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>A98/269</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>A46/423</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>A8/434</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>7B/444</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>59/391</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>109/460</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>65/423</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>A8/432</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>A98/404</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>A17/399</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>A12/436</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>A21/445</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>A10/458</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>A6/364</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>A10/458</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>A13/438</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>A53/465</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>A97/375</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>36/458</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>29/424</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>27/425</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>A17/467</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

Frequency: 62.5% 37.5% 84.2% 80% 38.1% 41.7% 52.4% 63.6% 33.3% 0

---

Fig. 2. RT-PCR analysis of egfr transcripts. RT-PCR analysis on egfr mRNA was performed using different primer sets, which were designed to identify a possible exon 2/exon 7 deletion (egfr-vIII) as described in “Material and Methods.” Exons 1–8 of egfr are shown, and the primer pairs are indicated (top panel). Amplified RT-PCR fragments were separated on agarose gels, and lanes 1–6 represent samples of patients with AI at CA-SSR I (bottom middle and right panel). As a positive control plasmid DHAE801P with exon 2/exon 7 deletion was used, which generates a 352-bp fragment using primer pair V3A/V3B (bottom left panel). (Size standards VII from Roche: 124, 147, 190, 242, 320, 404, 489 + 501, 692, 900, and 1114 bp, and 100 bp ladder from MBI: 100, 200, 300, 400, 500, 600, 700, 800, 900, 1031, 1200, 1500, 2000, and 3000 bp).
samples, classified as benign proliferative breast lesions or intraductal proliferations with nuclear atypia, respectively, AI was detected. Furthermore, 30% of the microdissected samples of normal, noncancerous breast tissue showed an AI at the CA-SSR I of egfr. Twelve of 16 patients (75%) revealed AI at normal breast lobules, and at least two lobules were affected.

**AI at CA-SSR I Indicated a Subtype of Breast Cancer with Poor Clinical Outcome.** Patients with AI at CA-SSR I showed significantly higher risk of recurrence than the group with tumors diploid for egfr alleles (Fig. 5; \( P < 0.001 \)). The odds ratio for the 5-year probability of recurrence of the 82 patients with a follow-up >5 years reached 4.9 (95% confidence interval, 1.4–17.6). In the subgroup of 35 young early stage patients (<55 years) the odds ratio increased to 7.9 (95% confidence interval, 1.4–45.8).

**Discussion**

In a multistep tumor progression model of human breast cancer the first events at the molecular level are still unknown. Genomic instability leading to a higher mutation rate accounts for only a small proportion of breast cancers. Natural selection acting to increase the frequency of advantageous alleles in breast lobular cells can be the driving force behind tumor growth. Here, we give evidence that the segregation of alleles associated with a high transcriptional activity of the egfr gene may represent early events that promote progression from normal breast lobular cells to invasive and metastatic breast cancer cells.

Amplifications of the entire egfr gene are a frequent event in high malignant tumors of the central nervous system, and EGFR overexpression has been shown on the cytogenetic level in renal carcinoma (23–25). However, gene dosage studies show very little evidence for egfr gene rearrangements or amplifications (26). Against this background, the overexpression of EGFR was regarded mainly as a result of disturbed transcriptional mechanisms. We have shown previously that low-level gene dosage deviations at the boundary of exon 1 and intron 1 of the egfr gene occur frequently and identify patients with poor clinical outcome (14, 17). The deviations were not detectable on a cytogenetic level, or at most they were represented by low-level chromosomal gains in comparative genomic hybridization (CGH) (12).

In a recent study, we gave evidence for the first time that AIs...
affecting regulating sequences of egfr transcription in the intron 1 of egfr were correlated with EGFR expression in the tumor (12).

The data presented in this work clearly show that these genetic alterations involve the first 1200 bases of intron 1 of egfr. Although the frequency of AI on 7p differed substantially, a distinct peak in the CA-SSR I region as shown by microsatellite analysis using 16 additional markers was observed.

To exclude that this AI represents only a surrogate marker for the well-defined exon 2/exon 7 deletion mutation resulting in the constitutive active EGFR isoform vIII (6), we performed RT-PCR analysis. Using a nested approach by applying primers for the cDNA of exon 1 to exon 8, as well as from exon 7 to exon 8, mutant egfr mRNAs could not be detected in tissues harboring a CA-SSR I AI. The integrity of mRNA was approved by β-actin cDNA amplification. From a subgroup of tumors, Western blotting was also performed using a COOH-terminal binding antibody and an exon 2 to exon 7 sequence-specific antibody as a control. In agreement with the RT-PCR results, no truncated receptor proteins were detected.

From a comparison study with microsatellite PCR and CGH analysis we know that AI at distinct polymorphic loci can also represent amplifications (27). Therefore, we developed a quantitative real-time PCR assay for gene dosage measurements within the six CA-SSRs in intron 1 of egfr. Interestingly, an increase in CA-SSR I gene dosages could be detected in cases presenting with an AI score <0.56 and >1.3. The sensitivity of the method did not allow us to assess a gene amplification to all of the cases presenting with an AI score significantly different from diploids at the egfr locus, but in no case could a decreased gene dosage be measured. We interpret these data as a first suggestion that distinct amplifications of egfr occur in a substantial subset of invasive breast cancer cases. A basal gene dosage-dependent regulation might be regarded also for egfr as already established for c-erbB-2 (3).

Searching for published data that could explain an increase of probability for the formation of an amplification at that locus, we found that fragile sites were suggested at 7p11.2 and 7p14.2 (21), which have not been characterized in detail, thus far. Therefore, based on the knowledge that the induction of fragile sites triggers the formation of amplification we investigated chromosome 7p in normal lymphocyte metaphase spreads after drug treatment by cytogenetics and FISH analysis using an egfr-specific probe. In fact, a single-strand break could be detected in 30% of the metaphase spreads, and the FISH analysis confirmed the location near the egfr gene region (Fig. 4). The high frequency of this event suggests that this is a common fragile site. We are investigating currently the exact structure of the strand break at this fragile site to substantiate our working hypothesis that it is the initial event in the formation of egfr amplifications e.g., by breakage-fusion-bridge cycles (22). Of note, AI at CA-SSR I could be detected in morphologically normal breast tissue in three quarters of the patients harboring this AI at the corresponding invasive carcinoma. This was not unexpected because of the fact that AIs at normal breast lobules in women with or without synchronous breast cancer (28) have already been described. Therefore, we conclude that this genetic alteration might be a primary event in breast cancer carcinogenesis.

Furthermore, 20% of nontumorous hyperplasias and metaplasias, and 80% of the atypical ductal hyperplasias presented with an AI of the CA-SSR I. To our knowledge no other genomic alteration leading to gene deregulation has thus far been reported with such a high frequency in nontumorous breast tissue. The segregation of the AI was additionally detected in all of the cases of associated DCIS and also in all of the lymph node metastases.

For squamous cell carcinoma of the lower jaw, we could already demonstrate that egfr and c-erbB-2 alterations in morphologically normal mouth tissue have predictive value concerning the recurrence rate (29). Recently, a high-risk marker of local recurrence in early stage breast cancer patients has been narrowed to a region at chromosome 3p24.3 (30). Therefore, it could be hypothesized that the AI at CA-SSR I of egfr is a genetic marker, which is also important for breast cancer prognosis. Our retrospective cohort study using the CA-SSR I AI at the primary tumor as a prognostic marker demonstrated an increased risk of recurrence for patients harboring the AI. This leads to the assumption that tumors with a CA-SSR I AI present with a distinct phenotype, which is at least determined genetically in the early steps of carcinogenesis.

However, an overexpression of EGFR was not observed in all cases of CA-SSR I AI, thereby indicating only basal impact of such a mutation in an intronic regulatory sequence of the gene. Factors with repressor function may counteract the enhancement of egfr transcription activity by the amplifications.

In consequence a better understanding of this phenotype might lead to the discovery of a target for new therapeutic or even preventive approaches especially counteracting egfr expression and signaling.

Acknowledgments

We thank Katrin Blaschei and Ute Hinz for outstanding technical assistance. We also thank Dr. Rita Exeler, Institute for Human Genetics, Muenster, Germany, for helpful suggestions and experimental advices.

References


![Fig. 5. Life table analysis of egfr AI. Metastasis-free survival of breast cancer patients dichotomized by the occurrence of AI in the egfr CA SSR I. The differences between the life table curves were tested by log-rank test.](image-url)


Distinct Amplification of an Untranslated Regulatory Sequence in the egfr Gene Contributes to Early Steps in Breast Cancer Development

Nicola Tidow, Almuth Boecker, Hartmut Schmidt, et al.


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/63/6/1172

Cited articles
This article cites 28 articles, 16 of which you can access for free at:
http://cancerres.aacrjournals.org/content/63/6/1172.full#ref-list-1

Citing articles
This article has been cited by 11 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/63/6/1172.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, contact the AACR Publications Department at permissions@aacr.org.