Identification of Novel Gene Amplifications in Breast Cancer and Coexistence of Gene Amplification with an Activating Mutation of PIK3CA

Mitsutaka Kadota, Misako Sato, Beverly Duncan, Akira Ooshima, Howard H. Yang, Natacha Diaz-Meyer, Sheryl Gere, Shun-Ichiro Kageyama, Junya Fukuo, Takuya Nagata, Kazuhiro Tsukada, Barbara K. Dunn, Lalage M. Wakefield, and Maxwell P. Lee

Abstract
To identify genetic events that characterize cancer progression, we conducted a comprehensive genetic evaluation of 161 primary breast tumors. Similar to the “mountain-and-hill” view of mutations, gene amplification also shows high- and low-frequency alterations in breast cancers. The frequently amplified genes include the well-known oncogenes ERBB2, FGFR1, MYC, CCND1, and PIK3CA, whereas other known oncogenes that are amplified, although less frequently, include CCND2, EGF, FGFR2, and NOTCH3. More importantly, by honing in on minimally amplified regions containing three or fewer genes, we identified six new amplified genes: POLD3, IRAK4, IRX2, TBL1XR1, ASPH, and BRD4. We found that both the IRX2 and TBL1XR1 proteins showed higher expression in the malignant cell lines MCF10CA1h and MCF10CA1a than in their precursor, MCF10A, a normal immortalized mammary epithelial cell line. To study oncogenic roles of TBL1XR1, we performed knockdown experiments using a short hairpin RNA approach and found that depletion of TBL1XR1 in MCF10CA1h cells resulted in reduction of cell migration and invasion as well as suppression of tumorigenesis in mouse xenografts. Intriguingly, our mutation analysis showed the presence of activation mutations in the PIK3CA gene in a subset of tumors that also had DNA copy number increases in the PIK3CA locus, suggesting an additive effect of coexisting activating amino acid substitution and dosage increase from amplification. Our gene amplification and somatic mutation analysis of breast primary tumors provides a coherent picture of genetic events, both corroborating and novel, offering insight into the genetic underpinnings of breast cancer progression. [Cancer Res 2009;69(18):7357–65]

Introduction
Human cancers are characterized by gene mutations and chromosomal aberrations (1, 2). Breast cancer is the most common cancer of women in the United States and other western countries, with an accumulated life time incidence rate of about 11%.

Materials and Methods
DNA, RNA isolation, and DNA microarrays. Tissue samples were provided by the Cooperative Human Tissue Network, which is funded by the National Cancer Institute. The study was approved by the Institutional Review Board of the National Cancer Institute. The clinical pathologic data are described in Supplementary Table S1. Genomic DNA was prepared using the Qiagen DNA Mini kit (Qiagen, Inc.). RNA was isolated from the tissues using RNAzol B (Tel-Test, Inc.). We followed the original Affymetrix 500K or...
SNP5 protocols to obtain genotype data and copy number values. The Gene Expression Omnibus (GEO) accession number for these array data is GSE16619. All the primers used in PCR and sequencing are described in Supplementary Table S2.

### Cell culture, immunohistochemistry, and Western blots

MCF10A and MCF10AT cell lines are maintained in DMEM/F12 supplemented with 5% horse serum, 10 μg/mL mitomycin C, and 100 ng/mL cholera toxin. Culture media for cell proliferation. One straight scratch line was produced using a p200 pipette tip. The cells were washed once with PBS and then changed to culture medium without drug. A green fluorescent protein–targeting particle was isolated and used to transduce MCF10CA1h cells for 24 h.

### Short hairpin RNA knockdown of TBL1XR1 and functional analysis of the knockdown cells

Short hairpin RNA (shRNA) against TBL1XR1 RNA (TRCN0000060743), purchased from Open Biosystems, was transfected into the 293FT producer cell line using pPACKH1 Lentivector Packaging kit (System Biosciences) and Lipofectamine 2000 (Invitrogen). Particles were isolated and used to transduce MCF10CA1h cells for 24 h. The cells were grown for 48 h after transduction and then selected for stable transduced cells by the addition of 4 μg/mL puromycin for 7 d followed by culturing in medium without drug. A green fluorescent protein–targeting shRNA was processed similarly and used as a control (control-shRNA). Experiments for evaluating phenotypes of TBL1XR1 knockdown were performed within 2 wk using a pool of the transduced cells. Cell motility was analyzed with the scratch assay (13). Confluent cultures of the cells in six-well plates were treated with 10 μg/mL mitomycin C for 2 h to inhibit cell proliferation. One straight scratch line was produced using a p200 pipette tip. The cells were washed once with PBS and then changed to culture medium without mitomycin C. The width of the scratched area was measured at 0, 12, and 24 h after scratching. Cell invasion was analyzed using a tumor cell invasion system (BD Biosciences). Cells (2.5 × 10^5) were seeded in each insert well. After 60 h, the insert wells were washed and scraped according to the manual. The cells were fixed in methanol, and the nuclei were stained with hematoxylin and counted under a microscope. For in vivo mouse studies, cells were suspended in serum-free DMEM/F12 medium, and 5 × 10^5 cells were injected into the no. 2 and no. 7 mammary fat pads of 6- to 8-wk-old female athymic NCr nu/nu mice. Tumors were measured weekly with calipers and the volumes were calculated using the following formula: (short × long dimensions) / 2.

### Data analysis

Affymetrix CNAT4.0 was used to normalize microarray data and to generate CNstate and log2 ratios. Generally, we used the following default parameters: bandwith of 100 kb, transition_decay at 0.52 (14). The numbers of mice used were 5, 5, and 7 for MCF10CA1h, control-shRNA, and TBL1XR1-shRNA cell lines, respectively.

### Results and Discussion

#### Novel focal amplification regions

To gain a comprehensive understanding of the genetic events that delineate multiple stages of tumor progression, including hyperplasia, invasion, and metastasis, we performed DNA copy number analysis using the Affymetrix 500K or SNP5 SNP arrays on breast primary tumors. DNA copy number analysis was performed on 161 tumors, including 10 DCIS and 151 invasive breast cancers, 90 of which...
Figure 1. DNA copy number analysis of 161 primary breast tumors identified novel gene amplification events. We used the Affymetrix SNP arrays to identify potential oncogenes in regions of focal amplification. A, copy number estimation of ERBB2 by SNP array and qPCR was conducted using 76 tumors that are the subset of the 161 tumors. The regression line is described by $y = 0.1635x + 0.7479$ with $R^2 = 0.7965$. B, examples of three loci exhibiting gene amplification listed in Table 1, including TBL1XR1, IRX2, and NOTCH3/BRD4. Two tumor samples are shown for each locus. The graphs were generated by the Partek Genomics Suite. X axis, genomic position; Y axis, log2 ratio (tumor hybridization intensity divided by normal reference samples from HapMap project). The red lines highlight gene amplification regions. Gene annotation encompassing each amplification region is provided at the top of the graphs. C, gene expression measured by RT-qPCR. Gene expression was measured in 14 tumors. For seven of the tumors, adjacent normal samples were also analyzed. The expression values of tumors are normalized by the average value of the seven normal samples; hence, gene expression is indicated on the Y axis relative to this average value. The matched normal sample is connected to each corresponding tumor from the same patient by a straight line. *, matched normal and tumor samples; □, tumors without matched normal. Red symbols and lines, tumors showing amplification and their adjacent normal samples.
were positive for lymph node metastases (see Supplementary Table S1 for clinical information). We are interested in identifying genomic regions that showed copy number gain or loss. Given that there have been extensive studies on DNA copy number alterations in breast cancer, we chose to target our search to focal amplification events that affect a few hundred kb regions, in breast cancer, we chose to target our search to focal amplification regions, five loci contained single genes (Table 1; Fig. 1B; Supplementary Fig. S3) and two loci had three genes in the minimal overlapping region of amplified DNA fragments (Table 1). Some of these infrequently amplified genes are well-known oncogenes: CCND2, EGFR, FGFR2, and NOTCH3. Rare amplification of EGFR and FGFR2 in breast tumors was reported in a recent publication (10). The minimal region of amplification at the NOTCH3 locus also contains BRD4 and ABHD9. A recent study suggests that genes whose expression is regulated by BRD4 activation might correlate with breast cancer survival (17). To investigate the effect of DNA copy number gain on gene expression, we measured expression of BRD4 and NOTCH3 using reverse transcription-qPCR (RT-qPCR). BRD4 gene expression was frequently elevated in tumors (Fig. 1C). When BRD4 copy number gain is present, gene expression up-regulation is almost always observed (Fig. 1C). The expression level of BRD4 in normal tissue was always low, which suggested that up-regulation of BRD4 gene expression is relevant to tumorigenesis (Fig. 1C; the difference between normal and tumor in gene expression has \( P = 0.01151 \) by \( t \) test). NOTCH3 gene expression remained at a low level, comparable with the average value from the seven normal tissues, even for three tumors with high-level copy number gain. Moderate increases in gene expression were noted in a small subset of tumors but did not correlate with copy number gain (Fig. 1C).

In the preceding paragraphs, we have discussed some results of high-level focal amplification events, a few of them involving well-known oncogenes. Next, we concentrate on the characterization of novel oncogenes within the newly identified amplification loci.

Characterization of novel oncogenes. Four high-level focal amplification regions contained single potentially novel oncogenes: IRX2, TBL1XR1, POLD3, and ASPH (Fig. 1B; Supplementary Fig. S3). We focused detailed molecular characterization on two of these genes: TBL1XR1 and IRX2.

IRX2 is a member of the Iroquois homeobox transcription factor family, which is involved in developmental pattern formation in multiple organs such as the brain and heart (18, 19). The expression of IRX2 in mammary gland development is particularly interesting because the gene is expressed only in epithelial cells during development; IRX2 expression is absent from stromal cells and is reduced in differentiated ductal epithelial cells (20). In contrast, some breast cancers exhibit high levels of IRX2 expression (20). Our gene expression analysis of IRX2 also showed that IRX2 was up-regulated in some breast tumors, in at least one case in association with gene amplification (Fig. 1C). To characterize IRX2 protein expression in the MCF10A series of cell lines, we performed Western blot analysis and found that IRX2 protein was expressed at higher levels in the malignant cell lines MCF10CA1h and MCF10CA1a than in their precursor, MCF10A, a normal immortalized mammary epithelial cell line (Fig. 2A), suggesting that up-regulation of IRX2 might be involved in cancer progression. This
observation was further corroborated by immunohistochemistry, with more intense nuclear staining in MCF10CA1a cells than MCF10A cells (Supplementary Fig. S4).

To study oncogenic mechanisms of the IRX2 gene, we undertook RNA interference experiments using small interfering RNA as well as shRNA. Despite numerous attempts, we were not able to generate breast cancer cell lines that could maintain a stable low-level expression of the IRX2 protein. A possible explanation is that knockdown of IRX2 inhibits proliferation or survival of the breast epithelial cells. To gain insight into potential oncogenic functions of IRX2, we performed immunohistochemical studies on TMAs to investigate IRX2 protein expression in primary breast tumors. Positive staining was observed in 66 of 85 tumors (77.6%), with 20 moderately positive and 46 strongly positive tumors, suggesting an association of high-level IRX2 expression with breast carcinogenesis (Fig. 2B); 19 of the 85 tumors showed negative staining. We did not detect a statistically significant association of IRX2 expression (presence versus absence) with any of the clinical phenotypes, including stage, tumor size, and lymph node invasion (data not shown). However, when different degrees of expression intensity were analyzed among the 66 positively stained tumors, comparison of the 20 IRX2+ to the 46 IRX2++ cases revealed a positive correlation of degree of IRX2 staining with tumor size ($P = 0.0288$ by generalized linear model). This suggests that IRX2 may play a role in tumor cell proliferation and progression.

The second focally amplified gene that we characterized is TBL1XR1. Two recent studies showed that TBL1XR1 plays a pivotal role in releasing the repressive complex of corepressors NcoR and SMRT following oncogenic activation of multiple pathways, including the Wnt, Notch, NF-$\kappa$B, and nuclear receptor pathways (21, 22). Our reverse transcription-PCR analyses showed relatively constant, low levels of TBL1XR1 gene expression in most breast tumor and normal breast samples (Fig. 1C). Similar to IRX2, TBL1XR1 protein was primarily located in nuclei (Supplementary Fig. S4) and was detected in breast tumors that showed gene amplification (Supplementary Fig. S5). Western blot analysis showed that TBL1XR1 expression increased progressively from MCF10A to the malignant cell lines (Fig. 2A), suggesting a role for TBL1XR1 in cancer progression.

![Figure 3. Characterization of cell transformation in breast cancer cell lines with shRNA knockdown of the TBL1XR1 gene.](https://www.aacrjournals.org/)
protein bands were detected, corresponding to the α form (56 kDa) and β form (60 kDa) of TBL1XR1, which differ in their carboxyl end due to alternative splicing (23).

We further characterized TBL1XR1 in terms of its oncogenic functions using a lentiviral vector system to transduce MCF10CA1h cells with a shRNA targeting the TBL1XR1 gene. TBL1XR1 protein expression in shRNA-containing cells was examined by Western blot. Compared with parental cells or cells containing a control-shRNA, TBL1XR1-shRNA knockdown cells showed a nearly complete loss of TBL1XR1 protein expression (Fig. 3A).

In vitro cell growth was minimally reduced in TBL1XR1-shRNA cells (Supplementary Fig. S6); however, a more prominent change was

Figure 4. A, clustering analysis of gene amplification in 161 primary breast tumors. The log2 ratios of the 17 genes listed in Table 1 for the 161 tumors were used to perform clustering analysis, generating the dendrogram at the top of the figure and the heat map at the bottom. The tested genes from Table 1 are listed to the left of the heat map. The labels to the left of the middle portion of the figure are as follows: node (red, lymph node positive; blue, negative), invasion (red, invasive breast cancer; blue, noninvasive or DCIS), size (red, tumor size >5 cm; blue, <5 cm), and mutation status for PIK3CA, TP53, and AKT1 (red, presence of mutation; blue, absence of mutation). B, survival plots calculated by Kaplan-Meier analysis using expression data of four genes: POLD3, CCND1, FGFR1, and FGFR2. The four publicly available gene expression data sets from the GEO database are GSE4922, GSE2034, GSE1456, and GSE3494. Red and blue curves, good and poor survival patient groups, respectively. The two patient groups were determined by hierarchical clustering analysis based on gene expression values of these four genes as described in Results and Discussion.
observed in cell migration, as analyzed by the scratch assay (Fig. 3B). The difference in the cell migration between TBL1XR1-shRNA and control-shRNA experiments, quantified by the width of the scratched area, was highly significant \((P < 0.0001, t\) test; Fig. 3B). Because cell migration is related to tumor cell invasion, we further characterized the ability of the cells to invade a basement membrane using Matrigel Matrix system (BD Biosciences). TBL1XR1-shRNA knockdown cells showed a marked reduction in cell invasion when compared with control-shRNA cells (Fig. 3C). The difference in invasive cell numbers between control-shRNA and TBL1XR1-shRNA was highly significant \((P < 0.0001, t\) test). Given that tumor cell invasion is a hallmark of carcinoma cells, the loss of cell invasion associated with the TBL1XR1 knockdown is consistent with an oncogenic role for this gene. A more rigorous test for tumorigenesis is examination of \textit{in vivo} tumor growth. Therefore, we injected parental MCF10CA1h cells, control-shRNA cells, or TBL1XR1-shRNA cells into the mammary fat pads of nude mice. Mice injected with either the MCF10CA1h or control-shRNA cells started to develop tumors around 2 weeks (Fig. 3D). In contrast, mice injected with the cells containing TBL1XR1-shRNA showed a marked reduction in tumor growth \((P < 0.001, t\) test). Thus, our \textit{in vitro} and \textit{in vivo} studies of TBL1XR1 knockdown experiments provide strong supporting evidence that TBL1XR1 is a novel breast cancer oncogene.

Characterization of clinical phenotypes in relation to gene amplification. To characterize the relationship between gene amplification and clinical pathologic data, we performed two-way clustering analysis (Fig. 4A). Among the noteworthy observations, one cluster of tumors was defined by ERBB2 amplification (Fig. 4A, yellow). A small subset of these tumors also showed high or moderate MYC amplification (Fig. 4A, yellow or green). Some of the infrequent high-level amplification events involve sets of genes, which identify a small group of samples. One such set comprises CCND2, IRX2, IRAK4, PRDM1, PIK3CA, and TBL1XR1 (six-gene set) and another includes POLD3, CCND1.

![Figure 5](cancerres.aacrjournals.org) Concomitant activation mutation and gene amplification of PIK3CA in breast cancer. A, examples of PIK3CA amplification in five tumors. Formatting of A is the same as for Fig. 1B.
FGFR1, and FGFR2 (four-gene set). We explored the relationship of these amplification features to survival in an indirect manner. Having shown that a positive correlation exists between some of the high-level amplification events and increased gene expression (Fig. 1C), we used gene expression data from the public domain for patient samples for which survival data were available (24–27). We evaluated survival differences between the two groups of patients that formed at the top of the clustering tree using either the six-gene set or the four-gene set by Kaplan-Meier analysis. In Kaplan-Meier analysis, the six-gene set showed no difference in survival between the two groups (data not shown). In contrast, the four-gene set containing POLD3, CCND1, FGFR1, and FGFR2 showed a significant difference in survival in two of the four public data sets (Fig. 4B).

Because we have the clinical phenotypes for the 161 tumors analyzed for copy number variation, we performed association tests between gene amplification and clinical phenotype (using the data in Supplementary Tables S1, S3, and S4; only those with a P value of <0.05 in the Fisher’s exact test are included in Supplementary Table S5). As expected, ERBB2 amplification was positively associated with HER2+ status. We also found that FGFR1 amplification was positively associated with HER2+ status; CCND1 and POLD3 amplification was positively associated with PR+ status, whereas MYC amplification was negatively associated with PR+ status; ERBB2 amplification was positively associated with tumor size.

**Mutation analysis of cancer genes.** Recent large-scale mutation analyses of the genomes of multiple breast cancers revealed a mutation landscape consisting of mountains and hills (4). TP53 and PIK3CA were the only two genes that existed as mountains, with high mutation frequencies, whereas hundreds of other genes making up the hills showed rare mutations in the breast tumors. We conducted mutation analysis for the following five genes in 161 breast tumors: TP53 (exons 4–9), PIK3CA (exons 10 and 21), BRAF (exons 11 and 15), AKT1 (exon 3), and HRAS (exons 1 and 2). Consistent with published studies, only TP53, PIK3CA, and AKT1 showed frequent mutations. We identified 44 (27.3% of the 161 tumor samples) mutations in TP53, 25 (15.5%) mutations in PIK3CA, and 11 (6.8%) mutations in AKT1. The result of mutation analyses for TP53, PIK3CA, and AKT1 is shown in Fig. 4A (red colors mark tumors with a mutation) and Supplementary Table S4. Analysis of mutations and gene amplifications revealed TP53 mutations to be positively associated with gene amplification of PIK3CA, CCND2, and NCALD (Supplementary Table S5), which is consistent with the notion that the loss of TP53 causes genomic instability. Interestingly, PIK3CA mutation is also positively associated with PIK3CA amplification (Supplementary Table S5), a point that will be further discussed in the next section.

**Synergistic effect of PIK3CA amplification and mutations.** To evaluate whether an interaction also exists between activating mutations (Fig. 5B) and copy number gain of PIK3CA in primary breast tumors (Fig. 5A), we sequenced exons 10 and 21 of PIK3CA in the 161 tumors. We detected PIK3CA mutations in 25 of 161 tumors (Fig. 5B); 19 of 25 were H1047R and 4 were
E545K. This 15.5% mutation rate was comparable with that noted in previously published works. When we analyzed PIK3CA mutation in relation to copy number gain, we found that 5 of 10 tumors with copy number gain also harbored activation mutations (Fig. 5C). The simultaneous occurrence of an activating mutation and copy number gain was highly significant \((P = 0.008968, \text{Fisher's test; odds ratio, 6.4})\). Interestingly, those tumors with both copy number gain and mutation had moderate levels of gain and were enriched for E545K and other non-H1047R mutations (Fig. 5C). We noted that 3 of 3 tumors with non-H1047R mutations had copy number gain, whereas only 2 of 17 tumors with H1047R mutations had copy number increase (Fig. 5C). The result suggests that the H1047R mutation may have oncogenic features that are distinct from other PIK3CA mutations.

There are three recent published studies that characterize PIK3CA mutations extensively. Two show no difference in growth rate (28) and enzymatic activities (29) between H1047R and E545K mutations. But the third study describes the difference in growth rate (28) and enzymatic activities (29) of the PIK3CA gene, suggesting that an additive effect of point mutation and copy number gain can contribute to oncogenesis.

### Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

### Acknowledgments

Received 1/7/09; revised 5/29/09; accepted 7/6/09; published OnlineFirst 8/25/09.

Grant support: Intramural Research Program of the NIH and the National Cancer Institute.

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.

We thank the Cooperative Human Tissue Network for providing tumor tissues, Drs. Kent Hunter and Davud Meierzaman for critical reading of the manuscript, Dr. Robert Clifford for providing bioinformatics support, and Dr. Stephen Hewitt for providing TMA slides.

### References


In conclusion, we have identified the 17 loci focally amplified in primary breast tumors, 6 of which contain potential oncoprogens and reflect novel findings in this study. Among the genes representing these six loci, only rarely was amplification observed in primary tumors. However, these rare amplification events provided signposts that allowed us to functionally evaluate the potential oncogenic roles of these genes. To this end, we used the experimental approach of RNA interference to characterize the effect of gene knockdowns. This strategy can be applied to the other candidate cancer-causing genes identified in our study. We have also described a finding of simultaneous gene amplification and mutation of the PIK3CA gene, suggesting that an additive effect of point mutation and copy number gain can contribute to oncogenesis.
Identification of Novel Gene Amplifications in Breast Cancer and Coexistence of Gene Amplification with an Activating Mutation of PIK3CA

Mitsutaka Kadota, Misako Sato, Beverly Duncan, et al.