Published in final edited form as:
Cancer Res. 2009 September 15; 69(18): 7357–7365. doi:10.1158/0008-5472.CAN-09-0064.

Identification of novel gene amplifications in breast cancer and coexistence of gene amplification with an activating mutation of *PIK3CA*

Mitsutaka Kadota[1], **Misako Sato**[2], **Beverly Duncan**[1], **Akira Ooshima**[2], **Howard H. Yang**[1], **Natacha Diaz-Meyer**[1], **Sheryl Gere**[1], **Shun-Ichiro Kageyama**[4], **Junya Fukuoka**[4], **Takuya Nagata**[5], **Kazuhiro Tsukada**[5], **Barbara K Dunn**[3], **Lalage Wakefield**[2], and **Maxwell P Lee**[1,*]

[1] Laboratory of Population Genetics, CCR, NCI, USA

[2] Laboratory of Cancer Biology and Genetics, CCR, NCI, USA

[3] Basic Prevention Science Research Group, DCP, NCI, USA

[4] Laboratory of Pathology, Toyama University Hospital, Toyama, Japan

[5] Department of Surgery, University of Toyama, Toyama, Japan

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 *PIK3*CA, whereas other known oncogenes that are amplified, though less frequently, include *CCND2*, *EGFR*, *FGFR2*, and *NOTCH3*. More importantly, by honing in on minimally amplified regions containing ≤ 3 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 shRNA 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 co-existing 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.

Keywords

DNA copy number; gene amplification; oncogene; somatic mutation; breast cancer

* Correspondence to Maxwell P. Lee, Laboratory of Population Genetics, Bldg. 41, Rm D702, 41 Library Dr., Bethesda, MD 20892. Tel.301-435-8956, Fax 301-435-8963, e-mail address: leemax@mail.nih.gov.

Introduction

Human cancers are characterized by gene mutations and chromosomal aberrations (1,2). Breast cancer is the most common cancer of women in the U.S. and other western countries, with an accumulated life time incidence rate of about 11%. Approximately 180,000 new cases were estimated to occur in the U.S. in 2008. About 10% of breast cancers are inherited, mostly caused by mutations in *BRCA1* and *BRCA2*. The rest are sporadic breast cancers caused by somatic mutations and chromosome instability in the breast tissue. Breast cancer development is marked by multiple histopathologically discernable stages, including hyperplasia of mammary duct epithelial cells, ductal carcinoma in situ (DCIS), invasive tumor confined to the breast, lymph node involvement, and metastases to distant organs.

Several large initiatives have identified somatic mutations in breast cancers. These include screening for mutations located in protein kinase genes (3) as well as the more global approach of analyzing a nearly complete set of human genes (4,5). The latter investigations have demonstrated a bimodal distribution of mutations in breast and colon cancers. With regard to breast cancer these observations led to the proposal that the genomic landscape consists of "mountains" and "hills", the mountains corresponding to the most frequently mutated genes, specifically *TP53* and *PIK3CA*, and the hills consisting of hundreds of less frequently mutated cancer-associated genes. In addition to the identification of somatic mutations involving single base changes or small regions of DNA alteration, many studies of breast cancer have investigated genomic instability such as copy number alteration and DNA amplification and deletion affecting larger regions. Most of these studies of genomic alterations were conducted using array CGH (comparative genome hybridization) or cDNA arrays (some examples are described in references (6-9)). A couple of recent studies have used high density oligo arrays (10,11). The most commonly amplified regions include 8p11, 8q24, 11q13, 12q14, 17q11, 17q21, and 20q13, with amplification of oncogenes such as *ERBB2*, *MYC*, *CCND1*, and *MDM2* noted in multiple studies. However, most of these global genomic studies have not revealed any additional genes that contain alterations that potentially impact breast cancer development. Therefore, in this study we decided to look for focal amplification events that affect relatively small regions of genomic DNA, spanning a few hundred kb to a couple of Mb, with the goal of identifying novel oncogenes.

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 US National Cancer Institute. The clinical pathological data are described in Supplementary Table S1. Genomic DNA was prepared using the QIAamp DNA mini kit (Qiagen, Inc., Valencia, CA, USA). RNA was isolated from the tissues using RNAzol B (Tel-Test, Inc., Friendswood, TX, USA). We followed the original Affymetrix 500K or SNP5 protocols to obtain genotype data and copy number values. The 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 (IHC), and Western blots

MCF10A and MCF10AT cell lines are maintained in DMEM/F12 supplemented with 5% horse serum, 10 μg/ml insulin, 20 ng/ml EGF, 0.5 μg/ml hydrocortisone, and 100 ng/ml cholera toxin. Culture media for MCF10CA1h and MCF10CA1a cell lines are DMEM/F12 plus 5% horse serum. The tissue microarray (TMA) slides used in this paper were constructed at Toyama University according to a previously described method (12). Rabbit polyclonal anti-IRX2

antibody (cat: ARP32188_T100; Aviva System Biology, San Diego, CA, USA) and mouse monoclonal anti-TBL1XR1 antibody (cat: sc-100908; Santa Cruz Biotechnology, Santa Cruz, CA, USA) were used.

shRNA knockdown of *TBL1XR1* and functional analysis of the knockdown cells

Short hairpin RNA (shRNA) against *TBL1XR1* RNA (TRCN0000060743), purchased from Openbiosystems (Huntsville, AL, USA), was transfected into the 293FT producer cell line using pPACKH1 Lentivector Packaging kit (System Biosciences, Mountain View, CA, USA) and Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA). Pseudoviral particles were isolated and used to transduce MCF10CA1h cells for 24 hours. The cells were grown for 48 hours post-transduction and then selected for stable transduced cells by the addition of 4 μg/ml puromycin for 7 days followed by culturing in medium without drug. A GFP-targeting shRNA was processed similarly and used as a control (control-shRNA). Experiments for evaluating phenotypes of *TBL1XR1* knockdown were performed within 2 weeks using a pool of the transduced cells. Cell motility was analyzed with the scratch assay (13). **C**onfluent cultures of the cells in 6-well plates were treated with 10 μg/ml mitomycin C for 2 hours to inhibit cell proliferation. One straight scratch line was produced using a p200 pipet 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 hours post-scratching. Cell invasion was analyzed using a tumor cell invasion system (BD Biosciences, San Jose, CA, USA). 2.5×10^4 cells were seeded in each insert well. After 60 hours, the insert wells were washed and scrubbed 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 6-8 week-old female athymic NCr nu/nu mice. Tumors were measured weekly with calipers and the volumes were calculated using the formula: (short × short × long dimensions) ×0.52 (14). The numbers of mice used were 5, 5, and 7 for MCF10CA1h, control-shRNA, and TBL1XR1-shRNA cell lines, respectively.

Data analysis

Affymetrix CNAT4.0 was used to normalize microarray data and to generate CNstate and log2ratios. Generally, we used the default parameters: bandwith of 100 kb, transition_decay at 1e-7, and no outlier_smoothing. For focal amplification, we used the bandwith of 1 kb. The CNstate ranges from 0 to 4: normal CN corresponds to CNstate 2; CNstates 0 and 1 indicate copy number loss; CNstates 3 and 4 correspond to copy number gain. For the data generated with the Affymetrix SNP5.0, we formatted the log2ratio generated by quantile normalization so that the data could be visualized with the Affymetrix Genotyping Console Browser. For gene-level copy number estimation, we simply calculated the average log2ratio for all the probesets mapped within a gene, between the transcription start and termination sites of the gene. The gene-level log2ratio was used to identify gene amplification/deletion in each tumor. All the statistical analyses (clustering, survival analysis, Fisher's exact test, generalized linear models) were conducted using the R package.

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 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, anticipating that this search would facilitate identification of novel oncogenes. Table 1 summarizes a list of the genes that mapped to the regions with amplification events. Our selection criteria required that an amplification event be present in a minimum of 2 tumors, with the amplification being focal in at least one of these tumors. Among the 17 loci listed in Table 1, six genes, *ERBB2*, *FGFR1*, *MYC*, *CCND1*, *PIK3CA*, and *NCALD*, had frequent amplification (amplified in at least 10 tumors). We used quantitative PCR (qPCR) to evaluate independently the level of gene amplification. A correlation was observed between DNA copy number estimated from SNP arrays compared to qPCR (as an example, R^2 is 0.8 for *ERBB2*, Figure 1A. The example for *MYC* appears in Supplementary Figure S1).

We identified genetic alterations in the genes listed in Table 1. These included amplification of *ERBB2*, *FGFR1*, *MYC*, and *CCND1*, which have been extensively described in the literature. We also saw *PIK3CA* amplification, which has occasionally been noted in breast tumors; however, the primary reported alteration in this gene is somatic point mutation (15,16). Although overall our results are consistent with previous studies, we report specific novel findings below. For example, amplification of the *NCALD* locus on chromosome 8q22 is very frequent in our samples (16 out of 161 tumors); importantly, the involved region is distinct from the *MYC* region at 8q24 (Supplementary Figure S2). The minimal amplification region of *NCALD* spans 1.6 Mb and still contains six other genes (Table 1). The frequent amplification of this region suggests that it potentially harbors a novel oncogene.

Next, we turned our attention to high-level (i.e. high copy number) focal amplification events, with the intention of identifying novel oncogenes, even though these regions were amplified only infrequently in our tumor samples. Among the 11 infrequent amplification regions, five loci contained single genes (Table 1 and Figure 1B and Supplementary Figure 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 RT-qPCR. *BRD4* gene expression was frequently elevated in tumors (Figure 1C). When *BRD4* copy number gain is present, gene expression up-regulation is almost always observed (Figure 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 (Figure 1C, the difference between normal and tumor in gene expression has p-value=0.01151 by t-test). *NOTCH3* gene expression remained at a low level, comparable to 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 (Figure 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* (Figure 1B and Supplementary Figure 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, since 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 (Figure 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 (Figure 2A), suggesting that up-regulation of IRX2 might be involved in cancer progression. This observation was further corroborated by IHC, with more intense nuclear staining in MCF10CA1a cells than MCF10A cells (Supplementary Figure S4).

To study oncogenic mechanisms of the *IRX2* gene, we undertook RNA interference experiments using siRNA 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 IHC 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 (Figure 2B); 19 out 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 staining 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 co-repressors NcoR and SMRT following oncogenic activation of multiple pathways, including the Wnt, Notch, NF-κB, and nuclear receptor pathways (21,22). Our RT-PCR analyses showed relatively constant, low levels of *TBL1XR1* gene expression in most breast tumor and normal breast samples (Figure 1C). Similar to IRX2, TBL1XR1 protein was primarily located in nuclei (Supplementary Figure S4) and was detected in breast tumors that showed gene amplification (Supplementary Figure S5). Western blot analysis showed that TBL1XR1 expression increased progressively from MCF10A to the malignant cell lines (Figure 2A), suggesting a role for TBL1XR1 in cancer progression. Two 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 to parental cells or cells containing a control-shRNA, *TBL1XR1*-shRNA knockdown cells showed a nearly complete loss of TBL1XR1 protein expression (Figure 3A). *In vitro* cell growth was minimally reduced in *TBL1XR1*-shRNA cells (Supplementary Figure S6); however, a more prominent change was observed in cell migration, as analyzed by the scratch assay (Figure 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 (Figure 3B, p-value < 0.0001, t-test). Since 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 to control-shRNA cells (Figure 3C). The difference in invasive cell numbers between control-shRNA and TBL1XR1-shRNA was highly significant (Figure 3C, p-value < 0.0001). 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 *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 (Figure 3D). In contrast, mice injected with the cells containing TBL1XR1-shRNA showed a marked reduction in tumor growth (p-value<0.001, t-test). Thus, our *in vitro* and *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 pathological data, we performed two-way clustering analysis (Figure 4A). Among the noteworthy observations, one cluster of tumors was defined by *ERBB2* amplification (depicted as yellow in Figure 4A). A small subset of these tumors also showed high or moderate *MYC* amplification (depicted as yellow or green in Figure 4A). 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* (6-gene), and another includes *POLD3*, *CCND1*, *FGFR1*, and *FGFR2* (4-gene). 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 (Figure 1C), we utilized 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 6-gene set or the 4-gene set by Kaplan-Meier analyses. In Kaplan-Meier analysis the 6-gene set showed no difference in survival between the two groups (data not shown). In contrast, the 4-gene set containing *POLD3*, *CCND1*, *FGFR1*, and *FGFR2* showed a significant difference in survival in 2 out of the 4 public datasets (Figure 4B).

Since 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 less than 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 Figure 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 consist 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 (Figure 5B) and copy number gain of *PIK3CA* in primary breast tumors (Figure 5A), we sequenced exon 10 and exon 21 of *PIK3CA* in the 161 tumors. We detected *PIK3CA* mutations in 25 out of 161 tumors (Figure 5B); 19 of 25 were H1047R and 4 were E545K. This 15.5% mutation rate was comparable to that noted in previously published works. When we analyzed *PIK3CA* mutation in relation to copy number gain, we found that 5 out of 10 tumors with copy number gain also harbored activation mutations (Figure 5C). The simultaneous occurrence of an activating mutation and copy number gain was highly significant (p-value=0.008968, odds ratio 6.4, Fisher's test). 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 (Figure 5C). We noted that 3 out of 3 tumors with non-H1047R mutations had copy number gain whereas only 2 out of 17 tumors with H1047R mutations had copy number increase (Figure 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 demonstrates that the two mutations are associated with different prognoses for disease-free survival (30), suggesting different oncogenic mechanisms, which cannot be explained by the similar enzymatic activity and *in vitro* cell growth rate. The relevance of both qualitative and quantitative changes of *PIK3CA* to tumor progression was also supported by our observations that all 10 DCIS lesions, in contrast to multiple invasive breast cancers, had neither *PIK3CA* mutation nor copy number gains (Supplementary Tables S1, S3, and S4).

In conclusion, we have identified the 17 loci focally amplified in primary breast tumors, 6 of which contain potential oncogenes and reflect novel findings in this study. Among the genes representing these 6 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Cooperative Human Tissue Network for providing tumor tissues. We thank Drs. Kent Hunter and Daoud Meerzaman for critical reading of the manuscript. We thank Dr. Robert Clifford for providing bioinformatics support.

We thank Dr. Stephen Hewitt for providing tissue microarray (TMA) slides. This research was supported by the Intramural Research Program of the NIH and the National Cancer Institute.

References

1. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell 2000;100:57–70. [PubMed: 10647931]
2. Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. Nat Med 2004;10:789–99. [PubMed: 15286780]
3. Stephens P, Edkins S, Davies H, et al. A screen of the complete protein kinase gene family identifies diverse patterns of somatic mutations in human breast cancer. Nat Genet 2005;37:590–2. [PubMed: 15908952]
4. Sjoblom T, Jones S, Wood LD, et al. The consensus coding sequences of human breast and colorectal cancers. Science 2006;314:268–74. [PubMed: 16959974]
5. Wood LD, Parsons DW, Jones S, et al. The genomic landscapes of human breast and colorectal cancers. Science 2007;318:1108–13. [PubMed: 17932254]
6. Chin K, DeVries S, Fridlyand J, et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell 2006;10:529–41. [PubMed: 17157792]
7. Loo LW, Grove DI, Williams EM, et al. Array comparative genomic hybridization analysis of genomic alterations in breast cancer subtypes. Cancer Res 2004;64:8541–9. [PubMed: 15574760]
8. Naylor TL, Greshock J, Wang Y, et al. High resolution genomic analysis of sporadic breast cancer using array-based comparative genomic hybridization. Breast Cancer Res 2005;7:R1186–98. [PubMed: 16457699]
9. Pollack JR, Perou CM, Alizadeh AA, et al. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat Genet 1999;23:41–6. [PubMed: 10471496]
10. Adelaide J, Finetti P, Bekhouche I, et al. Integrated profiling of basal and luminal breast cancers. Cancer Res 2007;67:11565–75. [PubMed: 18089785]
11. Haverty PM, Fridlyand J, Li L, et al. High-resolution genomic and expression analyses of copy number alterations in breast tumors. Genes Chromosomes Cancer 2008;47:530–42. [PubMed: 18335499]
12. Fukuoka J, Fujii T, Shih JH, et al. Chromatin remodeling factors and BRM/BRG1 expression as prognostic indicators in non-small cell lung cancer. Clin Cancer Res 2004;10:4314–24. [PubMed: 15240517]
13. Cha D, O'Brien P, O'Toole EA, Woodley DT, Hudson LG. Enhanced modulation of keratinocyte motility by transforming growth factor-alpha (TGF-alpha) relative to epidermal growth factor (EGF). J Invest Dermatol 1996;106:590–7. [PubMed: 8617990]
14. Tang B, Vu M, Booker T, et al. TGF-beta switches from tumor suppressor to prometastatic factor in a model of breast cancer progression. J Clin Invest 2003;112:1116–24. [PubMed: 14523048]
15. Samuels Y, Wang Z, Bardelli A, et al. High frequency of mutations of the PIK3CA gene in human cancers. Science 2004;304:554. [PubMed: 15016963]
16. Wu G, Xing M, Mambo E, et al. Somatic mutation and gain of copy number of PIK3CA in human breast cancer. Breast Cancer Res 2005;7:R609–16. [PubMed: 16168105]
17. Crawford NP, Alsarraj J, Lukes L, et al. Bromodomain 4 activation predicts breast cancer survival. Proc Natl Acad Sci U S A 2008;105:6380–5. [PubMed: 18427120]
18. Christoffels VM, Keijser AG, Houweling AC, Clout DE, Moorman AF. Patterning the embryonic heart: identification of five mouse Iroquois homeobox genes in the developing heart. Dev Biol 2000;224:263–74. [PubMed: 10926765]
19. Matsumoto K, Nishihara S, Kamimura M, et al. The prepattern transcription factor Irx2, a target of the FGF8/MAP kinase cascade, is involved in cerebellum formation. Nat Neurosci 2004;7:605–12. [PubMed: 15133517]
20. Lewis MT, Ross S, Strickland PA, Snyder CJ, Daniel CW. Regulated expression patterns of IRX-2, an Iroquois-class homeobox gene, in the human breast. Cell Tissue Res 1999;296:549–54. [PubMed: 10370142]
21. Li J, Wang CY. TBL1-TBLR1 and beta-catenin recruit each other to Wnt target-gene promoter for transcription activation and oncogenesis. Nat Cell Biol 2008;10:160–9. [PubMed: 18193033]

22. Perissi V, Scafoglio C, Zhang J, et al. TBL1 and TBLR1 phosphorylation on regulated gene promoters overcomes dual CtBP and NCoR/SMRT transcriptional repression checkpoints. Mol Cell 2008;29:755–66. [PubMed: 18374649]

23. Zhang XM, Chang Q, Zeng L, Gu J, Brown S, Basch RS. TBLR1 regulates the expression of nuclear hormone receptor co-repressors. BMC Cell Biol 2006;7:31. [PubMed: 16893456]

24. Ivshina AV, George J, Senko O, et al. Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. Cancer Res 2006;66:10292–301. [PubMed: 17079448]

25. Miller LD, Smeds J, George J, et al. An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. Proc Natl Acad Sci U S A 2005;102:13550–5. [PubMed: 16141321]

26. Pawitan Y, Bjohle J, Amler L, et al. Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. Breast Cancer Res 2005;7:R953–64. [PubMed: 16280042]

27. Wang Y, Klijn JG, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 2005;365:671–9. [PubMed: 15721472]

28. Zhang H, Liu G, Dziubinski M, Yang Z, Ethier SP, Wu G. Comprehensive analysis of oncogenic effects of PIK3CA mutations in human mammary epithelial cells. Breast Cancer Res Treat 2008;112:217–27. [PubMed: 18074223]

29. Sugita H, Dan S, Kong D, Tomida A, Yamori T. A new evaluation method for quantifying PI3K activity by HTRF assay. Biochem Biophys Res Commun 2008;377:941–5. [PubMed: 18952065]

30. Barbareschi M, Buttitta F, Felicioni L, et al. Different prognostic roles of mutations in the helical and kinase domains of the PIK3CA gene in breast carcinomas. Clin Cancer Res 2007;13:6064–9. [PubMed: 17947469]

The abbreviations used are

SNP
: single nucleotide polymorphism

MEC
: mammary epithelial cell

qPCR
: quantitative PCR

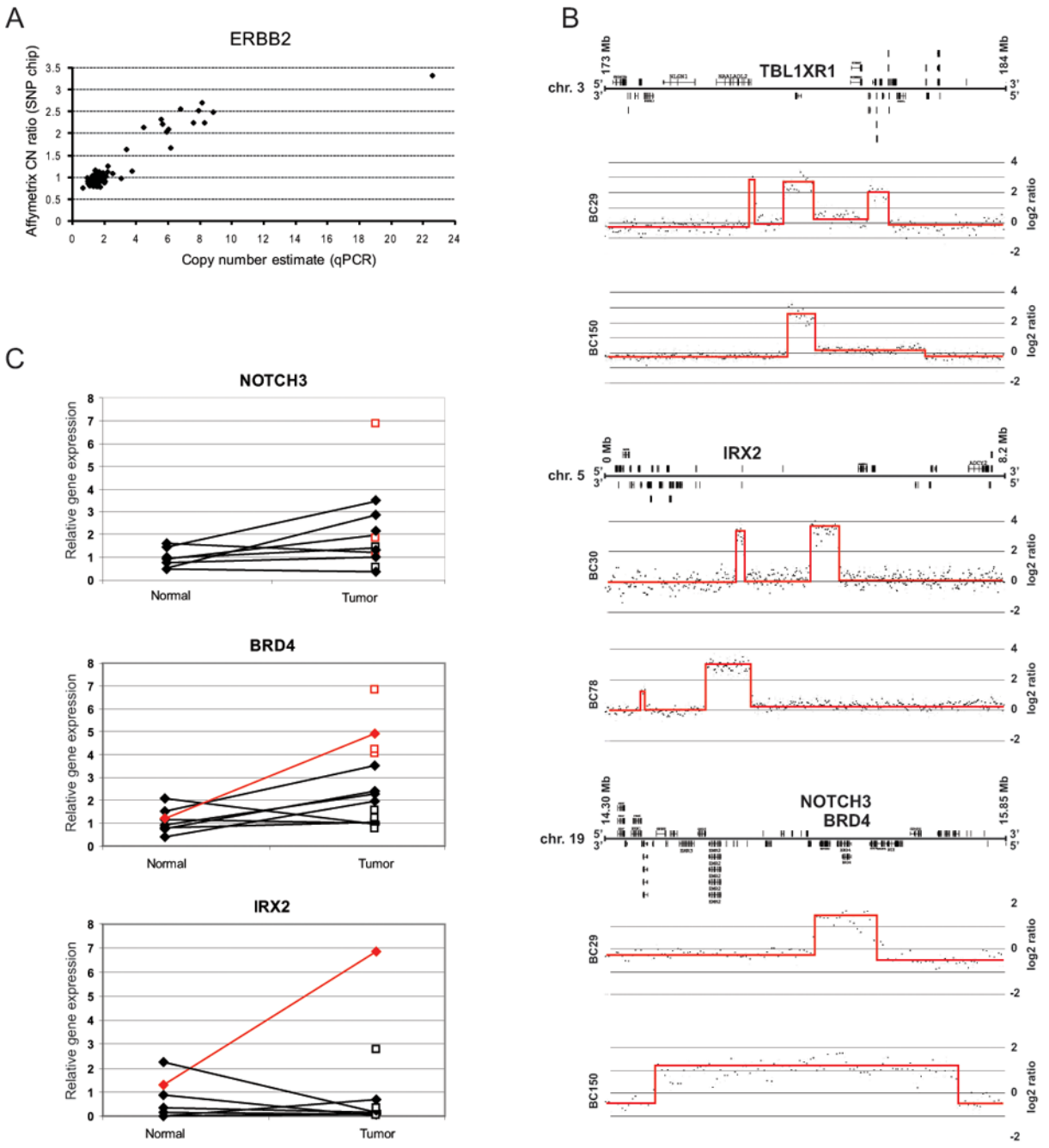

Figure 1. DNA copy number analysis of 161 primary breast tumors identified novel gene amplification events

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 were conducted using 76 tumors that are the subset of the 161 tumors. The regression line is described by y = 0.1635x + 0.7479 with R2 = 0.7965. **B**) Examples of three loci exhibiting gene amplification listed in Table 1 are shown here, including *TBL1XR1*, *IRX2*, and *NOTCH3*/*BRD4*. Two tumor samples are shown for each locus. The graphs were generated by the Partek Genomics Suite. Genomic position is displayed on x-axis and log2ratio (tumor hybridization intensity divided by normal reference samples from HapMap project) is on y-axis. 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 is shown. Gene expression was measured in 14 tumors. For 7 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. The matched normal and tumor samples are represented by filled diamonds, whereas tumors without matched normal are indicated by open squares. The tumors showing amplification and their adjacent normal samples are indicated by red symbols and lines.

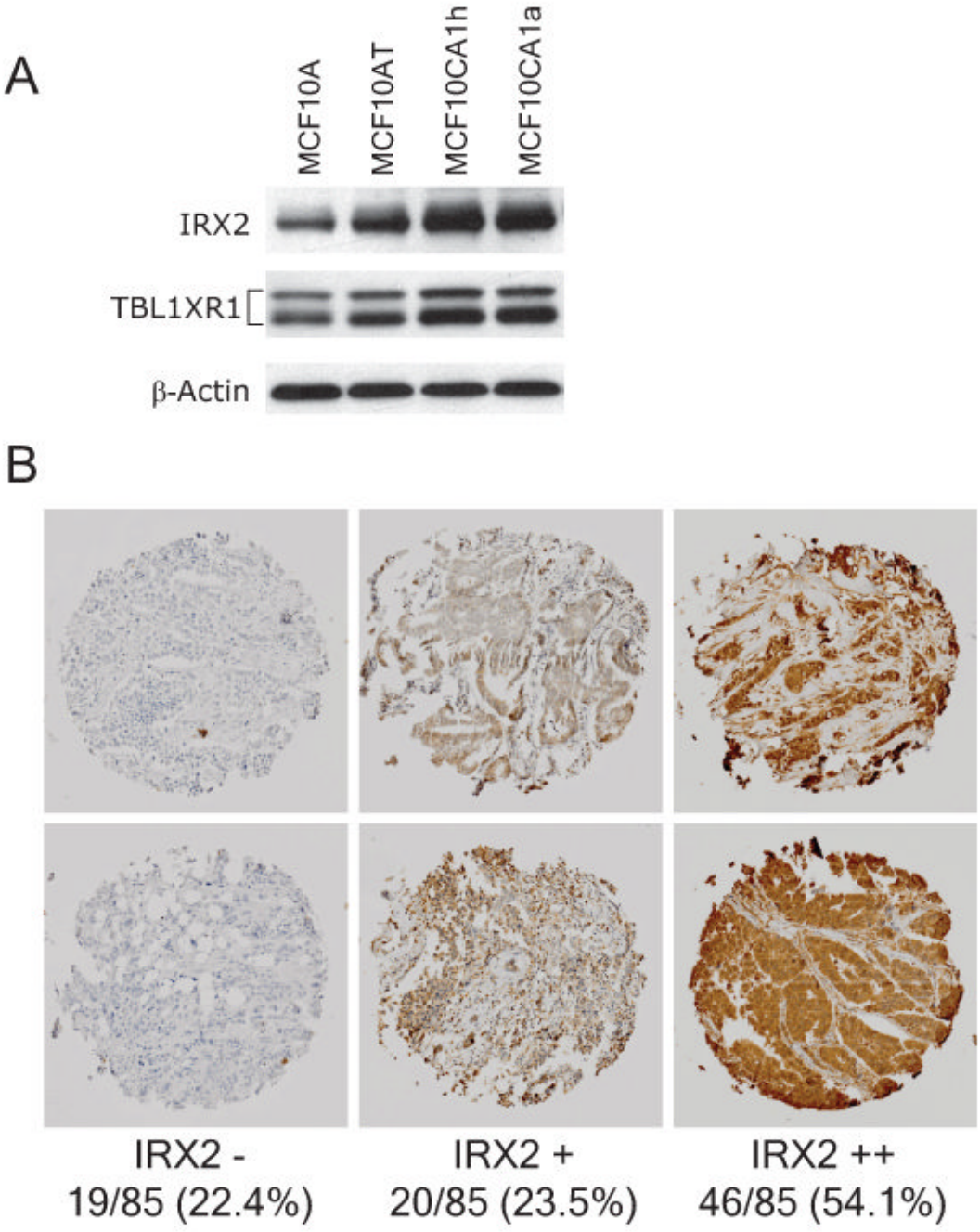

Figure 2. Western blot and Immunohistochemical (IHC) analyses of IRX2 and TBL1XR1

A) Protein expression, analyzed by Western blot, showed that both IRX2 and TBL1XR1 protein expression increased progressively from MCF10A to MCF10AT to the malignant cells, MCF10CA1h and MCF10CA1a. B) IHC with antibodies against IRX2 on tissue microarray (TMA) slides is shown. Two tumor cores for each negative ($IRX2^{-}$), weak-positive ($IRX2^{+}$) and strong-positive ($IRX2^{++}$) are shown together with numbers and percentage of each staining category among the 85 tumor samples.

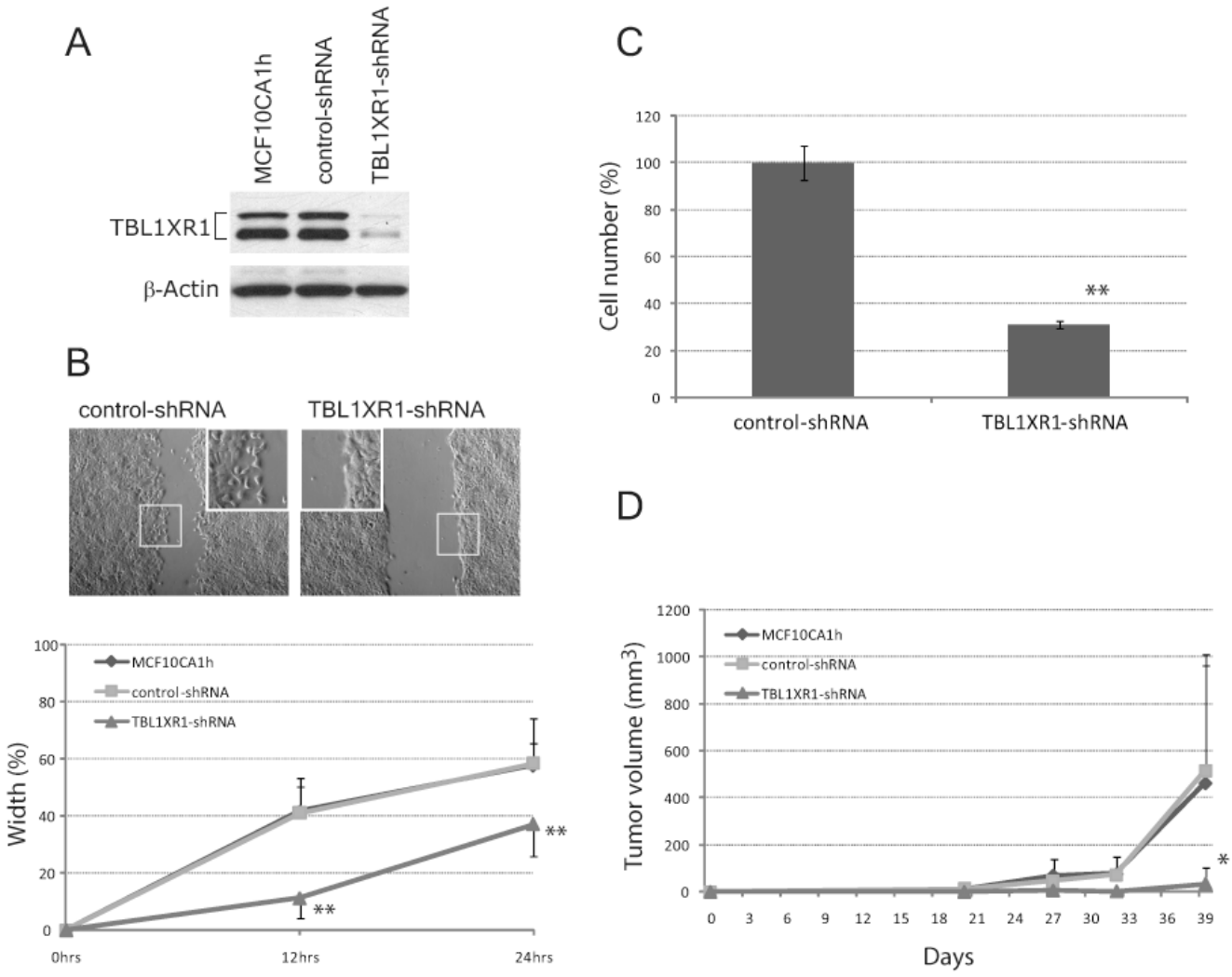

Figure 3. Characterization of cell transformation in breast cancer cell lines with shRNA knockdown of the *TBL1XR1* gene

TBL1XR1 knockdown in MCF10CA1h cells resulted in reduction of cell migration and loss of cell invasion examined by an *in vitro* cell culture system. A) Western blot analysis showed reduced TBL1XR1 protein in TBL1XR1-shRNA-transduced cells (lane labeled "TBL1XR1-shRNA"). B) *In vitro* scratch assays showed reduction of cell migration in TBL1XR1-shRNA-containing cells. Pictures were taken at 24 hours after scratching. In the image labeled "control-shRNA", the inset (higher-magnification) shows individual migrating cells at the front edge of the cell mass; in contrast, the edge of the scratched area appears smooth in the TBL1XR1-shRNA-transduced cells. The width of the scratched area was measured at 12 and 24 hours after scratching. Each point represents a mean width with SD (standard deviation) based on 6 measurements. C) Cancer cell invasion was assayed using the Matrigel Matrix system at 60 hours after cell plating. Numbers of cells were counted in 10 randomly selected areas under the microscope. The histogram illustrates the mean of these cell numbers +/- SD. D) Knockdown of *TBL1XR1* inhibited breast carcinoma development in mouse xenografts. MCF10CA1h cells begin forming detectable tumors as early as 14 days after injection, with tumor volumes increasing rapidly afterwards. A similar tumor growth curve was observed for control-shRNA. In contrast, the mice injected with the cells containing TBL1XR1-shRNA

showed marked reduction in tumor growth (t-test, p-value<0.001) compared to control-shRNA (marked by *).

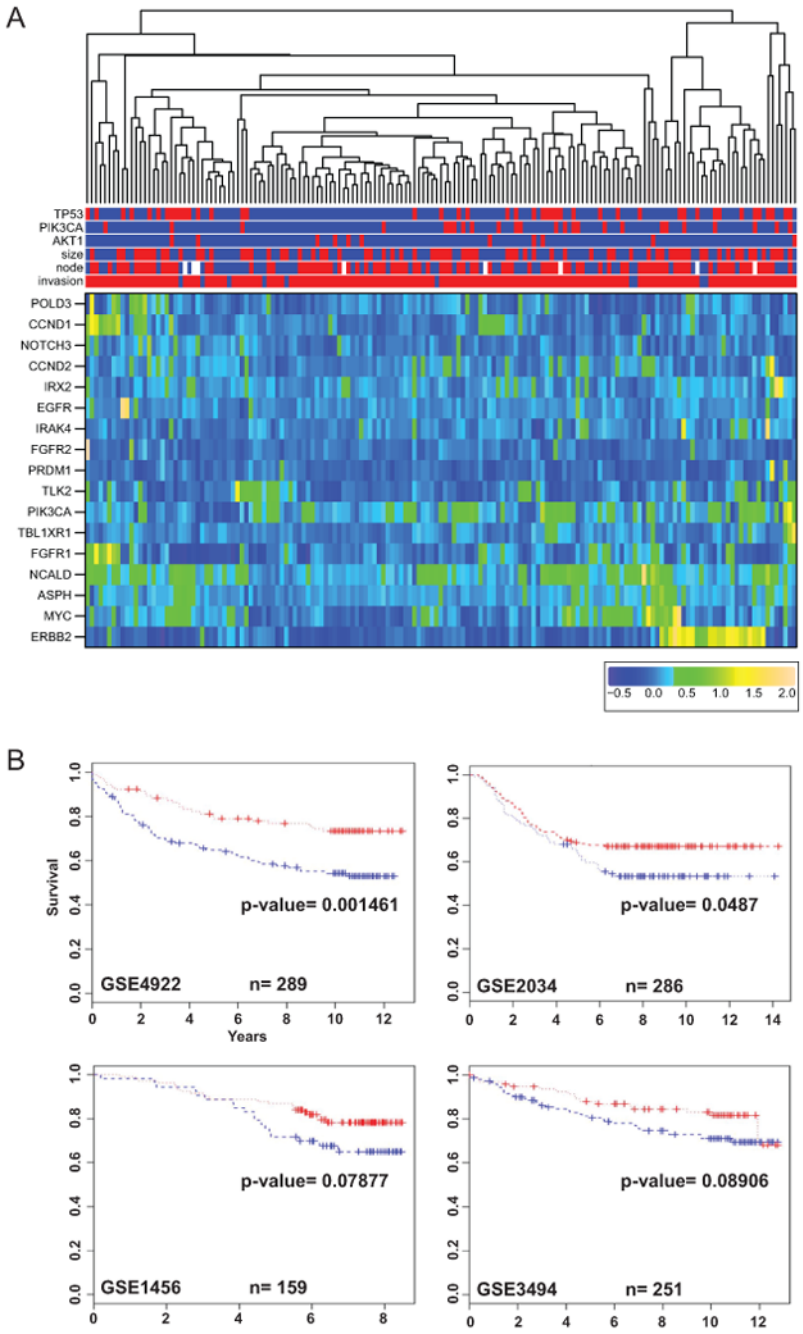

Figure 4.

Figure 4A. Clustering analysis of gene amplification in 161 primary breast tumors. The log2ratios 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: node (red for lymph node positive and blue for negative), invasion (red for invasive breast cancer and blue for noninvasive, or DCIS), size (red for tumor size over 5 cm and blue for less than 5 cm), and mutation status for *PIK3CA*, *TP53*, and *AKT1* (red for the presence of mutation and blue for absence of mutation). **Figure 4B. Survival plots calculated by Kaplan-Meier analysis using expression data of 4 genes *POLD3,***

CCND1, FGFR1,* and *FGFR2. The four publicly available gene expression datasets from the GEO database are GSE4922, GSE2034, GSE1456, and GSE3494. The red and blue curves represent good and poor survival patient groups, respectively. The two patient groups were determined by hierarchical clustering analysis based on gene expression values of these 4 genes as described in Results and Discussion section.

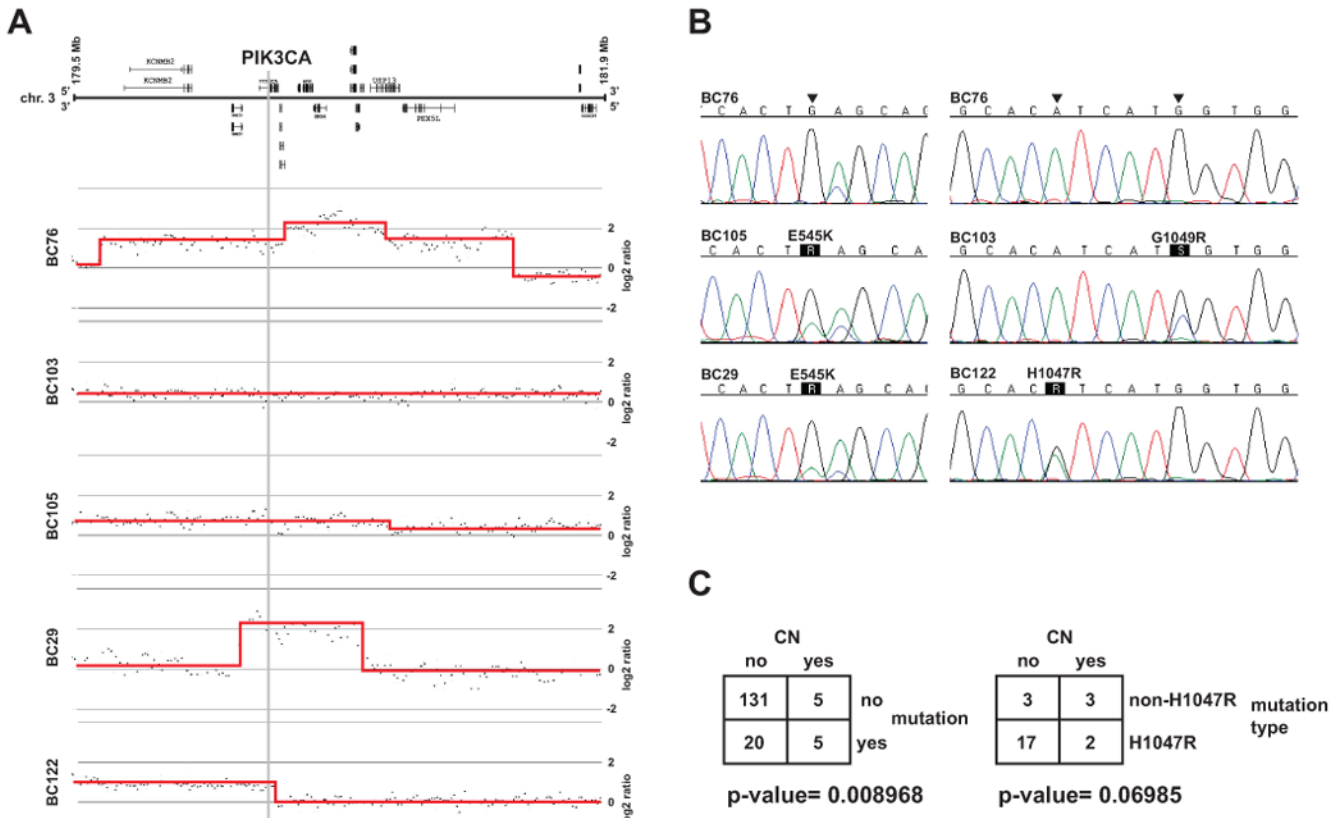

Figure 5. Concomitant activation mutation and gene amplification of *PIK3CA* in breast cancer

A) Examples of *PIK3CA* amplification in five tumors are illustrated. Formatting of Figure 5A is the same as for Figure 1B. **B**) The mutation analysis for these five tumors is depicted. Two exons (10 and 21) of *PIK3CA* showed mutations. The tumors in Figures 5A and 5B are matched and displayed in the same order from top to bottom. The first tumor (at the top), which exhibits high-level copy number gain, lacks a *PIK3CA* mutation in both exons 10 and 21. The arrowheads in the first tumor mark the bases with the mutations, E545K (exon 10), H1047R (exon 21), and G1049R (exon 21), identified in other tumors. The positions of the mutations in these last four tumors are highlighted by black boxes, with the mutations labeled by the amino acid substitutions. **C**) Association of *PIK3CA* mutation and gene amplification are summarized here. On the left side, it displays a 2×2 contingency table showing the number of tumors in each of the four categories: copy number gain only; *PIK3CA* mutation only; copy number gain plus *PIK3CA* mutation; neither copy number gain nor *PIK3CA* mutation. On the right side, it displays a 2×2 contingency table showing only tumors with a *PIK3CA* mutation. Copy number gain is depicted in relation to type of *PIK3CA* mutation. The following four categories are included: copy number gain with an H1047R mutation; copy number gain with a non-H1047R mutation; an H1047R mutation without copy number gain: a non-H1047R mutation without copy number gain.

Table 1

Chromosome	minimally overlapping amplification region (kb)			loci marked by gene*	gene_ID	number of tumors with gene amplification^	other genes in minimally overlapping amplificatin region
	region_start	region_end	interval				
3	178000	178800	800	TBL1XR1	79718	3	
3	180200	180800	600	PIK3CA	5290	10	KCNMB3 ZNF639 MFN1 GNB4 ACTL6A MRPL47
5	2700	3000	300	IRX2	153572	3	CEI
6	106200	108200	2000	PRDM1	639	2	ATG5 AIM1 QRSL1 MIRN587 PDSS2 SCML4
7	54900	55200	300	EGFR	1956	3	
8	36500	38800	2300	FGFR1	2260	12	ASH2L LSM1 etc ~ 20 genes
8	62400	62800	400	ASPH	444	5	
8	102200	103800	1600	NCALD	83988	16	ZNF706 GRHL2 RRM2B EDD1 ODF1 KLF10
8	128300	128800	500	MYC	4609	11	PVT1
10	122700	123700	1000	FGFR2	2263	2	ATE1
11	68900	69300	400	CCND1	595	11	ORAOV1 FGF19 FGF4
11	73900	74050	150	POLD3	10714	6	
12	4100	5300	1200	CCND2	894	3	FGF23 FGF6 etc ~ 10 genes
12	42200	42600	600	IRAK4	51135	3	PUS7L TWF1
17	35090	35170	80	ERBB2	2064	23	GRB7
17	57800	58500	700	TLK2	11011	3	EFCAB3 METTL2A MRC2 RNF190 MIRN633
19	15110	15260	150	BRD4	4854	2	NOTCH3 ABHD9

*Genes highlighted with red, green or blue are novel findings. Red indicates a single gene in the minimal amplification region/locus. Green indicates 3 genes in the minimal amplification region/locus. Blue indicates more than 3 genes in the minimal amplification region/locus.

^The number of tumors with gene amplification are tumors that have log2ratio > 0.6.