Systems consequences of amplicon formation in human breast cancer

Koichiro Inaki, Francesca Menghi, Xing Yi Woo, et al.

Supplemental Material	http://genome.cshlp.org/content/suppl/2014/08/06/gr.164871.113.DC1.html
References	This article cites 44 articles, 15 of which can be accessed free at: http://genome.cshlp.org/content/24/10/1559.full.html#ref-list-1
Open Access	Freely available online through the *Genome Research* Open Access option.
Creative Commons License	This article, published in *Genome Research*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**

Research

Systems consequences of amplicon formation in human breast cancer

Koichiro Inaki,[1,2,9] Francesca Menghi,[1,2,9] Xing Yi Woo,[1,9] Joel P. Wagner,[1,2,3]
Pierre-Étienne Jacques,[4,5] Yi Fang Lee,[1] Phung Trang Shreckengast,[2]
Wendy WeiJia Soon,[1] Ankit Malhotra,[2] Audrey S.M. Teo,[1] Axel M. Hillmer,[1]
Alexis Jiaying Khng,[1] Xiaoan Ruan,[6] Swee Hoe Ong,[4] Denis Bertrand,[4] Niranjan Nagarajan,[4]
R. Krishna Murthy Karuturi,[4,7] Alfredo Hidalgo Miranda,[8] and Edison T. Liu[1,2,7]

[1]Cancer Therapeutics and Stratified Oncology, Genome Institute of Singapore, Genome, Singapore 138672, Singapore; [2]The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut 06030, USA; [3]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA; [4]Computational and Systems Biology, Genome Institute of Singapore, Genome, Singapore 138672, Singapore; [5]Université de Sherbrooke, Sherbrooke, Québec, J1K 2R1, Canada; [6]Genome Technology and Biology, Genome Institute of Singapore, Genome, Singapore 138672, Singapore; [7]The Jackson Laboratory, Bar Harbor, Maine 04609, USA; [8]National Institute of Genomic Medicine, Periferico Sur 4124, Mexico City 01900, Mexico

Chromosomal structural variations play an important role in determining the transcriptional landscape of human breast cancers. To assess the nature of these structural variations, we analyzed eight breast tumor samples with a focus on regions of gene amplification using mate-pair sequencing of long-insert genomic DNA with matched transcriptome profiling. We found that tandem duplications appear to be early events in tumor evolution, especially in the genesis of amplicons. In a detailed reconstruction of events on chromosome 17, we found large unpaired inversions and deletions connect a tandemly duplicated ERBB2 with neighboring 17q21.3 amplicons while simultaneously deleting the intervening BRCA1 tumor suppressor locus. This series of events appeared to be unusually common when examined in larger genomic data sets of breast cancers albeit using approaches with lesser resolution. Using siRNAs in breast cancer cell lines, we showed that the 17q21.3 amplicon harbored a significant number of weak oncogenes that appeared consistently coamplified in primary tumors. Down-regulation of BRCA1 expression augmented the cell proliferation in ERBB2-transfected human normal mammary epithelial cells. Coamplification of other functionally tested oncogenic elements in other breast tumors examined, such as RIPK2 and MYC on chromosome 8, also parallel these findings. Our analyses suggest that structural variations efficiently orchestrate the gain and loss of cancer gene cassettes that engage many oncogenic pathways simultaneously and that such oncogenic cassettes are favored during the evolution of a cancer.

[Supplemental material is available for this article.]

Cancer arises from normal cells progressively accumulating both genetic and epigenetic changes that circumvent cellular regulatory controls (Hanahan and Weinberg 2011). Genetic changes represent the "hard-wiring" of a cancer resulting in a genomic configuration that provides an evolutionary advantage for tumor growth in the patient. Genetic changes through somatic mutations span from single nucleotide alterations to breakpoints interrupting gene contiguity, segmental copy number changes, and imbalances of entire chromosomes. The spectrum of point mutations and small insertions/deletions in breast cancers has been extensively studied and revealed a limited number of genes that are frequently mutated, such as TP53, PIK3CA, KRAS, and a larger number of rare mutations with many of unclear importance (Banerji et al. 2012; The Cancer Genome Atlas Network 2012; Shah et al. 2012; Stephens et al. 2012). The frequency of copy number gain and loss and their relationship with patient prognosis have also been examined in breast cancer (Chin et al. 2006; Hicks et al. 2006;

Beroukhim et al. 2010; Jönsson et al. 2010; Russnes et al. 2010; Staaf et al. 2010; Curtis et al. 2012). Although in earlier studies the major focus was to find specific driver oncogenes in amplicons and tumor suppressor genes in common regions of loss (primarily using loss of heterozygosity mapping), progressively there emerged an understanding that more than one driver oncogene may be present in any amplicon. Moreover, each amplicon or region of copy number loss alters the expression of many adjacent genes, some with proven conjoint cancer effects (Zhang et al. 2009; Curtis et al. 2012). Thus, any cancer is likely to be a composite of hundreds to thousands of gene changes that contribute to the cancer state. Although specific point mutations contribute to adaptive evolutionary processes, recent genomic analyses from controlled evolutionary experiments in model systems suggest that copy number changes through segmental duplications and rearrangements may play a more prominent role (Chang et al. 2013; Fares et al. 2013).

Previously, we investigated the transcriptional consequences of somatic structural variations in primary breast cancers and cell

lines (Hillmer et al. 2011; Inaki et al. 2011). Our observations showed that in one amplified locus, a single tandem duplication led to the selection of cancers with augmented expression of two adjacent oncogenic components. Generalizing from this observation, we hypothesize that conjoint expression of adjacent oncogenes or loss of adjacent cancer suppressors may provide a "systems genetic" strategy for activating cassettes of oncogenic signals. Herein, we seek to assess the importance of these structural variations in the establishment of a transcriptional state that might be optimal for the evolution of an individual cancer.

Results

Genomic organization and its transcriptional consequences in a set of breast tumors

To investigate recurrent molecular paths to breast oncogenesis, we carried out whole-genome sequencing of ~10-kbp insert size–mate pair libraries generated from the DNA of eight breast tumor samples and their normal counterparts, representing different breast cancer molecular subtypes (Supplemental Table 1). Because we were interested in the dynamics of clinically relevant amplicon formation, five of the eight tumors had histologic evidence for *ERBB2* amplification. The large insert size provided a significant depth of coverage (170× average physical coverage across the 16 sequenced genomes) (Supplemental Table 2) for the sensitive detection of genomic structural variations. We complemented the genomic data with transcriptomic information generated by either RNA sequencing or microarray (Supplemental Tables 2, 3). Circos plots representing somatic structural variations and copy number changes are shown in Supplemental Figure 1.

On average, 22.1% of each haploid cancer genome appeared to be affected by a focal change in copy number (from 1.9% to 43.1%) (Supplemental Methods), indicating that copy number variations are major contributors to the somatic changes in the genomic landscape of the analyzed breast tumors.

We confirmed previous evidence that copy number changes substantially affect the transcriptional output of cancer genomes (Curtis et al. 2012) by testing the expression dynamics of regions of copy number loss and gain: In each of the eight individual transcriptomes examined, genes affected by an increase in copy number were generally expressed at high absolute levels, whereas genes affected by a decrease in copy number were skewed toward the low expression range (Supplemental Fig. 2).

Structural variations and fusion genes

To gain insight into the biological relevance of specific patterns of copy number alterations, we focused on the functional genomic consequences of structural variations, which are likely to initiate massive copy number changes. Because the mate-pair sequencing method is able to detect all forms of structural variations quantitatively and with high sensitivity (Hillmer et al. 2011; Inaki and Liu 2012), we examined the rearrangements of each tumor in detail following the subtraction of its germline structural variations (Supplemental Methods).

Each tumor exhibited proclivity for specific kinds of somatic structural variations. For example, 74% of structural variations in BT122, which is a triple negative breast cancer, are tandem duplications scattered across entire chromosomal regions (Supplemental Fig. 3; Supplemental Table 4). A similar configuration has been recently described as the tandem duplicator phenotype (Ng et al.

2012). We also observed evidence of chromothripsis, another type of genome-scale structural pattern (Stephens et al. 2011), on chromosome 22 in BT55 (Supplemental Fig. 1), which appears as a large number of clustered rearrangements associated with frequent oscillation between two discrete copy number states. As an example of the high-resolution structural details afforded by the mate-pair sequencing approach, we detected an inversion at the *17p13.1* locus, which truncated the *TP53* gene body in BT122 and was associated with low expression of the gene (Supplemental Fig. 4A). In BT73, an unpaired inversion juxtaposed the promoter of a highly expressed gene, *EIF3H* (734.4 FPKM), next to *MYC*, which was overexpressed in this sample (114.7 FPKM) (Supplemental Fig. 4B).

Previous observations from our group suggested that somatic structural variations most frequently affect gene structures (Inaki et al. 2011). By cataloging fusion genes and gene truncations (fusion gene, 3′ truncation, 5′ truncation, and intragenic rearrangement) caused by structural variations in four of the sequenced breast cancers, we confirmed that nearly 60% of somatic structural variations affect gene structures (Supplemental Table 5). We tested 61 predicted gene fusions by RT-PCR and found that 37 of them (61%) were expressed (Supplemental Table 6). Of these, only eight fusion transcripts are in-frame and possess intact protein functional domains, although none are recurrent fusions, and the fusion partners have little plausible involvement in cancer, so that their functional consequences remain unknown. These findings are consistent with our previously published observations, which noted such fusion events that do not directly create a functional oncogene through the fusion but are indicative of a genomic process that initiates more critical downstream oncogenic mutations (Inaki et al. 2011; Inaki and Liu 2012).

Tandem duplications can function as initiators of gene amplification

We previously reported that genes mapping to tandem duplications' breakpoints are generally expressed at high levels (Inaki et al. 2011). Another possible oncogenic function of such segmental duplications is the consequent increased gene expression of the oncogenes they span. Indeed, we observed a 1.74-fold increase in the expression level of genes mapping inside compared to outside tandem duplications in the breast cancers analyzed herein (P-value $< 1.00 \times 10^{-4}$) (Fig. 1A).

We previously noted that tandem duplications appear frequently in regions of higher segmental genomic amplification and speculated that they may be integral to the formation of an amplicon (Hillmer et al. 2011). We used the cluster size of the structural variations, which is the number of discordantly mapped mate-pair reads corresponding to a given structural variation, as a measure of the prevalence of the mutated allele within the tumor cell population and therefore, as an indication of the probable chronological sequence by which different somatic rearrangements occurred in the cancer genome. This allowed us to estimate the genealogy of amplifications in the breast tumors examined.

Five of the eight breast tumors we analyzed (BT47, BT55, BT58, BT146, and BT191) displayed evident gene amplification (24, 14, 7, 26, and 22 copies, respectively) with associated high levels of expression of the *ERBB2* gene on chromosome 17 (Chr 17) (Supplemental Table 1). In four of these tumors, we found tandem duplications spanning the *ERBB2* locus and surrounding its amplicon; and in at least three cases, these events were associated with locally high cluster sizes (cluster sizes 207 in BT55; 359, 377, and 426 in BT146; and 45 in BT47) (Fig. 1B; Supplemental Fig. 5).
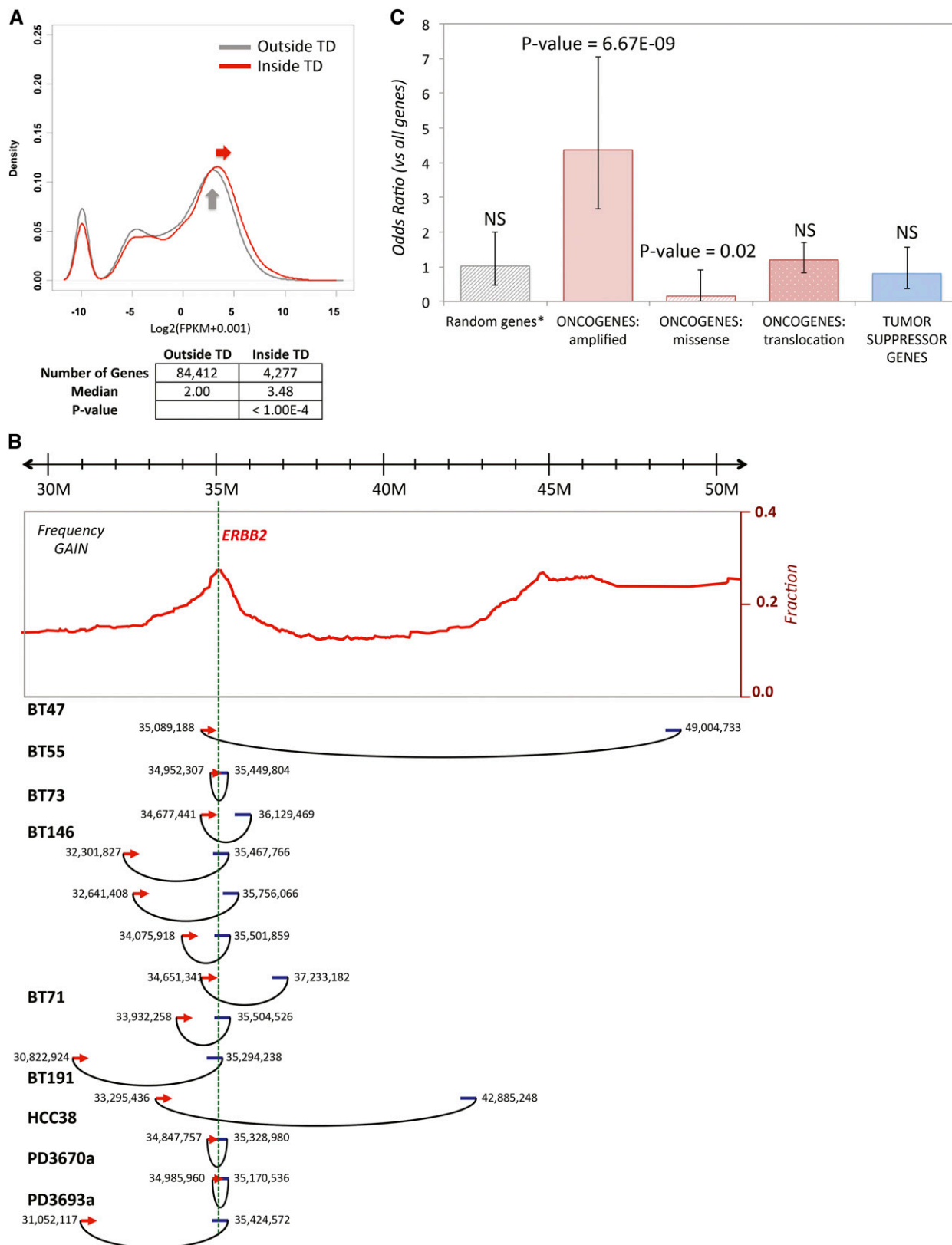
Figure 1. Tandem duplications are involved in amplicon formation in breast cancer. (*A*) Density plot of expression values of genes located inside and outside tandem duplications (TD) for the four breast tumors for which RNA-seq data is available. A *P*-value was computed based on 10,000 random gene permutations. (*B, upper*) frequency of copy number gain around the *ERBB2* locus on Chr 17 across a data set of 1174 breast cancers; (*lower*) genomic coordinates corresponding to tandem duplications spanning the *ERBB2* locus identified in nine different breast cancer samples. (*C*) Amplified oncogenes are significantly enriched for genes that are recurrently spanned by tandem duplications. Bars correspond to the fold enrichment compared to the all-gene population, with error bars indicating the 95% confidence interval. (NS) Not significant; (*) average and standard deviation values based on 100 gene permutations.

This suggests that a tandem duplication affecting the *ERBB2* gene may occur as an early event during *ERBB2* amplicon formation. Similarly, we observed tandem duplications spanning *ERBB2* in other samples—two ERBB2-positive tumors (PD3693a and PD3670a) and one triple negative breast cancer cell line (HCC38)—which have been independently sequenced at a lower physical coverage (Fig. 2B; Stephens et al. 2009). In BT71 and BT73, although no overt *ERBB2* amplification was clinically reported (Supplemental Fig. 1), we found evidence of a moderate copy number increase at the *ERBB2* locus (~3 gene copies), indicative of an early stage of amplicon formation. In both tumors, the *ERBB2* locus was spanned by a large tandem duplication (cluster size 141 in BT71 and 44 in BT73) (Fig. 2B; Supplemental Fig. 5).

Consistent with these results, when looking at the *CDK4*-amplified tumor sample BT122, we found that one of the first somatic rearrangements to take place on chromosome 12p is a large tandem duplication encompassing the *CDK4* locus and preceding a series of unpaired inversions and additional tandem duplications (Supplemental Results; Supplemental Fig. 6). Based on these observations, we hypothesize that one of the functional consequences of tandem duplications may be to nucleate and shape oncogenic amplicons.

If tandem duplications are indeed often involved in the earliest stages of amplicon formation, we would expect to find genes frequently amplified in cancer to be more often affected by tandem duplications than other functionally relevant sets of genes. We generated a list of genes that are most commonly affected by tandem duplications based on a catalog of 96 breast cancer genomes (Supplemental Methods) and calculated the level of enrichment for specific kinds of cancer-associated genes. Whereas no significant association was observed between genes in tandem duplications with oncogenes activated solely by somatic interchromosomal translocation or point mutation, a significant 4.4-fold enrichment was found in tandem duplications for oncogenes that are activated by focal amplifications (P-value = 6.67×10^{-9}) (Fig. 1C). This is further support for a model of amplicon formation in which a single tandem duplication spanning an oncogene is the initiating event in a series of structural rearrangements, which eventually generate additional copies of an oncogenic locus.

## Coordinated rearrangement of Chr 17 structure in breast cancer

In three cases (BT47, BT55, and BT146) with highly discernable *ERBB2* gene amplification, the copy number assessment suggested that a sharp peak of amplification at the *ERBB2* locus was immediately adjacent to a region of copy number loss that included the *BRCA1* gene on Chr 17 (Fig. 2). This region of DNA loss was physically followed by a distinct plateau of amplification at *17q21.3* extending toward the telomere (43M–46M, hg18), suggesting that a common complex chromosomal configuration involving the entire 17q arm may be part of the *ERBB2* amplification process. When we examined the breakpoints within this amplification-loss-amplification region on Chr 17, in all three cases (BT47, BT55, and BT146) the *ERBB2* amplicon was directly connected to the *17q21.3* amplicon by large unpaired inversions (Fig. 2B). We also found such coamplifications to be joined by an unpaired inversion in the HCC2218 breast cancer cell line, which shares a similar 17q copy number profile (Fig. 2B; Stephens et al. 2009), supporting the hypothesis of a recurrent genomic mechanism among breast cancers.

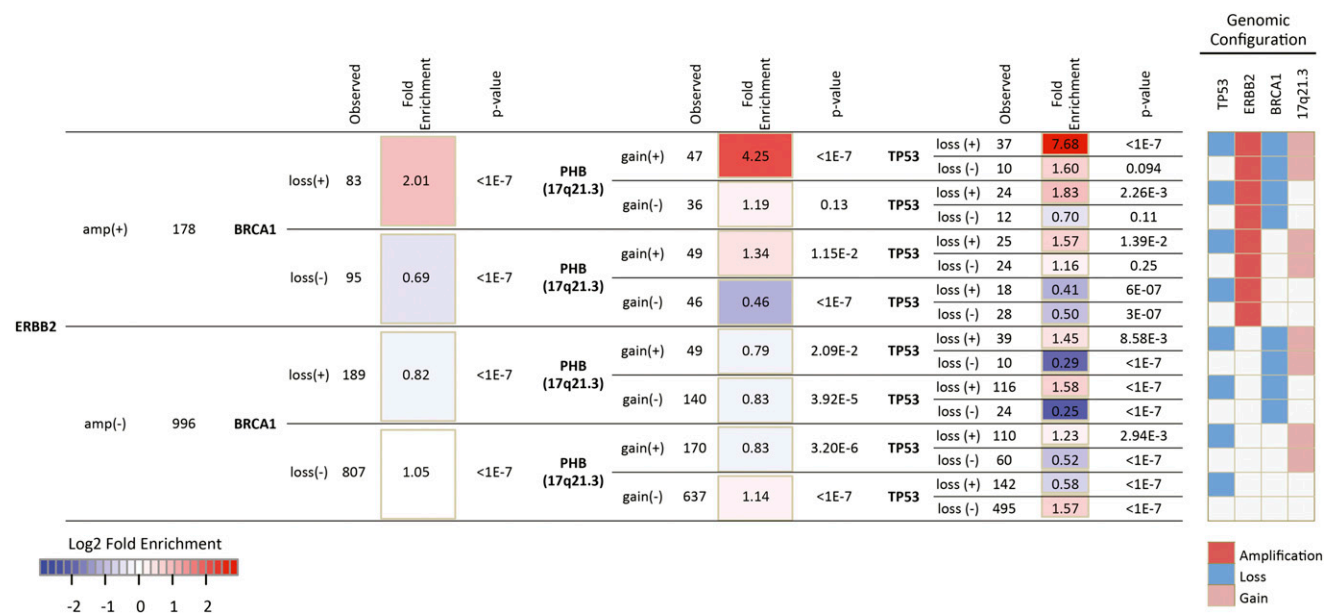To validate these observations, we systematically searched for recurrent structural variations across the eight sequenced breast cancers, seven of which show at least a minimal level of *ERBB2* amplification. We defined recurrent structural variations as events with >50% sequence overlap between the two breakpoint boundaries and whose breakpoints mapped within 3 Mbp of one another (Supplemental Methods). We found 45 loci with regionally recurrent rearrangements, corresponding to a total of 140 individual rearrangements (Supplemental Table 7). Seventeen recurrent events localize to Chr 17q, and at least six of them (comprising 38 individual structural variations and affecting six out of the eight samples analyzed) appear to contribute to the characteristic 17q genomic configuration described above (Supplemental Figs. 5, 7; Supplemental Table 7).

Having observed a possible consistent mechanism of *ERBB2* amplicon formation in this recurrent chromosomal rearrangement, which involves *BRCA1* loss and *17q21.3* coamplification, we addressed the significance of such a configuration by interrogating the copy number profile of a large number of breast cancer samples accessible in publicly available databases ($n = 1174$) (Supplemental Methods). We first classified all samples as *ERBB2*-amplified ($\log_2$ CN ratio > 1.0; $n = 178$) or nonamplified ($\log_2$ CN ratio < 1.0; $n = 996$). We then quantified the association between the *ERBB2* status and the copy number status of adjacent loci, *BRCA1* and *17q21.3*, by comparing the frequencies of different combinations of copy number patterns. As predicted, the frequencies of *BRCA1* loss and *17q21.3* gain (calculated using the *PHB* gene as a surrogate measure of the *17q21.3* amplicon) in *ERBB2*-amplified tumors were strikingly higher than those in *ERBB2* nonamplified samples (47% versus 19% for *BRCA1* loss, P-value = 4.134×10^{-14}; 54% versus 22% for *17q21.3* gain, P-value < 2.2×10^{-16}), indicating that *BRCA1* and *17q21.3* copy number changes are directly associated with the magnitude of *ERBB2* copy number (Supplemental Fig. 8A). Furthermore, we found that this specific configuration of *ERBB2*-amplification/*BRCA1*-loss/*17q21.3*-gain occurs 4.25 times more frequently than expected by chance (P-value < 1×10^{-7}) (Fig. 3). Focusing on samples that show an increase in the copy number of both the *ERBB2* and the *17q21.3* loci, we observed that in cases in which the *BRCA1* gene is also gained, the overall copy number profile is indicative of 17q trisomy (median copy number of 2.9 for *ERBB2, BRCA1*, and *17q21.3*) (Supplemental Fig. 8B). However, in cases where *BRCA1* is copy number neutral or, even more significantly, lost, *ERBB2* and *17q21.3* gains have much higher amplitudes (median copy number of 25.3 and 6.4 for *ERBB2* and *17q21.3*, respectively, in the presence of *BRCA1* loss) (Supplemental Fig. 8B). These data strongly suggest a physical dissociation of the *BRCA1* gene from the two flanking loci in the early events leading to their massive amplification.

We also noted that in seven out of the eight sequenced breast cancer samples, regions of 17p, including the *TP53* gene, appeared as one copy; and in three samples, *TP53* damaging point mutations affected the remaining allele (Supplemental Table 1). Remarkably, copy number frequencies across a larger data set of 1174 breast cancers (Supplemental Methods) showed that loss of *TP53* was more frequent in the *ERBB2*-amplified samples (58%) than in the *ERBB2*-negative samples (41%, P-value = 1.777×10^{-5}) (Supplemental Fig. 8A). When we included *TP53* copy number state in the analysis of Chr 17 configurations in breast cancer, we found that the loss of *TP53* was significantly associated with the oncogenic 17q copy number configuration as described above, increasing the overall fold enrichment for this specific genomic pattern (i.e., *TP53* loss, *ERBB2* amplification, *BRCA1* loss, and *17q21.3* gain) to 7.68-fold greater than expected by chance (Fig. 3). These data altogether suggest that there is a coordinated rearrangement involving the
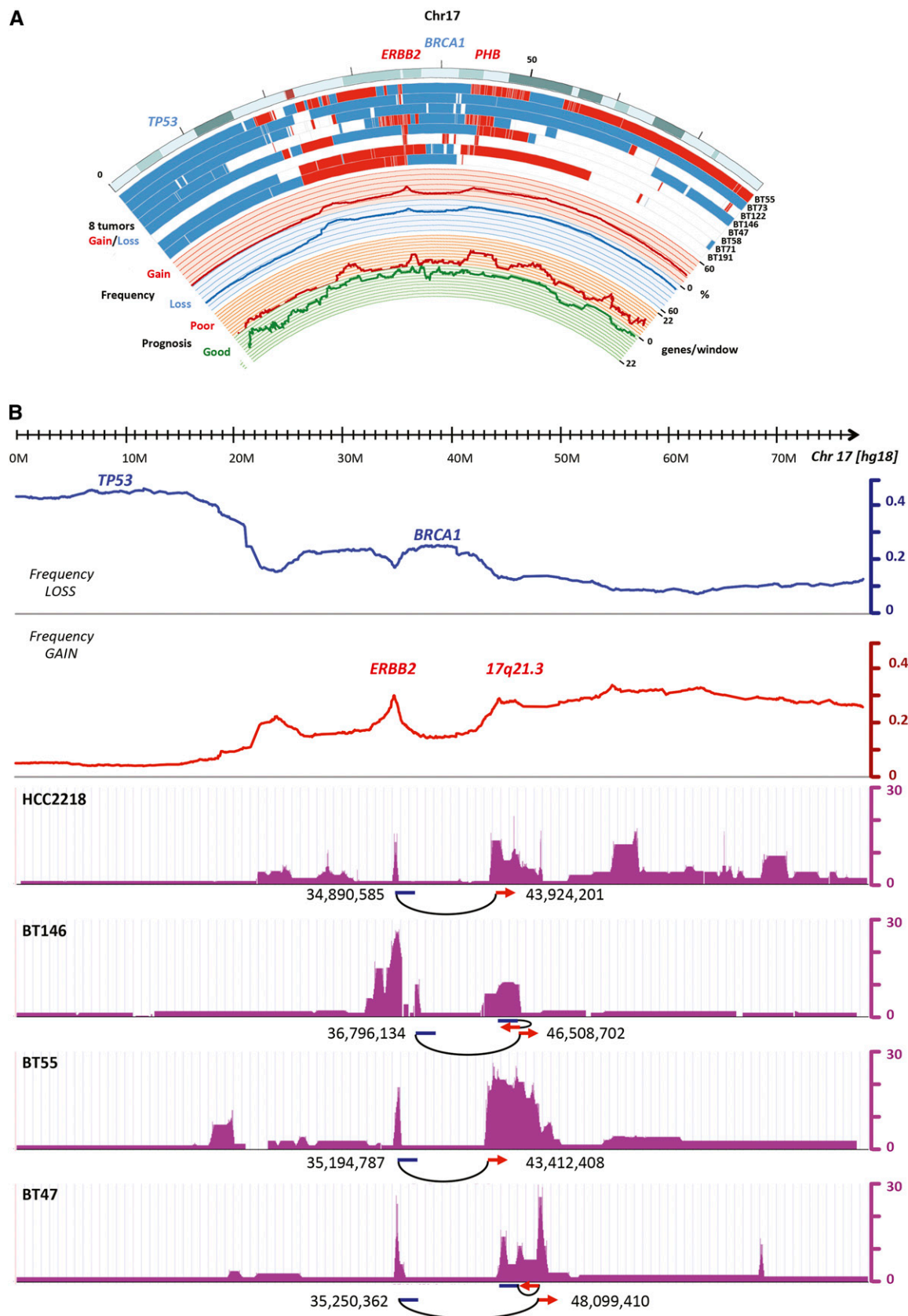
Figure 2. Oncogenic organization of Chr 17 in breast cancer. (*A*) Chromosomal map of Chr 17, showing copy number state of the eight breast tumors analyzed, the frequency of copy number gain and loss across a combined data set of 1174 breast cancers, and the number of breast cancer prognosis-associated genes per 40-gene windows. Clusters of poor and good prognosis associated genes are located in copy number gain and loss regions, respectively. (*B*) Linear map of Chr 17, showing the frequency of copy number gain and loss as in *A*, as well as the copy number profile of three breast tumors and a breast cancer cell line, HCC2218, with information on recurrent unpaired inversions (*below*).

Inaki et al.

	Observed	Fold Enrichment	p-value			Observed	Fold Enrichment	p-value			Observed	Fold Enrichment	p-value	Genomic Configuration TP53	ERBB2	BRCA1	17q21.3
ERBB2 amp(+) 178				BRCA1 loss(+) 83	2.01	<1E-7	PHB (17q21.3) gain(+)	47	4.25	<1E-7	TP53	loss (+) 37	7.68	<1E-7				
												loss (-) 10	1.60	0.094				
							gain(-)	36	1.19	0.13	TP53	loss (+) 24	1.83	2.26E-3				
												loss (-) 12	0.70	0.11				
				loss(-) 95	0.69	<1E-7	PHB (17q21.3) gain(+)	49	1.34	1.15E-2	TP53	loss (+) 25	1.57	1.39E-2				
												loss (-) 24	1.16	0.25				
							gain(-)	46	0.46	<1E-7	TP53	loss (+) 18	0.41	6E-07				
												loss (-) 28	0.50	3E-07				
amp(-) 996				BRCA1 loss(+) 189	0.82	<1E-7	PHB (17q21.3) gain(+)	49	0.79	2.09E-2	TP53	loss (+) 39	1.45	8.58E-3				
												loss (-) 10	0.29	<1E-7				
							gain(-)	140	0.83	3.92E-5	TP53	loss (+) 116	1.58	<1E-7				
												loss (-) 24	0.25	<1E-7				
				loss(-) 807	1.05	<1E-7	PHB (17q21.3) gain(+)	170	0.83	3.20E-6	TP53	loss (+) 110	1.23	2.94E-3				
												loss (-) 60	0.52	<1E-7				
							gain(-)	637	1.14	<1E-7	TP53	loss (+) 142	0.58	<1E-7				
												loss (-) 495	1.57	<1E-7				

Log2 Fold Enrichment: -2 -1 0 1 2

Amplification / Loss / Gain

Figure 3. Enrichment of *ERBB2*-amplification, *BRCA1*-loss, *PHB*-gain, and *TP53*-loss in breast cancer ($n = 1174$ breast cancers). Fold enrichments correspond to the ratio between observed and expected values, in which expected values were computed based on the individual frequencies of the four events examined (*ERBB2* amplification, *BRCA1* loss, *PHB* gain, and *TP53* loss). *P*-values were derived empirically from 10 million shuffled data sets. (amp) Amplification (log_2 copy number ratio > 1.0); (gain) copy number gain (log_2 copy number ratio > 0.3); (loss) copy number loss (log_2 copy number ratio < −0.3).

entire Chr 17, where significant segmental coamplification of *ERBB2* and genes on *17q21.3* occurs with the loss of the intervening segment containing *BRCA1*, in conjunction with copy number loss and disruptive mutations of *TP53*.

Mechanisms of the *ERBB2* amplicon formation on Chr 17q

The resolution of mate-pair sequencing in defining and quantifying structural variations provided a portal into the dynamics of amplicon formation that enabled us to model, in detail, the complex rearrangements on Chr 17q. In the breast tumor sample BT55, the structural variations with the highest cluster sizes ($N = 207$–331) are an ~500-kbp tandem duplication harboring *ERBB2* (TD207) and a recurrent unpaired inversion connecting the *ERBB2* and the *17q21.3* amplicons (UI331) (Fig. 4). The breakpoints of the two rearrangements coincide with a fivefold increase in the copy number of both the *ERBB2* and *17q21.3* loci. We speculate that the tandem duplication, TD207, which causes an increase in *ERBB2* copy number, occurred as the first event in the region and was followed by an unpaired inversion, UI331, which physically connects the *ERBB2* and *17q21.3* amplicons (Fig. 4). A subsequent deletion, Del51, caused the loss of a copy of the *BRCA1* gene. Similarly, in a detailed reconstruction of Chr 17q in BT146, three tandem duplications (TD359, TD377, and TD426), all of which include the *ERBB2* locus, were followed by two unpaired inversions (UI852 and UI626) connecting *ERBB2* with *17q21.3* and a deletion (Del285) affecting *BRCA1* (Supplemental Figs. 9A,B).

An analogous sequence of events characterizes *ERBB2* amplification in BT47 (Supplemental Fig. 9C), which shares a remarkably similar copy number profile on Chr 17 with BT55 and BT146. Furthermore, three samples with lower amplitude *ERBB2* gains (BT58, BT71, and BT73) also show evidence of at least some of these critical recurrent 17q rearrangements (Supplemental Fig. 5; Supplemental Table 7).

In both BT146 and BT73, we observed a second mechanism leading to *ERBB2* amplicon formation. In BT146, in addition to the unpaired inversion generating a conjoint *ERBB2-17q21.3* amplicon, the *ERBB2* locus translocated to chromosome 16 and was independently amplified (breakpoints: Inter591 and Inter826) (Supplemental Fig. 9A). A reconstruction of the breakpoints suggests the possibility that this alternative path of genomic amplification may have caused the *ERBB2* gene to be embedded in an episome structure (Supplemental Fig. 9B). In BT73, the *ERBB2* amplicon appears to result from a series of interchromosomal translocations between chromosomes 17 and 5, which follow an initial tandem duplication event (Supplemental Fig. 9D). In this setting, the *ERBB2* gene physically disconnects from the downstream *BRCA1* locus before being amplified in its new chromosomal location.

Taken together, this detailed structural variation mapping revealed the presence of conserved mechanistic paths involved in *ERBB2* amplification. Although different rearrangements may eventually be responsible for the final generation of the massive copy number increase at the *ERBB2* locus observed in ERBB2-positive breast cancers, we were able to delineate a more general sequence of events whereby an initial tandem duplication of the *ERBB2* locus is followed by specific rearrangements that physically separate *ERBB2* from *BRCA1* favoring the subsequent massive coamplification of the distal *17q21.3* region, thereby generating a putatively optimal genomic configuration.

Functional consequences of 17q rearrangements in breast cancer

Our results suggest that *ERBB2* overexpression, *BRCA1* loss, and overexpression of genes on *17q21.3* may represent an advantageous expression configuration for breast cancer progression. We hypothesized that these events would be at least additive in advancing the cancer phenotype. To examine the combined biological effects of *ERBB2* overexpression and *BRCA1* knockdown, we utilized non-transformed human mammary epithelial cells (HMECs). Although
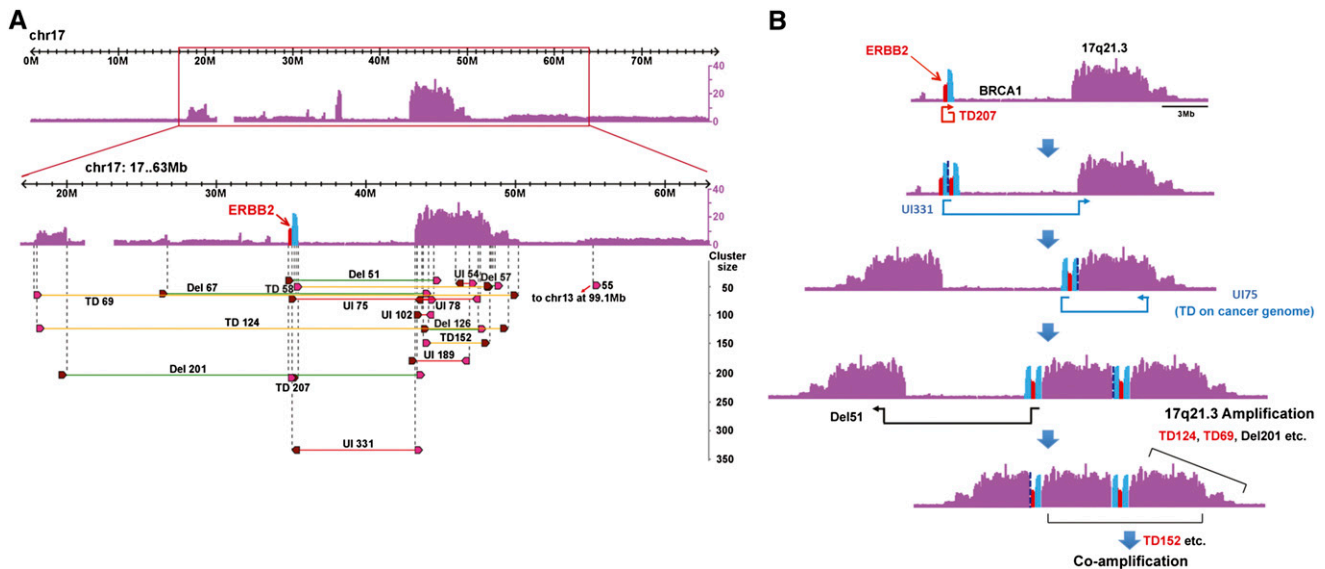
Figure 4. Reconstruction of Chr 17q rearrangements in BT55. (*A*) Copy number and structural variations mapping to Chr 17 (17–63 Mb) in the BT55 sample. For each structural variation, the brown and pink arrowheads correspond to the *left* and *right* mate-pair read clusters, respectively. SV breakpoints map to the end of the brown arrowhead and to the start of the pink arrowhead (as detailed in Hillmer et al. 2011) and are indicated by dashed vertical lines. (TD) Tandem duplication (yellow); (UI) unpaired inversion (red); (Del) deletion (green). The number associated with each structural variation corresponds to its cluster size. (*B*) A schematic description of the plausible structural events that contributed to the final shape of Chr 17 in the BT55 cancer genome. An initial tandem duplication (TD207) involving the *ERBB2* locus was followed by unpaired inversions UI331 and UI75, which is in effect a TD on the cancer genome, resulting in the joining of *ERBB2* and the *17q21.3* locus. Subsequent events caused the loss of the *BRCA1* locus and the massive coamplification of the *ERBB2* and *17q21.3* loci.

both *ERBB2* overexpression and *BRCA1* knockdown increased cell proliferation in HMECs (*P*-value = 1.3×10^{-5} and *P*-value = 3.2×10^{-6}, respectively), the combination of *ERBB2* overexpression with *BRCA1* knockdown caused an increase in cell proliferation greater than either perturbation individually (*P*-value = 3.7×10^{-2}) (Fig. 5A; Supplemental Fig. 10A). This supports the hypothesis that biological selection would favor ERBB2-positive tumors that concurrently lose the *BRCA1* tumor suppressor gene.

Next, we aimed to identify driver oncogenes mapping to Chr *17q21.3*. Because this locus appears to be a broad amplicon in breast cancer, it is likely to harbor a cluster of oncogenes (Supplemental Fig. 11). The *17q21.3* amplified regions in BT55 and BT146 overlap and demarcate a region frequently amplified in breast cancer (Supplemental Fig. 11). We selected candidate oncogenes on *17q21.3* and assessed the effects of their targeted knockdown by siRNAs on cell proliferation in 10 breast cancer cell lines, all of which show copy number gain corresponding to the *17q21.3* region, with the exception of SKBR3. All the tested cell lines are *ERBB2*-amplified, except MCF7 (Supplemental Table 8). Among the 76 genes mapping to the *17q21.3* amplicon, 39 were expressed (>1.5 FPKM) in both BT55 and BT146. We further narrowed down this list to 21 genes, based on their association with breast cancer patients' poor prognosis (distant metastasis free survival and relapse free survival) (Supplemental Methods), high expression in cancer, and coexpression with the *ERBB2* gene (Supplemental Table 8; Methods). We first tested *ERBB2* dependecies and found that *ERBB2* knockdown by siRNA induced a reduction in proliferation in all *ERBB2*-amplified cell lines, except MB361 (Supplemental Fig. 12; Supplemental Table 9). We then applied siRNAs against the 21 *17q21.3* gene candidates in all 10 cell lines and measured their effect on cell proliferation. We found that a maximal effect was achieved when *ERBB2* was knocked down, and that the silencing of 19 of the 21 candidate oncogenes also resulted in reduced

growth (Fig. 5B; Supplemental Table 9). Because siRNA effects might be more effective in highly amplified cell lines, we developed a "copy number dependency" measure of the siRNA effects on cell proliferation in each cell line: We multiplied the number of copies of each tested gene by its corresponding siRNA-induced reduction in cell proliferation (Supplemental Fig. 13A). As expected, the *ERBB2* gene appeared to cause the strongest progrowth effect (score = 57). The same 19/21 oncogene candidates on *17q21.3* also showed growth associated potential, although to a lesser degree when compared with *ERBB2* (*HOXB7* gene silencing generated the highest score of 25.9) (Supplemental Fig. 13A), supporting the hypothesis that a cluster of weak oncogenes maps to the broad peak region of the *17q21.3* amplicon.

Synergistic effects of *17q21.3* oncogene silencing and *ERBB2* inhibition

Next, we tested whether there may be interactive effects on cell proliferation between *ERBB2* and the *17q21.3* oncogenes, whose genomic regions are physically connected and coamplified in some breast cancers. To assess this question, we used Lapatinib, an inhibitor of the ERBB2 and EGFR kinase activities that is approved for the treatment of patients with *ERBB2*-overexpressing breast cancers (Arteaga et al. 2011). We treated each cell line with Lapatinib at the appropriate IC50 concentration while silencing each target gene on *17q21.3*. Each one of the 19 siRNAs that caused a decrease in cell proliferation when used individually caused a further reduction in cell growth when applied in combination with Lapatinib, suggesting that their targeted genes contribute at minimum a cellular effect that is additive to the effect induced by *ERBB2* (Supplemental Table 9). Following an adjustment for copy number effects across the cell lines (copy number dependency) (Supplemental Fig. 13B), we computed the difference in the
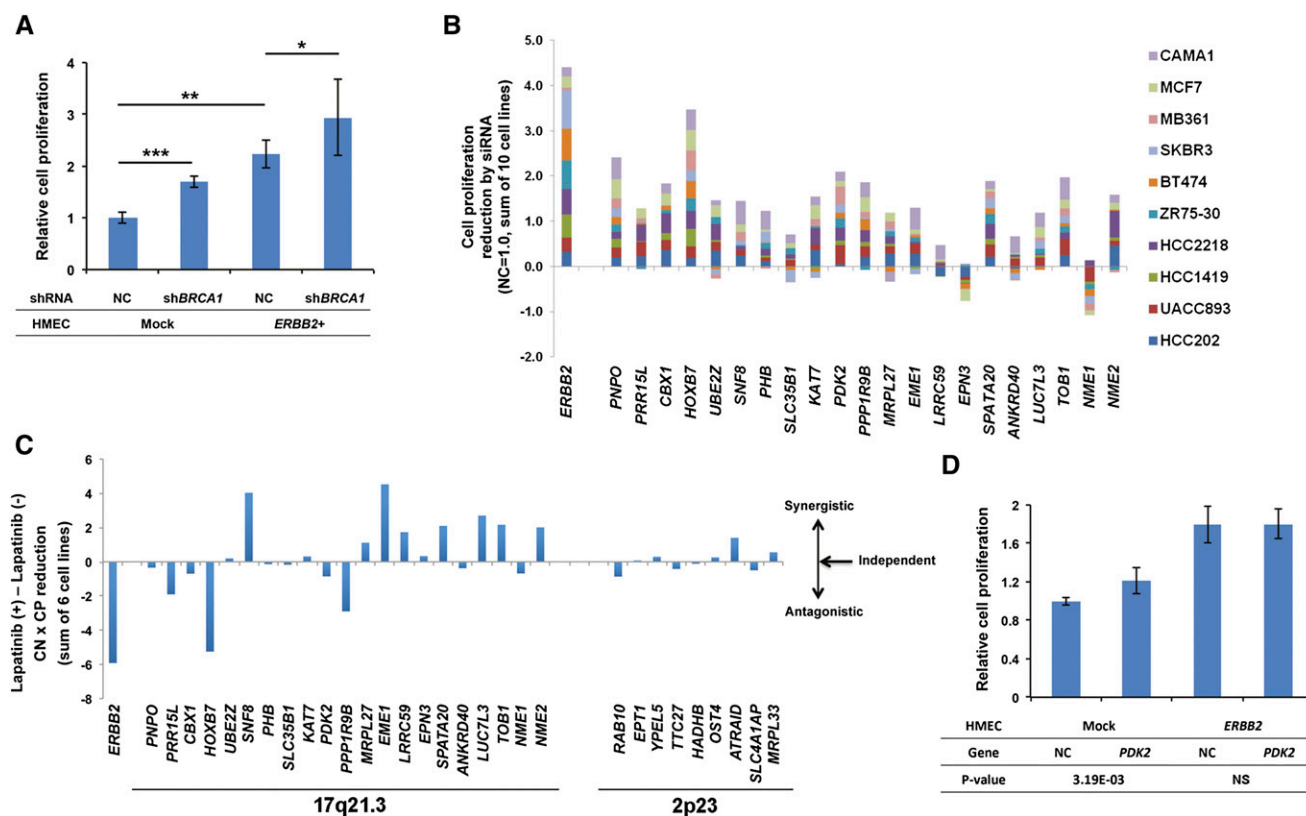
Figure 5. Validation of the cancerous function of genes on Chr 17q. (*A*) Cell proliferation assay using mock and *ERBB2*-overexpressing human mammary epithelial cells (HMECs), which were treated with shRNA against *BRCA1* or a nontargeting shRNA. Knockdown of *BRCA1* increases cell proliferation in both conditions. The average of two independent experiments ± standard deviation is shown. *P*-values (Student's *t*-test): (*) *P*-value = 3.7×10^{-2}; (**) *P*-value = 1.3×10^{-5}; (***) *P*-value = 3.2×10^{-6}. (*B*) Cell proliferation assay for 10 breast cancer cell lines transfected with siRNAs against the *17q21.3* genes and *ERBB2*. Each bar represents the sum of the reductions in cell proliferation caused by each individual siRNA across all cell lines examined. Gene symbols are ordered based on chromosomal location (lower coordinates on *left*). (*C*) Combined effect of Lapatinib treatment and cell transfection with siRNAs targeting *ERBB2*, *17q21.3* genes, or *2p23* control genes in six breast cancer cell lines. The observed reduction in cell proliferation (CP) is multiplied by the copy number (CN) in each cell line (copy number dependency) (Supplemental Fig. 13A,B). Bars correspond to the difference in cell proliferation reduction between cells cultured in the presence and in the absence of Lapatinib. Positive values indicate putative synergistic effects between the Lapatinib and the siRNA treatments, whereas negative values indicate antagonistic effects. Values close to zero represent cases in which the two treatments are independent. A selection of genes mapping to the *2p23* control locus was made to match the range of gene expression levels observed for the *17q21.3* candidate genes across the six experimental cell lines used in this experiment. (*D*) Cell proliferation assay using mock and *ERBB2*-overexpressing HMECs, which were transfected with a *PDK2*-overexpressing vector or a negative control vector (NC), showing enhanced cell proliferation in mock HMECs following *PDK2* overexpression (*P*-value = 3.19×10^{-3}, Student's *t*-test). The average of two independent experiments ± standard deviation is shown. qPCR data showing changes in expression levels following gene silencing/overexpression are shown in Supplemental Figures 10 and 12.

reduction of cell proliferation for cells cultured in the presence or in the absence of Lapatinib to measure the potential synergy or antagonism between the inhibition of *ERBB2* and the selected *17q21.3* genes (Fig. 5C). Seven of the 19 putative oncogenes showed potential for synergism with *ERBB2*, whereas three showed plausible partial antagonism with *ERBB2* in at least two cell lines (Supplemental Table 10). When genes on chromosome *2p23*, whose frequency of copy number gain does not correlate with *ERBB2* amplification in breast cancer (10.9% in *ERBB2*-amplified and 8.2% in *ERBB2*-nonamplified tumors based on the TCGA data set), were tested for their synergy with *ERBB2*, we found significantly fewer interactive genes compared to those on *17q21.3* (*P* = 1.91×10^{-7}) (Supplemental Fig. 13C–E). Thus, genes in the coamplified region of *17q21.3* appear to have discernible pharmacological interactions with ERBB2 inhibition over randomly selected genes.

One of the partially antagonistic genes, *PDK2*, was further tested. The overexpression of *PDK2* showed enhanced growth in

the native HMECs but not in *ERBB2*-overexpressing HMECs (Fig. 5D; Supplemental Fig. 10B). Thus, even though *PDK2* is a growth inductive gene, its effects are overshadowed by *ERBB2* action, suggesting that the growth promoting pathways elicited by *PDK2* and *ERBB2* are redundant. Similarly, although *HOXB7* has a strong isolated effect on growth individually, it appears to be antagonistic to *ERBB2*-associated growth. Therefore, *HOXB7* may have no positive cancer evolutionary advantages upon coamplification with *ERBB2*, even though it is a moderate progrowth gene when acting alone. This appears to be correlated with a lower frequency of copy number gain of *HOXB7* than other members of the *17q21.3* amplicon (Supplemental Fig. 11).

To generalize the portrayal of the *17q21.3* amplicon as an oncogene cluster that contributes additive oncogenic potential to a structurally connected locus, we asked if other broad amplicons could be indicative of biologically relevant gene coamplifications. By using the BT73 cancer genome as a model, we focused on the generally broad amplicon on the telomeric end of the long arm of

chromosome 8, which harbors the well-known *MYC* oncogene (Supplemental Results; Supplemental Fig. 14). Searching for putative oncogenes mapping to this region, we selected the serine/threonine protein kinase-coding *RIPK2* gene based on its association with poor breast cancer patients' outcome and on its statistically significant coamplification with *MYC* (Supplemental Results). SiRNA-based gene silencing confirmed that *RIPK2* functions as a strong breast cancer oncogene in *MYC*-amplified cancer cell lines (Supplemental Fig. 14).

Discussion

The maintenance of a full cancer phenotype is dependent on the complement of genetic and epigenetic changes affecting genes responsible for many cellular functions (Hanahan and Weinberg 2011). From whole-genome sequencing data, we learned that, although there are some common oncogenic drivers, each cancer shows combinations of genomic aberrations that often are unique for that individual tumor. These structural aberrations and their consequent expression profiles engage a range of biochemical and cellular functions that, as a whole, cannot be ascribed to a single oncogenic pathway. Using deep genomic analyses of breast cancers to reveal the detailed genetic landscape of each tumor, we sought to find coherence in how the many different somatic mutations may be related.

By examining the detailed structural genomic alterations associated with common amplicons, we found evidence for a sequence of structural variations affecting recurrent regions of genomic amplification. In five of the seven analyzed breast cancers with evidence of copy number gain at the *ERBB2* locus, large tandem duplications encompassing the *ERBB2* gene appeared to initiate the amplification of the region. In most cases, *ERBB2* amplification was associated with copy number loss of *BRCA1* through different genomic mechanisms. In three samples, unpaired inversions joined the *ERBB2* locus with the region on *17q21.3* that would be subsequently amplified; and in all three cases, we found structural evidence for a deletion of the intervening chromosomal segment containing the *BRCA1* gene (Supplemental Fig. 5). In the remaining samples, *ERBB2* is amplified through more complex rearrangements that included translocation to other chromosomes and multiple unpaired inversions.

To extend this analysis to a larger data set, we analyzed a collection of 1174 breast cancers and noted that despite the gene order on 17q being *ERBB2*, *BRCA1*, followed by *17q21.3*, among the tumors with concomitant amplification of *ERBB2* and of the *17q21.3* region, 81% (52/64 samples) showed either a decrease or no increase in the copy number of the intervening *BRCA1* locus. This implies that in the vast majority of cases, *BRCA1* is lost before the *ERBB2-17q21.3* linked region is amplified. In a different study, 60% of 200 *ERBB2*-amplified tumors show amplification of *17q21.3*, whereas 40% show loss of the *BRCA1* locus (Staaf et al. 2010). More recently, 2,000 breast tumors were divided into 10 subgroups based on their genomic and transcriptomic architecture (Curtis et al. 2012). The *ERBB2* subgroup (IntClust 5), consisting of 94 *ERBB2*-amplified tumors, showed coamplification of *17q21.3* and loss of *BRCA1* in 36% and 54% of cases, respectively (Curtis et al. 2012). Thus amplification of *17q21.3* and loss of *BRCA1* represent a frequent configuration among *ERBB2*-amplified tumors. Our detailed observations through deep sequencing for structural variations in the eight breast cancers suggest that rearrangements, such as unpaired inversions connecting the *ERBB2* and *17q21.3* loci or a direct deletion of *BRCA1*, followed by coamplification of the newly linked *ERBB2* and *17q21.3* loci, are among the key events leading such an optimal genomic rearrangement.

If one views the genetic landscape of a tumor as an indication of evolutionary optimization for any specific cancer, then the identification of frequent conjoint genetic events can be a means to discover additive or synergistic actions of oncogenic elements. We validated this hypothesis in vitro by showing that overexpression of *ERBB2* and enforced knockdown of *BRCA1* expression augmented the growth of HMECs at a level greater than modulating each gene alone. Recently, *ERBB2* amplification and *BRCA1* loss along with *MYC* amplification were assessed to be early events before establishment of the "most-recent common ancestor" clone, which undergoes subsequent expansion (Nik-Zainal et al. 2012). This is further evidence that the gain of *ERBB2* and the concomitant loss of *BRCA1* are early events in breast tumor evolution and may be advantageous to breast cancer growth. Furthermore, the genomic configuration of tumors harboring the *ERBB2* amplicon also suggested that there is an advantage for *ERBB2* to be coamplified with genes residing on *17q21.3*. We tested the biological role of 21 candidates on *17q21.3* using in vitro cell proliferation assays and found that 19 of the 21 genes exerted positive growth effects that were at least additive, and in seven cases possibly synergistic with those elicited by *ERBB2* activity. Finally, the observation that *TP53* copy number loss occurs in 79% of the cases in which *BRCA1* is also lost (Fig. 3) strongly suggests some functional interaction. Indeed, it has been noted that *TP53* and *BRCA1* interact together to optimally induce the expression of *SFN* (14-3-3ς), a major G_2/M checkpoint control gene that blocks progression of cells with DNA damage through mitosis (Aprelikova et al. 2001). The loss of *SFN* regulation would lead to abnormal checkpoint control, a mechanism for genomic instability.

Thus, the chromosomal "origami" generated by the conjoint rearrangements on 17q (*ERBB2* tandem duplication followed by unpaired inversions or deletions disrupting *BRCA1* and linking *ERBB2* with the downstream *17q21.3* amplicon) appears to be a topologically parsimonious strategy to generate a genomic state favoring tumor growth (Fig. 6). In addition, through only a few genomic rearrangements, the simultaneous activation by copy number gain and the attenuation by copy number loss of genes affecting cancer cell growth suggest that rearrangements significantly restructure the cancer genome to engage the maximal number of oncogenic elements with each rearrangement, so that even small copy number changes manifest to global changes in levels of gene expression.

It has been previously recognized that some cancer-associated genes are colocalized in the genome. A classic example of this is the overlapping localization of the p16[INK4a] and p14[ARF] tumor suppressors encoded within the *CDKN2A* locus on *9p21*. Deletions at this locus routinely inactivate both cell-cycle regulators (Bates et al. 1998; Stott et al. 1998). Recent evidence also suggests that oncogenes and tumor suppressor genes (TSGs) are clustered, including several TSGs on *8p21-23* in hepatocellular carcinoma, three oncogenes (*NKX2-1*, *NKX2-8*, and *PAX9*) on *14q13* in lung cancer, two oncogenes (*CCND1* and *FGF19*) on *11q13*, and two oncogenes (*BIRC2* and *YAP1*) on *11q22* in liver cancer (Zender et al. 2006; Kendall et al. 2007; Sawey et al. 2011; Solimini et al. 2012; Xue et al. 2012). Similar to the results presented here, phenotypes induced by the manipulation of individual genes are relatively weak, whereas the concerted deregulation of entire cancer gene clusters result in more significant effects. We found evidence that poor or good breast cancer prognosis associated genes, which include a significant number of oncogenes and TSGs, as well as markers of cancerous elements (Zhang et al. 2009; Soon et al. 2011; Lee et al. 2012), cluster in regions of aberrant copy number and gene expression in breast cancer (Fig. 2; Supplemental Figs. 6A, 14A), further supporting the hypothesis that clusters of cancer-relevant
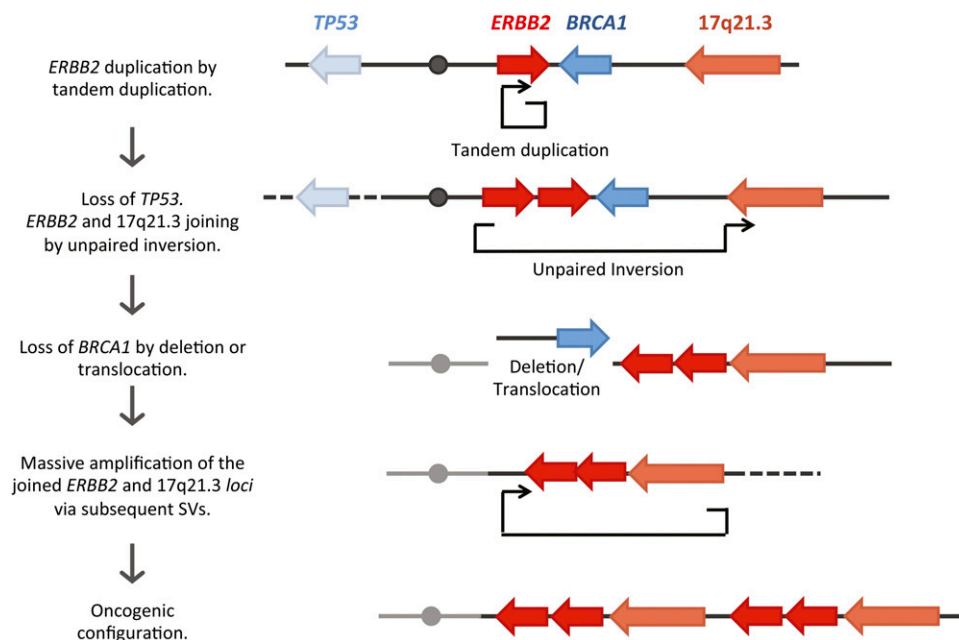
Figure 6. Schematic overview of sequential rearrangements leading to the oncogenic configuration on 17q.

genes are distributed over the entire genome and are activated or inactivated by copy number imbalances that contribute to the oncogenic state.

Our results also suggest that the genomic configuration of a breast cancer may identify conjointly expressed targets for combination therapy. One example is the coamplification of *ERBB2* and genes on the downstream *17q21.3* amplicon. The best possible targets would be those that exhibit additive or synergistic effects with Lapatinib treatment. Intriguing new candidates include *SNF8*, a component of the endosomal sorting complex (ESCRT-II), which is required for multivesicular body formation and involved in endosomal sorting (Malerød et al. 2007). The ESCRT-II complex is involved in the degradation of both endocytosed EGF and EGFR. *EME1*, another gene showing greater than additive effect with *ERBB2*, encodes a protein that is part of an endonuclease complex and may be involved in DNA damage repair and maintaining genomic stability (Sengerová et al. 2011). Other druggable candidate targets on *17q21.3* include *UBE2Z*, an ubiquitin ligase, and *KAT7* (previously known as *MYST2*), a histone acetyltransferase, both of which showed an additive effect.

In summary, our work suggests that putative oncogenes aggregate and separate from tumor suppressors in the human genome. This clustering of oncogenic elements augments the biological effect of segmental amplification and deletions commonly found in cancers as evidenced by their combined effects on tumor growth in vitro. Therefore structural variations in breast cancer act as systems organizers for gene cassettes that together have a significant role in maintaining the cancer phenotype. When viewed in evolutionary terms, the multitude of structural rearrangements and mutations are not necessarily randomly derived. Instead, it is a process of selection for the gain and loss of adjacent clusters of gene cassettes that, when combined, optimize tumor survivability and growth within a single individual. Given that the combinatorics are immense, a systems analysis focusing on multigene as opposed to single gene effects will be the best approach in deciphering this level of complexity.

Methods

Breast tissue samples

After obtaining a patient's written informed consent, breast tumor tissues and peripheral blood samples were collected during surgery at the Institute of Breast Diseases (FUCAM) in Mexico City. The protocol was reviewed and approved by the Ethics and Research committees of the National Institute of Genomic Medicine and the Institute of Breast Diseases in Mexico City (CE2009/11). After macroscopic review by a trained pathologist, tumor tissues were divided in half. One half was snap-frozen in liquid nitrogen and kept at −80°C until further processing. The second half was fixed in buffered formalin and embedded in paraffin. Tissue sections were generated from this half and following hematoxilin-eosin staining, they were independently reviewed by two pathologists. Only tumors with >80% tumor cells were selected, and DNA and RNA samples were isolated from their frozen tissue counterparts as well as the corresponding peripheral blood lymphocytes, using the AllPrep DNA/RNA mini kit (Qiagen).

Mate-pair DNA sequencing

Normal-tumor DNA paired samples from four breast cancer patients were processed and analyzed following the DNA paired-end tag (DNA-PET) sequencing technology, as previously described (BT55, BT73, BT122, and BT146) (Hillmer et al. 2011; Supplemental Table S1; Supplemental Methods), whereas another four pairs were sequenced using the Illumina HiSeq 2500 platform, following standard Nextera mate-pair library preparation (BT47, BT58, BT71, and BT191) (Supplemental Table S1; Supplemental Methods).

Mate-pair sequencing data for the 16 libraries were analyzed as described in Hillmer et al. (2011). Briefly, mate-pair reads were separated into concordant and discordant ones, based on whether they mapped back to the reference genome in the expected fashion or not. Discordant mate-pair reads were then used to predict structural

variations, based on their relative orientation and mapping coordinates. Overlapping discordant mate-pair reads were clustered together as they represent the same rearrangement. The number of mate-pair reads clustering together and pointing to the same structural variation was represented by the cluster size.

For each pair of normal and tumor libraries, discordant mate-pair read clusters were compared to search for significant overlaps (i.e., any overlap within a distance equivalent to the maximum library size from each end of any given cluster in both directions). Only clusters specific for the tumor library were considered to be indicative of somatic rearrangements and used as the input for structural variation classification, as described in Hillmer et al. (2011). Circos plots summarizing all the somatic structural variations in each cancer genome were generated using the Circos software (Krzywinski et al. 2009).

Selection of *17q21.3* oncogene candidates

We first identified the 76 genes mapping to the *17q21.3* amplified genomic region in BT55 and/or BT146. Of these, 39 genes were selected based on a level of gene expressions (RNA-seq) higher than 1.5 FPKM in both BT55 and BT146. We then scored the following parameters for the selection of the final list of 21 candidate oncogenes: (1) significant gene overexpression in ductal breast carcinoma compared with normal tissue (P-value < 0.05, fold change > 1.5, based on Oncomine data sets [http://www.oncomine.org/]); (2) significant negative correlation of gene expression with breast cancer patients' relapse free survival (P-value < 0.05, n = 2422, based on the KM plotter database [http://kmplot.com/analysis/]); (3) mapping to significant amplicon peaks, as defined based on the Tumorscape data portal (http://www.broadinstitute.org/tumorscape/, breast cancer data set, n = 240), frequent amplicon regions in *ERBB2*-amplified breast cancers (Staaf et al. 2010) (n = 200), and gene expression amplicons defined by the local singular value decomposition (Zhang et al. 2009); (4) significant positive correlation with *ERBB2* gene expression (P-value < 0.05), based on six independent microarray data sets (van de Vijver et al. 2002; Miller et al. 2005; Pawitan et al. 2005; Wang et al. 2005; Ivshina et al. 2006; Loi et al. 2007); and (5) negative gene expression correlation with distant metastasis free survival (DMFS) based on the union of the six expression data sets described in parameter (4).

Gene expression estimate values for the six breast cancer microarray studies indicated in (4) above were obtained by applying the MAS5.0 global mean method to Affymetrix CEL files downloaded from the Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo/). Probe set-based signal intensities were log-transformed and scaled by adjusting the mean intensity to a target signal value of log(500). To remove cohort-dependent array biases, arrays within each cohort were median-centered prior to combining them into one data set. Meta-analysis on the association of candidate gene expression with DMFS was performed using Partek Genomics Suite. The *ERBB2* status was determined by splitting samples by the median expression of *ERBB2* in each individual cohort.

Cell lines

Breast cancer cell lines were obtained from the ATCC. Cells were cultured in RPMI or DMEM supplemented with 10% FBS and maintained at 37°C and 5% CO_2. Details of the growth medium and Lapatinib concentration used for the in vitro experiments are as follows: HCC202 (RPMI, 130nM); UACC893 (DMEM, 300nM); HCC1419 (RPMI, 70nM); HCC2218 (RPMI, 40nM); ZR75-30 (RPMI, 35nM); BT474 (RPMI, 70nM); SKBR3 (DMEM, 40nM); MB361 (DMEM); MCF7 (DMEM); and CAMA1 (DMEM). Immortalized

human mammary epithelial cells (HMECs), stably transfected with empty vector or *ERBB2*, were established in the Yu Qiang laboratory (GIS, Singapore). HMECs were cultured in DMEM/F12 (Invitrogen) with 5% horse serum, 50 ng/mL EGF, 500 ng/mL hydrocortisone, 10 µg/mL insulin, 100 ng/mL cholera toxin, and 100 units/mL penicillin, maintained at 37°C with 5% CO_2.

Copy number data for the cell lines were obtained from a previous study (Kao et al. 2009) and smoothened using the *DNAcopy* package in R (Venkatraman and Olshen 2007), except CAMA1, whose copy number profile was obtained from the Tumorscape database.

Gene knockdown and cell proliferation assay

For the siRNA experiments using breast cancer cell lines, cells were plated at day 0 in 12-well plates, in triplicate, and at an appropriate cell density to assure that they would be in the log phase of cell growth at the time of the assay ($2–6 \times 10^4$ cells/mL). On day 1, cells were transfected with 20 or 40 pmol of ON-TARGET plus SMARTpool siRNA (Dharmacon) using Lipofectamine 2000 and Opti-MEM (Invitrogen) according to the manufacturer's instructions. In experiments testing the effect of Lapatinib, the drug (purchased from LC Laboratories) was added to the culture at the cell line-specific IC50 concentration on day 2. Cell proliferation was assessed on day 6 using Cell Counting Kit-8 (Dojindo). Cells transfected with a nontargeting siRNA were used as a negative control (NC).

HMECs were transfected with plasmid-based shRNAs (*BRCA1*: ggagatcaagaattgttacaa; negative control: ggaatctcattcgatgcatac; Qiagen, 336312 KH00322H) using Lipofectamine 2000. Forty-eight hours after transfection, the cells were replated and stably selected with 250 µg/mL of hygromycin (Merck Calbiochem). For the proliferation assay, cells were plated in quadruplicate in 96-well plates, at a density of 2500 cells/well. Cell proliferation was measured after 48 h using WST-1 (Roche).

PDK2 stable overexpression

pEB Multi-Neo vector (Wako), in which genes encoding *PDK2* with a C terminus HA tag and IRES-DsRed-express2 (Clontech) were inserted, was transfected into HMECs using Lipofectamine 2000. The pool of transfected cells was passaged once before a double sorting of DsRed-expressing cells via fluorescence activated cell sorting. The sorted cells were passaged once before proceeding to the WST-1 proliferation assay. A pEB Multi-Neo vector with IRES-DsRed-express2 was used as the negative control vector.

Additional methods

Additional methods for RNA-seq, DNA-PET, and Nextera mate-pair library preparation and sequencing as well as source details for the publicly available data sets utilized in this study are provided in the Supplemental Methods.

Data access

Raw sequencing and processed RNA-seq data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE57914. Whole-genome sequencing data have been submitted to the NCBI Sequence Read Archive (SRA; http://www.ncbi.nlm.nih.gov/sra) under accession numbers PRJNA253369 and PRJNA251354.

Acknowledgments

References

Aprelikova O, Pace AJ, Fang B, Koller BH, Liu ET. 2001. BRCA1 is a selective co-activator of 14-3-3ς gene transcription in mouse embryonic stem cells. *J Biol Chem* **276:** 25647–25650.

Arteaga CL, Sliwkowski MX, Osborne CK, Perez EA, Puglisi F, Gianni L. 2011. Treatment of HER2-positive breast cancer: current status and future perspectives. *Nat Rev Clin Oncol* **9:** 16–32.

Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, Lawrence MS, Sivachenko AY, Sougnez C, Zou L, et al. 2012. Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* **486:** 405–409.

Bates S, Phillips AC, Clark PA, Stott F, Peters G, Ludwig RL, Vousden KH. 1998. p14^ARF links the tumour suppressors RB and p53. *Nature* **395:** 124–125.

Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M, et al. 2010. The landscape of somatic copy-number alteration across human cancers. *Nature* **463:** 899–905.

The Cancer Genome Atlas Network. 2012. Comprehensive molecular portraits of human breast tumours. *Nature* **490:** 61–70.

Chang SL, Lai HY, Tung SY, Leu JY. 2013. Dynamic large-scale chromosomal rearrangements fuel rapid adaptation in yeast populations. *PLoS Genet* **9:** e1003232.

Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo WL, Lapuk A, Neve RM, Qian Z, Ryder T, et al. 2006. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell* **10:** 529–541.

Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, et al. 2012. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486:** 346–352.

Fares MA, Keane OM, Toft C, Carretero-Paulet L, Jones GW. 2013. The roles of whole-genome and small-scale duplications in the functional specialization of *Saccharomyces cerevisiae* genes. *PLoS Genet* **9:** e1003176.

Hanahan D, Weinberg RA. 2011. Hallmarks of cancer: the next generation. *Cell* **144:** 646–674.

Hicks J, Krasnitz A, Lakshmi B, Navin NE, Riggs M, Leibu E, Esposito D, Alexander J, Troge J, Grubor V, et al. 2006. Novel patterns of genome rearrangement and their association with survival in breast cancer. *Genome Res* **16:** 1465–1479.

Hillmer AM, Yao F, Inaki K, Lee WH, Ariyaratne PN, Teo AS, Woo XY, Zhang Z, Zhao H, Ukil L, et al. 2011. Comprehensive long-span paired-end-tag mapping reveals characteristic patterns of structural variations in epithelial cancer genomes. *Genome Res* **21:** 665–675.

Inaki K, Liu ET. 2012. Structural mutations in cancer: mechanistic and functional insights. *Trends Genet* **28:** 550–559.

Inaki K, Hillmer AM, Ukil L, Yao F, Woo XY, Vardy LA, Zawack KF, Lee CW, Ariyaratne PN, Chan YS, et al. 2011. Transcriptional consequences of genomic structural aberrations in breast cancer. *Genome Res* **21:** 676–687.

Ivshina AV, George J, Senko O, Mow B, Putti TC, Smeds J, Lindahl T, Pawitan Y, Hall P, Nordgren H, et al. 2006. Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer Res* **66:** 10292–10301.

Jönsson G, Staaf J, Vallon-Christersson J, Ringnér M, Holm K, Hegardt C, Gunnarsson H, Fagerholm R, Strand C, Agnarsson BA, et al. 2010. Genomic subtypes of breast cancer identified by array-comparative genomic hybridization display distinct molecular and clinical characteristics. *Breast Cancer Res* **12:** R42.

Kao J, Salari K, Bocanegra M, Choi YL, Girard L, Gandhi J, Kwei KA, Hernandez-Boussard T, Wang P, Gazdar AF, et al. 2009. Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery. *PLoS ONE* **4:** e6146.

Kendall J, Liu Q, Bakleh A, Krasnitz A, Nguyen KC, Lakshmi B, Gerald WL, Powers S, Mu D. 2007. Oncogenic cooperation and coamplification of developmental transcription factor genes in lung cancer. *Proc Natl Acad Sci* **104:** 16663–16668.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res* **19:** 1639–1645.

Lee YF, Miller LD, Chan XB, Black MA, Pang B, Ong CW, Salto-Tellez M, Liu ET, Desai KV. 2012. JMJD6 is a driver of cellular proliferation and motility and a marker of poor prognosis in breast cancer. *Breast Cancer Res* **14:** R85.

Loi S, Haibe-Kains B, Desmedt C, Lallemand F, Tutt AM, Gillet C, Ellis P, Harris A, Bergh J, Foekens JA, et al. 2007. Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J Clin Oncol* **25:** 1239–1246.

Malerød L, Stuffers S, Brech A, Stenmark H. 2007. Vps22/EAP30 in ESCRT-II mediates endosomal sorting of growth factor and chemokine receptors destined for lysosomal degradation. *Traffic* **8:** 1617–1629.

Miller LD, Smeds J, George J, Vega VB, Vergara L, Ploner A, Pawitan Y, Hall P, Klaar S, Liu ET, et al. 2005. An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci* **102:** 13550–13555.

Ng CK, Cooke SL, Howe K, Newman S, Xian J, Temple J, Batty EM, Pole JC, Langdon SP, Edwards PA, et al. 2012. The role of tandem duplicator phenotype in tumour evolution in high-grade serous ovarian cancer. *J Pathol* **226:** 703–712.

Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, Raine K, Jones D, Marshall J, Ramakrishna M, et al. 2012. The life history of 21 breast cancers. *Cell* **149:** 994–1007.

Pawitan Y, Bjohle J, Amler L, Borg AL, Egyhazi S, Hall P, Han X, Holmberg L, Huang F, Klaar S, et al. 2005. Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. *Breast Cancer Res* **7:** R953–R964.

Russnes HG, Vollan HK, Lingjaerde OC, Krasnitz A, Lundin P, Naume B, Sørlie T, Borgen E, Rye IH, Langerød A, et al. 2010. Genomic architecture characterizes tumor progression paths and fate in breast cancer patients. *Sci Transl Med* **2:** 38ra47.

Sawey ET, Chanrion M, Cai C, Wu G, Zhang J, Zender L, Zhao A, Busuttil RW, Yee H, Stein L, et al. 2011. Identification of a therapeutic strategy targeting amplified *FGF19* in liver cancer by Oncogenomic screening. *Cancer Cell* **19:** 347–358.

Sengerová B, Wang AT, McHugh PJ. 2011. Orchestrating the nucleases involved in DNA interstrand cross-link (ICL) repair. *Cell Cycle* **10:** 3999–4008.

Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, Turashvili G, Ding J, Tse K, Haffari G, et al. 2012. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* **486:** 395–399.

Solimini NL, Xu Q, Mermel CH, Liang AC, Schlabach MR, Luo J, Burrows AE, Anselmo AN, Bredemeyer AL, Li MZ, et al. 2012. Recurrent hemizygous deletions in cancers may optimize proliferative potential. *Science* **337:** 104–109.

Soon WW, Miller LD, Black MA, Dalmasso C, Chan XB, Pang B, Ong CW, Salto-Tellez M, Desai KV, Liu ET. 2011. Combined genomic and phenotype screening reveals secretory factor SPINK1 as an invasion and survival factor associated with patient prognosis in breast cancer. *EMBO Mol Med* **3:** 451–464.

Staaf J, Jönsson G, Ringnér M, Vallon-Christersson J, Grabau D, Arason A, Gunnarsson H, Agnarsson BA, Malmström PO, Johannsson OT, et al. 2010. High-resolution genomic and expression analyses of copy number alterations in HER2-amplified breast cancer. *Breast Cancer Res* **12:** R25.

Stephens PJ, McBride DJ, Lin ML, Varela I, Pleasance ED, Simpson JT, Stebbings LA, Leroy C, Edkins S, Mudie LJ, et al. 2009. Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* **462:** 1005–1010.

Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, Pleasance ED, Lau KW, Beare D, Stebbings LA, et al. 2011. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144:** 27–40.

Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, Nik-Zainal S, Martin S, Varela I, Bignell GR, et al. 2012. The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486:** 400–404.

Stott FJ, Bates S, James MC, McConnell BB, Starborg M, Brookes S, Palmero I, Ryan K, Hara E, Vousden KH, et al. 1998. The alternative product from the human *CDKN2A* locus, p14^ARF, participates in a regulatory feedback loop with p53 and MDM2. *EMBO J* **17:** 5001–5014.

van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AAM, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, et al. 2002. A gene-

expression signature as a predictor of survival in breast cancer. *N Engl J Med* **347:** 1999–2009.

Venkatraman ES, Olshen AB. 2007. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23:** 657–663.

Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F, Talantov D, Timmermans M, Meijer-van Gelder ME, Yu J, et al. 2005. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* **365:** 671–679.

Xue W, Kitzing T, Roessler S, Zuber J, Krasnitz A, Schultz N, Revill K, Weissmueller S, Rappaport AR, Simon J, et al. 2012. A cluster of cooperating tumor-suppressor gene candidates in chromosomal deletions. *Proc Natl Acad Sci* **109:** 8212–8217.

Zender L, Spector MS, Xue W, Flemming P, Cordon-Cardo C, Silke J, Fan ST, Luk JM, Wigler M, Hannon GJ, et al. 2006. Identification and validation of oncogenes in liver cancer using an integrative oncogenomic approach. *Cell* **125:** 1253–1267.

Zhang J, Liu X, Datta A, Govindarajan K, Tam WL, Han J, George J, Wong C, Ramnarayanan K, Phua TY, et al. 2009. *RCP* is a human breast cancer–promoting gene with Ras-activating function. *J Clin Invest* **119:** 2171–2183.