NEOPLASIA
www.neoplasia.com

Volume 14 Number 8 August 2012 pp. 702–708 **702**

Gene Fusions Associated with Recurrent Amplicons Represent a Class of Passenger Aberrations in Breast Cancer[1,2]

Shanker Kalyana-Sundaram[*,†,‡], Sunita Shankar[*,†], Scott DeRoo[*,†], Matthew K. Iyer[*,§], Nallasivam Palanisamy[*,†], Arul M. Chinnaiyan[*,†,§,¶,3] and Chandan Kumar-Sinha[*,†,3]

*Michigan Center for Translational Pathology, University of Michigan, Ann Arbor, MI; [†]Department of Pathology, University of Michigan, Ann Arbor, MI; [‡]Department of Environmental Biotechnology, Bharathidasan University, Tiruchirappalli, India; [§]Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, MI; [¶]Howard Hughes Medical Institute, University of Michigan Medical School, Ann Arbor, MI

Abstract

Application of high-throughput transcriptome sequencing has spurred highly sensitive detection and discovery of gene fusions in cancer, but distinguishing potentially oncogenic fusions from random, "passenger" aberrations has proven challenging. Here we examine a distinctive group of gene fusions that involve genes present in the loci of chromosomal amplifications—a class of oncogenic aberrations that are widely prevalent in breast cancers. Integrative analysis of a panel of 14 breast cancer cell lines comparing gene fusions discovered by high-throughput transcriptome sequencing and genome-wide copy number aberrations assessed by array comparative genomic hybridization, led to the identification of 77 gene fusions, of which more than 60% were localized to amplicons including 17q12, 17q23, 20q13, chr8q, and others. Many of these fusions appeared to be recurrent or involved highly expressed oncogenic drivers, frequently fused with multiple different partners, but sometimes displaying loss of functional domains. As illustrative examples of the "amplicon-associated" gene fusions, we examined here a recurrent gene fusion involving the mediator of mammalian target of rapamycin signaling, *RPS6KB1* kinase in BT-474, and the therapeutically important receptor tyrosine kinase *EGFR* in MDA-MB-468 breast cancer cell line. These gene fusions comprise a minor allelic fraction relative to the highly expressed full-length transcripts and encode chimera lacking the kinase domains, which do not impart dependence on the respective cells. Our study suggests that amplicon-associated gene fusions in breast cancer primarily represent a by-product of chromosomal amplifications, which constitutes a subset of passenger aberrations and should be factored accordingly during prioritization of gene fusion candidates.

Neoplasia (2012) 14, 702–708

Address all correspondence to: Assistant Professor Chandan Kumar-Sinha, PhD, Department of Pathology, University of Michigan Medical School, 1400 E Medical Center Dr 5316 CCGC, Ann Arbor, MI 48109-0602. E-mail: chakumar@med.umich.edu

[1]This project was supported in part by Department of Defense Breast Cancer Research Program (W81XWH-08-0110), American Association for Cancer Research Stand Up to Cancer (SU2C) award, the National Functional Genomics Center (W81XWH-11-1-0520) from Department of Defense to A.M.C., and, in part, by the US National Institutes of Health through the University of Michigan's Cancer Center Support grant 5 P30 CA46592. A.M.C. is supported by the US National Cancer Institute's Early Detection Research Network (U01 CA111275). A.M.C. is supported by the Doris Duke Charitable Foundation Clinical Scientist Award and the Burroughs Welcome Foundation Award in Clinical Translational Research. A.M.C. is an American Cancer Society Research professor and Taubman Scholar. S.D. is supported by Howard Hughes Medical Institute Medical Student Fellowship.

[2]This article refers to supplementary materials, which are designated by Tables W1 and W2 and Figures W1 to W4 and are available online at www.neoplasia.com.

[3]These authors share senior authorship.

Received 6 June 2012; Revised 9 July 2012; Accepted 10 July 2012

DOI 10.1593/neo.12914

Introduction

Chromosomal amplifications and translocations are among the most common somatic aberrations in cancers [1,2]. Gene amplification is an important mechanism for oncogene overexpression and activation. Numerous recurrent loci of chromosomal amplifications have been characterized in breast cancer, which result in gain of copy number and overexpression of oncogenes such as *ERBB2* on 17q12 (the definitive molecular aberration in 20%-30% of all breast cancers) [3,4], as well as many other oncogenic drivers including *Myc* [5], *EGFR* [6], *FGFR1* [7], *CyclinD1* [8], *RPS6KB1* [9], and others [10]. Chromosomal translocations leading to generation of gene fusions represent another prevalent mechanism for the expression of oncogenes in epithelial cancers [11]. Recently, we described the discovery and characterization of recurrent gene fusions in breast cancer involving MAST family serine threonine kinases and Notch family of transcription factors [12]. Interestingly, we also observed a large number of gene fusions, including some recurrent fusions involving known oncogenes localized at loci of chromosomal amplifications.

Here we carried out a systematic analysis of the association between gene fusions and genomic amplification by integrating RNA-Seq data with array comparative genomic hybridization (aCGH)–based whole-genome copy number profiling from a panel of breast cancer cell lines. We examined a set of "amplicon-associated gene fusions" that refer to all the fusions where one or both gene partners are localized to a site of chromosomal amplification. Specifically, we assessed the functional relevance of two amplicon-associated fusion genes involving oncogenic kinases *EGFR* and *RPS6KB1* in the context of prioritizing fusion candidates important in tumorigenesis. Our results suggest that recurrent gene fusions localized to recurrent amplicons, displaying allelic imbalance between the fusion partners, may represent an epiphenomenon of genomic amplification cycles not essential for cancer development.

Materials and Methods

Gene Fusion Data Set

Chimeric transcript candidates were primarily obtained from paired-end transcriptome sequencing of breast cancer from a total of more than 49 cell lines and 40 tissue samples described previously [12]. aCGH data were generated using Agilent Human Genome 244A CGH Microarrays (Agilent Technologies, Santa Clara, CA) according to the manufacturer's instructions, and data were analyzed using CGH Analytics (Agilent Technologies). Copy number alterations were assessed using ADM-2, with the threshold a setting of 6.0 and a bin size of 10.

RNA Isolation and Complementary DNA Synthesis

Total RNA was isolated using TRIzol and RNeasy Kit (Invitrogen, Carlsbad, CA) with DNase I digestion according to the manufacturer's instructions. RNA integrity was verified on an Agilent Bioanalyzer 2100 (Agilent Technologies). Complementary DNA was synthesized from total RNA using Superscript III (Invitrogen) and random primers (Invitrogen).

Quantitative Real-time Polymerase Chain Reaction

Primers for validation of candidate gene fusions were designed using the National Center for Biotechnology Information Primer Blast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/), with primer pairs spanning exon junctions amplifying 70- to 110-bp products for every chimera tested. Quantitative polymerase chain reaction (QPCR) was performed using SYBR Green MasterMix (Applied Biosystems, Carlsbad, CA) on an Applied Biosystems StepOne Plus Real-Time PCR System. All oligonucleotide primers were obtained from Integrated DNA Technologies and are listed in Table W1. *GAPDH* was used as endogenous control. All assays were performed twice, and results were plotted as average fold change relative to *GAPDH*.

Cell Proliferation Assays

Cells were transfected with small interfering RNAs (siRNAs) using Oligofectamine reagent (Life Sciences, Carlsbad, CA), and 3 days after transfection, the cells were plated for proliferation assays. At the indicated times, cell numbers were counted using Coulter Counter (Indianapolis, IN).

Western Blot

Cell pellets were sonicated in NP-40 lysis buffer (50 mM Tris-HCl, 1% NP-40, pH 7.4; Sigma, St. Louis, MO), complete protease inhibitor mixture (Roche, Indianapolis, IN), and phosphatase inhibitor (EMD Bioscience, San Diego, CA). Immunoblot analysis was carried out using antibodies for *ERBB2* (MS-730-PABX; Thermo Scientific, Fremont, CA) and *RPS6KB1* (2708S; Cell Signaling, Danvers, MA). Human β-actin antibody (Sigma, St. Louis, MO) was used as a loading control.

Knockdown Assays

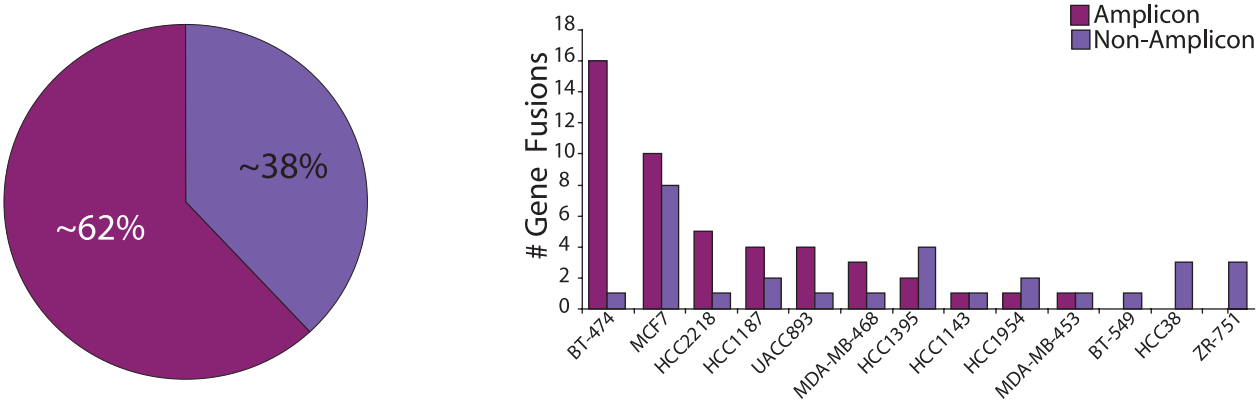
Short hairpin RNAs (shRNAs; Table W1) were transduced in presence of 1 μg/ml polybrene. All siRNA transfections were performed using Oligofectamine reagent (Life Sciences). For siRNA knockdown experiments, multiple custom siRNA sequences targeting the *ARID1A-MAST2* fusion (Thermo, Lafayette, CO) were used [12].

Results

Paired-end transcriptome sequencing of breast cancer cell lines and tissues led to the identification of an average of more than four gene fusions per breast cancer sample [12]. Interestingly, we observed that some of the cell lines with the largest number of gene fusions also harbored many well-known chromosomal amplifications, prompting us to examine a likely association between genomic amplifications and gene fusions. To assess copy number alterations at the chromosomal coordinates of the fusion genes, we analyzed aCGH (244K Agilent array) data in a set of 14 cell lines (Table W2) and observed that as many as 62% of the total number of fusions were associated with regions of amplifications (Figure 1*A*). The genes involved in fusions were found to be significantly associated with their genomic amplification status based on Fisher exact *t* test ($P < .0004$), in four of six cell lines with the maximum number of fusions, including BT-474, MCF7, HCC2218, and UACC893 (Figure 1*B*).

Examining the distribution of fusion genes in individual samples revealed that a majority of the gene fusions were associated with 17q12 amplicon harboring *ERBB2* and 17q23 amplicon that includes genes such as *BCAS3*, *RPS6KB1*, and *TMEM49*, 20q13 amplicon with *BCAS4* and the chr8q amplicon commonly found amplified in breast cancer (Table W2 and Figures 2 and W1). Interestingly, the breast cancer cell line BT-474 that harbors both the chr17 amplicons and the chr20 amplicon and MCF7 with prominent amplifications in chr17, chr20, and chr8q showed the maximum number of gene fusions observed in a sample, accounting for as many as 26 gene fusions associated with amplicons compared against only 9 in unamplified loci (Figures 1 and 2 and Table W2).

A



B

Sample	Fusion Partners	Number of Amplifications	Total Measureable Genes	Amplified Fusion Partners	Not-amplified Fusion Partners	Amplified without Fusion Partner	Unaffected Genes	p-value
MCF7	27	426	17691	11	16	415	17249	1.28E-11
HCC2218	9	446	17691	7	2	439	17243	2.13E-10
BT-474	31	5648	17691	25	6	5623	12038	3.19E-08
UACC-893	8	2902	17691	6	2	2896	14787	0.0004016
HCC1187	10	780	17691	2	8	778	16903	0.06908
HCC1395	12	2111	17691	1	11	2110	15569	0.783

Figure 1. Distribution of gene fusions across breast cancer cell lines. (A) Pie chart representation of the relative proportion of gene fusions associated with loci of genomic amplifications compared to unamplified loci (left) and bar graph representation of the relative distribution of gene fusions across different breast cancer cell lines (right). (B) Table summarizing the statistical significance of association between gene fusions and chromosomal amplifications in breast cancer cell lines with the highest number of gene fusions in A (using Fisher exact *t* test, sorted by *P* value).

In the backdrop of a large number of somatic aberrations seen in cancers, any "recurrent" events observed across samples are generally regarded as potentially "driving" tumorigenesis. Interestingly, among the more than 380 gene fusions reported in our compendium of breast cancer fusions [12], as many as 62 genes were found to be recurrent partners (appear at least twice). Among these, whereas the *MAST* and *Notch* fusions were shown to be functionally recurrent and potentially driving aberrations in up to 5% to 7% of breast cancers, 33 of other recurrent gene fusions were found to be associated with known frequent amplicons, including *ERBB2*, *BCAS3/4*, and chr8q. Among these, three fusions each involved the ikaros family zinc finger protein 3 transcription factor (*IKZF3* on chr17q12 amplicon) and breast carcinoma amplified sequence 3 (*BCAS3* on chr17q23 amplicon) as 3′ partners—all with different 5′ partners. Similarly, tripartite motif containing 37 (*TRIM37* on chr17q23) was a common 5′ partner in three distinct gene fusions with different 3′ partners (Table W2). To further expand our integrative analysis of copy number aberrations and gene fusions, next we used the breast cancer aCGH data [13,14] and observed gene fusion–associated amplicons in MCF7, BT-474, and MDA-MB-468, HCC-1187 as seen in our data as well as in an additional panel of cell lines, including ZR-75-30, SUM190, MDA-MB-361, HCC-1428, and HCC-1569 (Figure W2). Clearly, apart from triggering overexpression of constituent genes, our observations strongly suggest that the loci of chromosomal amplifications also serve as "hotspots" for the generation of recurrent gene fusions.

Next, to assess whether amplicon-associated gene fusions impart oncogenic phenotypes on the cells, we examined the open reading frames (ORFs), functional domains/motifs, and conservation of fusion architecture across different samples. Among recurrent fusion candidates within amplicons, we focused on known cancer-associated partner genes such as kinases, oncogenes, tumor suppressors, or known fusion partners in the Mitelman Database of chromosomal aberrations in cancer [15] and observed several functionally plausible gene fusions. Here we describe our observations with two specific examples of gene fusions involving oncogenic kinases.

The triple-negative breast cancer cell line MDA-MB-468 is known to show an overexpression of epidermal growth factor receptor (EGFR) [16]. In our transcriptome sequencing compendium of 89 breast cancer cell lines and tissues, the highest expression of *EGFR* is observed in MDA-MB-468 (Figure 3*A*), potentially resulting from a focal amplification at chr7p12 (Figure 2). In addition, we detected an *EGFR* fusion transcript (*EGFR-POLD1*) in this cell line, encoding the N-terminal portion of EGFR, completely devoid of the tyrosine kinase domain (Figure 3*A*, *inset*). However, the uniform read-coverage observed across the full length of the *EGFR* transcript in this sample (Figure 3*B*), precluded the existence of any exon imbalance, suggesting that even as the kinase domain is lost in the fusion, the full-length EGFR protein is expressed in this cell line. Further, we observed a remarkable mismatch between the copy numbers of *EGFR* and its fusion partner *POLD1* (Figure 3*C*) that supports a predominant expression of full-length *EGFR* compared with the *EGFR-POLD1* chimera. This is unlike the observation in case of *MAST* kinase fusions in breast cancer characterized in our previous study [12], in which case a marked exon imbalance in coverage was observed (Figure W3). Considering that the

MDA-MB-468 harbors both *MAST2* and *EGFR* fusions, we were intrigued to assess its relative "dependence" on both the kinases. Surprisingly, a profound reduction in cell proliferation was observed on siRNA knockdown of *MAST2*, whereas *EGFR* knockdown showed little effect (Figure 3*D*). Next, testing the possibility of *EGFR* amplicon potentially cooperating with *MAST2*, we found that the effect of combined knockdown of *EGFR* and *MAST2* was comparable with that of *MAST2* knockdown alone (Figure 3*D*), further suggesting that *EGFR* amplification does not signify a driver aberration. In this context, the EGFR fusion transcript that represents a miniscule fraction of overall EGFR expression and encodes only the N-terminal portion lacking the kinase domain was reckoned to be inconsequential.

Next, we looked at recurrent gene fusions involving oncogenic serine threonine kinase ribosomal protein S6 kinase on chr17q23 frequently amplified in breast cancers [17–20] identified in BT-474 (*RPS6KB1-SNF8*) and MCF7 (*RPS6KB1-VMP1*). Both of these cell lines harbor amplifications at the *RPS6KB1* locus and express the highest levels of *RPS6KB1* among all the samples examined (Figure 4*A*). Both the chimeric transcripts retain only the first exon of *RPS6KB1* and the respective open reading frames show a complete loss of the kinase domain (Figure 4*A*, *inset*). We also observed an even read coverage across the *RPS6KB1* transcript in both fusion-positive cell lines, similar to a representative benign mammary epithelial cell line, albeit at a much higher level, indicating that full-length *RPS6KB1* protein is encoded in these samples (Figures 4*B* and W4*A*). Further, the difference between the copy number observed between the fusion partners in both the *RPS6KB1* fusions (Figures 4*C* and W4*B*) indicates an allelic imbalance between the full-length and the putative fusion genes. Next, considering that BT-474 is an *ERBB2*-positive cell line, we tested potential dependence of these cells on the RPS6KB1 protein. Surprisingly,

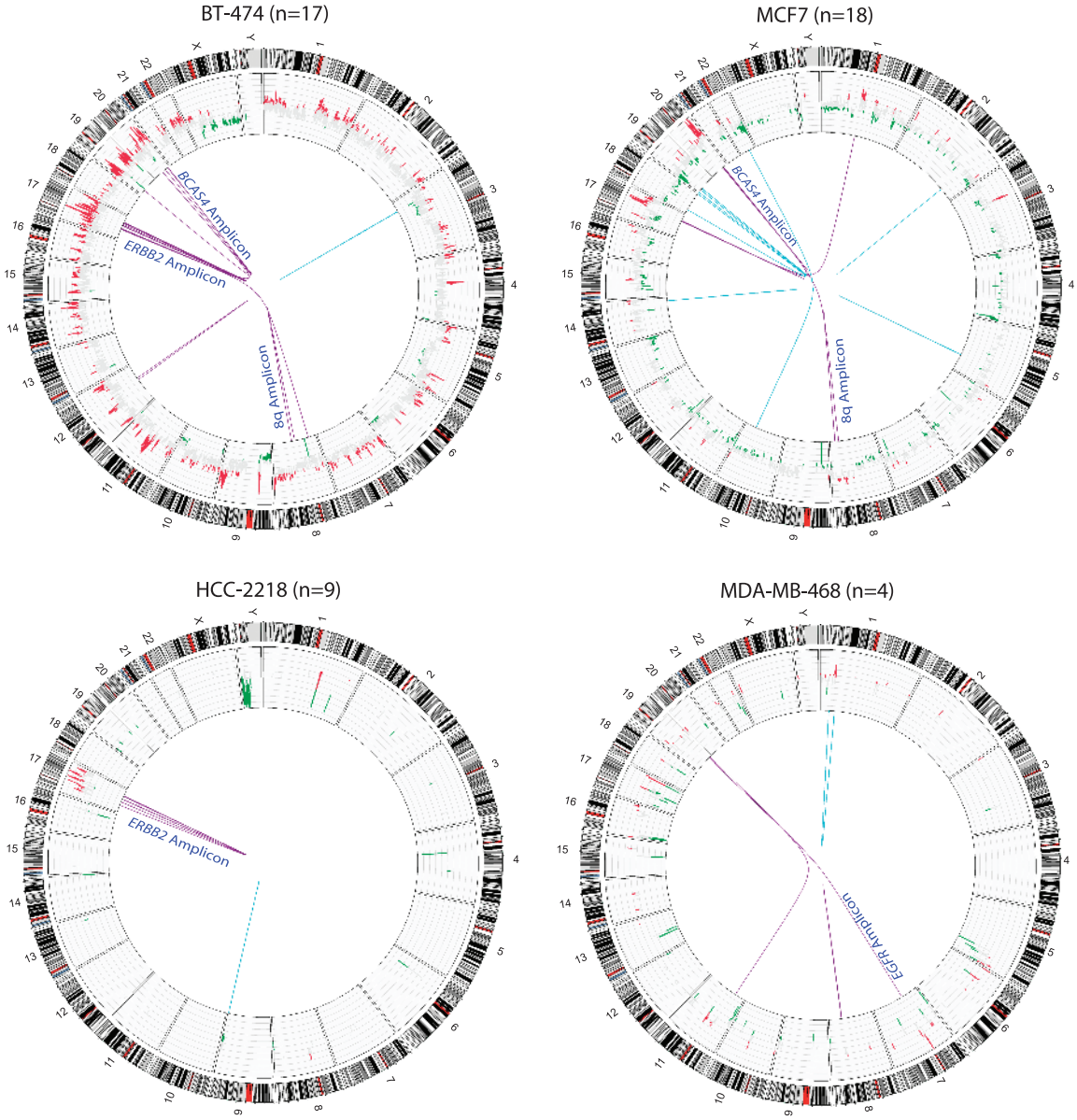


Figure 2. Graphical representation of integrative analysis of gene fusions with copy number analysis. Circos plots of the genome-wide distribution of gene fusions along with status of copy number alterations. Red and green peaks represent amplifications and deletions; purple and cyan lines represent the fusions associated with amplicons and nonamplicons, respectively. "*n*" refers to the total number of fusions identified.

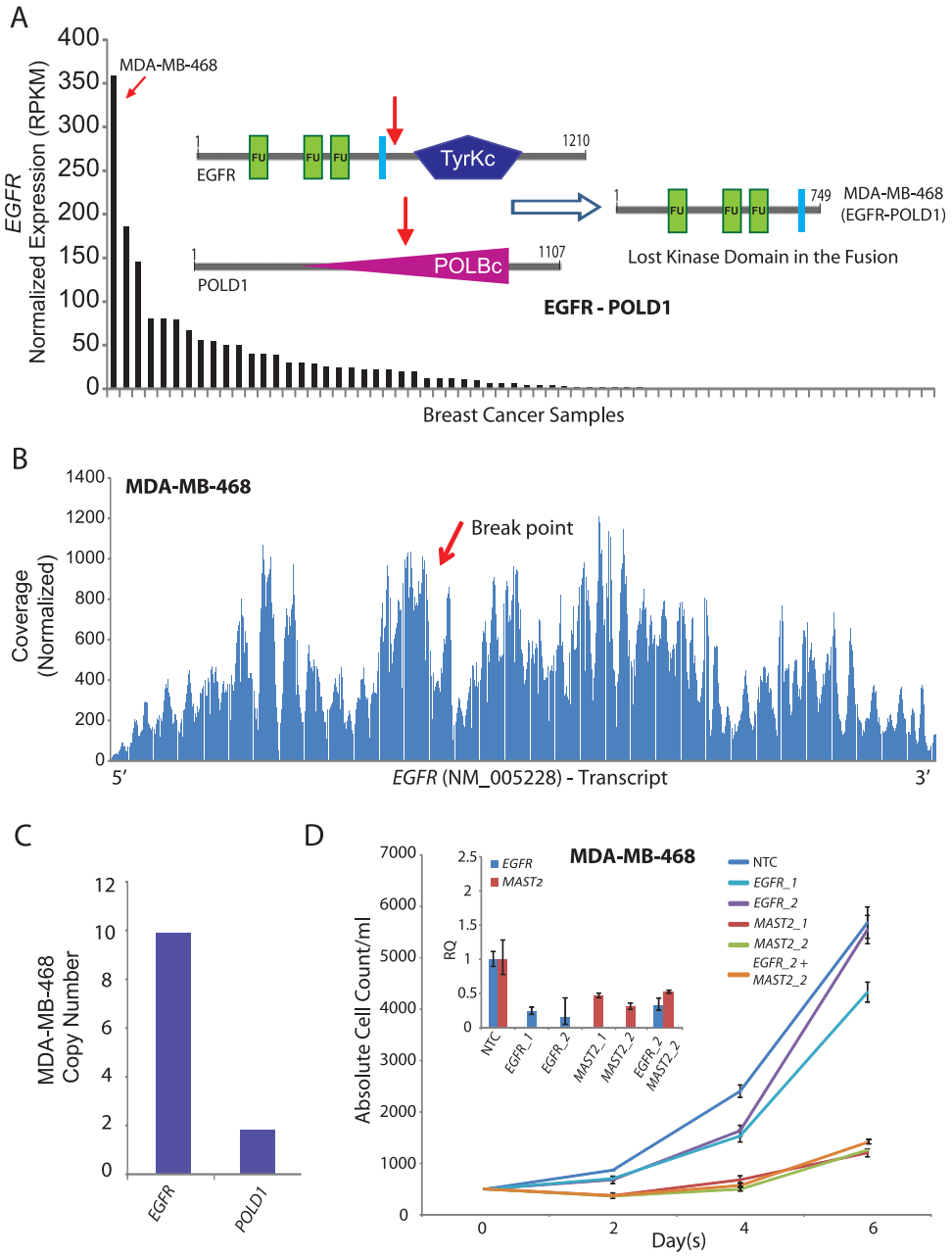


Figure 3. (A) Normalized expression (RPKM) of *EGFR* in descending order of expression in a panel of breast cancer samples obtained from RNA-Seq. Schematic representation of wild-type EGFR and POLD1 proteins with putative breakpoints indicated by red arrows and the domain structure of the putative fusion protein (inset). (B) Plot of normalized coverage of *EGFR* transcript in MDA-MB-468 cell line showing the location of the breakpoint (indicated by red arrow). (C) Bar graph representing the copy number of *EGFR* and *POLD1* in MDA-MB-468. (D) Proliferation assay showing absolute cell count (*y* axis) over a time course (*x* axis) after knockdown with *EGFR* and/or *MAST2* siRNAs in MDA-MB-468. QPCR assessment of knockdown efficiencies relative to nontargeted control (NTC; inset).

similar to our observations with *EGFR* knockdown in MDA-MB-468 cells, here we observed only a small effect on cell proliferation after shRNA knockdown of *RPS6KB1*, in dramatic contrast to the effect of *ERBB2* knockdown (Figure 4*D*). Notably, the shRNA knockdown of RPS6KB1 led to a significant depletion of the full-length protein yet it did not affect cell proliferation compared with ERBB2 protein depletion (Figure 4*D*, *inset*). Therefore, BT-474 cells do not display a dependence on RPS6KB1 protein, and considering that the RPS6KB1 fusion product is completely devoid of all functional domains of RPS6KB1, including the kinase domain, this fusion also likely represents a passenger event.

Discussion

In our systematic search for gene fusions in breast cancer using high-throughput transcriptome sequencing, we observed a notably large number of fusion genes associated with many well characterized recurrent amplicons, including 17q12, 17q23, 20q13, and 8q, among others. Amplicon-associated gene fusions were found to involve complex and cryptic rearrangements, involving one or both partners within the amplicon site, with the chimeric transcript expression apparently concealed in the backdrop of highly expressed wild-type genes. The gene fusions considered here include only "expressed" chimeric transcripts derived from known/annotated fusion partners. Chromosomal rearrangements

that do not express chimeric transcripts or that involve unannotated fusion partners are excluded from this analysis. This likely accounts for the variability observed in the number of gene fusions scored across multiple samples with known amplicons. Because many of the fusions at the amplicons appeared to be recurrent, although frequently fused with multiple different partners, it led us to examine whether the recurrence was incidentally associated with recurrent amplicons or signified functionally important aberrations.

MDA-MB-468 represents a prototype triple-negative breast cancer cell line with a "basal-like" gene expression profile that shows an overexpression of the oncogenic kinase *EGFR* due to a focal amplification at chr7p12. Here we discovered a chimeric transcript involving *EGFR*. However, careful examination of this transcript revealed that the fusion encodes N-terminal *EGFR* protein, without the kinase domain. Transcriptome sequencing did not show evidence of fusion-associated exon imbalance in *EGFR* expression, suggesting that full-length *EGFR* is expressed in this cell line. In addition, the significantly higher genomic copy number of *EGFR* compared to its fusion partner *POLD1* suggests that a minor allelic fraction of the *EGFR* is involved in fusion with *POLD1*, whereas other amplified copies of the gene

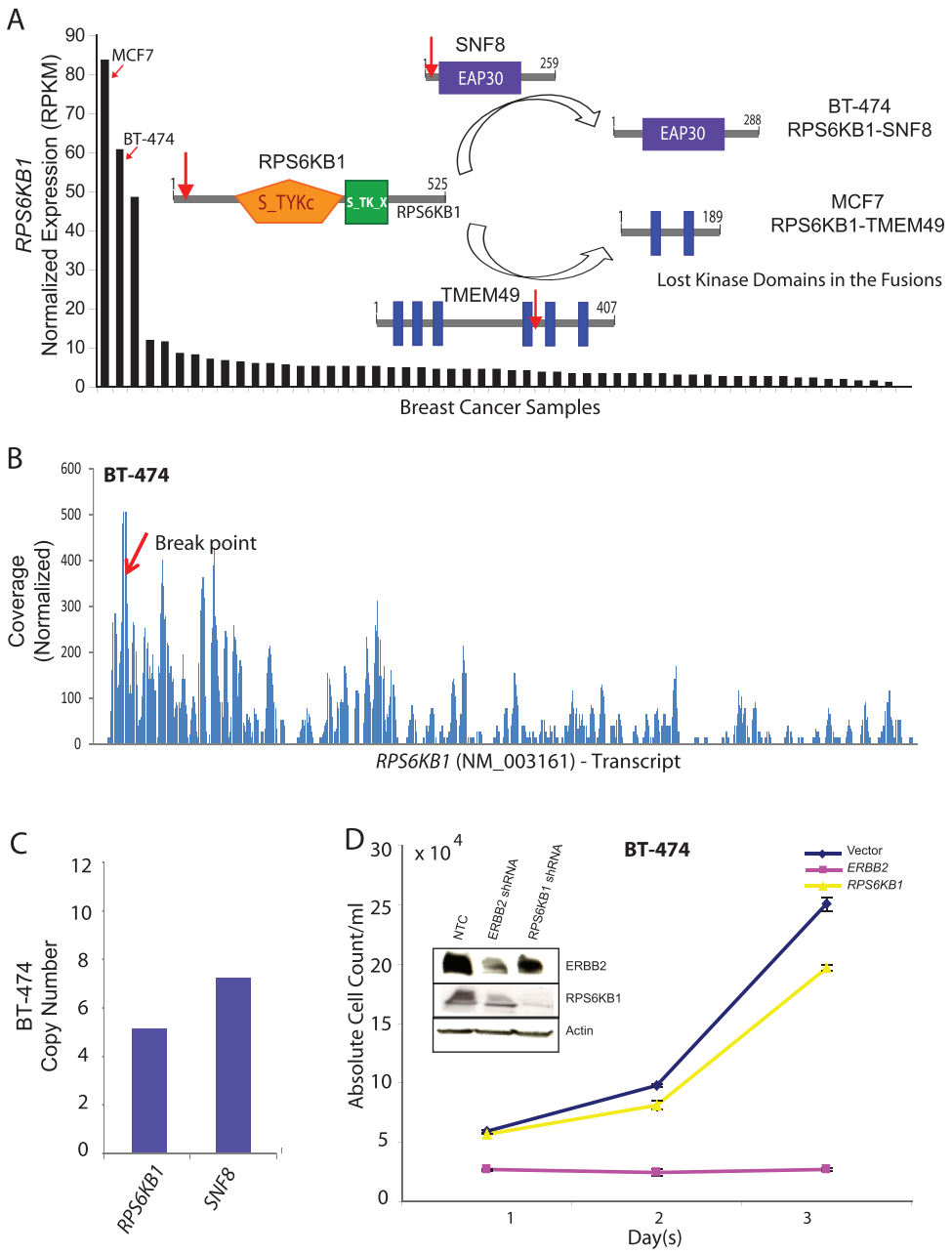


Figure 4. (A) Normalized expression (RPKM) of *RPS6KB1* in descending order of expression in a panel of breast cancer samples obtained from RNA-Seq. Schematic representation of wild-type RPS6KB1, TMEM49, and SNF8 proteins with putative breakpoints indicated by red arrows and the domain structure of the putative fusion proteins in BT-474 and MCF7 (inset). (B) Plot of normalized coverage of *RPS6KB1* transcript in BT-474 cell line showing the location of the breakpoint (indicated by red arrow). (C) Bar graph representing the copy number of *RPS6KB1* and *SNF8* in BT-474 (D) Proliferation assay showing absolute cell count (*y* axis) over a time course (*x* axis) after knockdown with *RPS6KB1* and/or *ERBB2* shRNAs in BT-474. Western blot assessment of the knockdown efficiency relative to nontargeted control (NTC). Actin was used as a loading control (inset).

express the full-length molecule. Technically, the detection and monitoring of the *EGFR* fusion transcript in the backdrop of extremely high levels of wild-type *EGFR* transcript is challenging; therefore, we chose to assess the dependency imparted by full-length *EGFR*. Interestingly, the knockdown of *EGFR* had only a slight effect on the proliferation of MDA-MB-468 cells, whereas a profound reduction in cell proliferation was observed on the knockdown the fusion gene *MAST2*. Combined knockdown of *MAST2* and *EGFR* produced the same effect as that by MAST2 alone, further calling into question the credentials of *EGFR* as a driver aberration in MDA-MB-468 cells. Interestingly, MDA-MB-468 is known to be insensitive to EGFR inhibitors like erlotinib [21] and gefitinib [22].

Similarly, the recurrent gene fusions involving *RPS6KB1* retain only the first exon, and the chimeric ORFs show a complete loss of the kinase domain in breast cancer cell lines BT-474 and MCF7. Similar to the *EGFR* fusion, DNA copy number analysis and RNA-Seq data provided the evidence that full-length RPS6KB1 protein is encoded in both these cell lines. Notably, both BT-474 and MCF7 have been shown to express high levels of full-length RPS6KB1 protein [23], suggesting that these cells exhibit elevated activity of RPS6KB1 as a result of amplification, independent of the fusion. Again, similar to *EGFR* knockdown in MDA-MB-468, *RPS6KB1* knockdown in BT-474 (an ERBB2-positive cell line) showed an insignificant effect on cell proliferation compared to *ERBB2* knockdown. Interestingly, in a previous study, knockdown of *RPS6KB1* was found to have no effect on apoptosis in both BT-474 and MCF7 breast cancer cells [24].

In the light of our observations, we surmise that repeated breaks and rejoining of chromosomes during chromosomal amplifications led to the generation of amplicon-associated gene fusions. Loci of recurrent genomic amplifications thus engender "pseudo" recurrent gene fusions that may largely represent passenger aberrations involving random breakpoints. The two cell lines with established drivers—*ERBB2* in BT-474 and *MAST2* in MDA-MB-468—made it possible for us to assess the relative importance of amplicon fusions involving *RPS6KB1* and *EGFR*, respectively. In cases where a driver is not clearly apparent, a more careful examination of all plausible fusion candidates will be required. Importantly, even as our study primarily pertains to breast cancers based on available data and a well-documented preponderance of copy number aberrations in breast cancers [10], we expect the association between amplicons and gene fusions to be consistent in other cancers as well. We argue here for a measure of caution in considering the functional implications of recurrent gene fusions associated with amplifications because these may be simply a result of massive chromosomal upheaval at the amplicons, not representing clonally selected oncogenic events.

Acknowledgments

The authors thank Vishal Kothari and Nikita Consul for technical help.

References

[1] Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, and Stratton MR (2004). A census of human cancer genes. *Nat Rev Cancer* **4**, 177–183.

[2] Santarius T, Shipley J, Brewer D, Stratton MR, and Cooper CS (2010). A census of amplified and overexpressed human cancer genes. *Nat Rev Cancer* **10**, 59–64.

[3] Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, and McGuire WL (1987). Human breast cancer: correlation of relapse and survival with amplification of the HER-2/*neu* oncogene. *Science* **235**, 177–182.

[4] Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, Fleming T, Eiermann W, Wolter J, Pegram M, et al. (2001). Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* **344**, 783–792.

[5] Deming SL, Nass SJ, Dickson RB, and Trock BJ (2000). C-*myc* amplification in breast cancer: a meta-analysis of its occurrence and prognostic relevance. *Br J Cancer* **83**, 1688–1695.

[6] Bhargava R, Gerald WL, Li AR, Pan Q, Lal P, Ladanyi M, and Chen B (2005). EGFR gene amplification in breast cancer: correlation with epidermal growth factor receptor mRNA and protein expression and HER-2 status and absence of EGFR-activating mutations. *Mod Pathol* **18**, 1027–1033.

[7] Elbauomy Elsheikh S, Green AR, Lambros MB, Turner NC, Grainge MJ, Powe D, Ellis IO, and Reis-Filho JS (2007). FGFR1 amplification in breast carcinomas: a chromogenic *in situ* hybridisation analysis. *Breast Cancer Res* **9**, R23.

[8] Elsheikh S, Green AR, Aleskandarany MA, Grainge M, Paish CE, Lambros MB, Reis-Filho JS, and Ellis IO (2008). CCND1 amplification and cyclin D1 expression in breast cancer and their relation with proteomic subgroups and patient outcome. *Breast Cancer Res Treat* **109**, 325–335.

[9] 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.

[10] Curtis C, Shah SP, Chin S-F, 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.

[11] Chinnaiyan AM and Palanisamy N (2010). Chromosomal aberrations in solid tumors. *Prog Mol Biol Transl Sci* **95**, 55–94.

[12] Robinson DR, Kalyana-Sundaram S, Wu YM, Shankar S, Cao X, Ateeq B, Asangani IA, Iyer M, Maher CA, Grasso CS, et al. (2011). Functionally recurrent rearrangements of the MAST kinase and Notch gene families in breast cancer. *Nat Med* **17**, 1646–1651.

[13] Pollack JR, Sorlie T, Perou CM, Rees CA, Jeffrey SS, Lonning PE, Tibshirani R, Botstein D, Borresen-Dale AL, and Brown PO (2002). Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc Natl Acad Sci USA* **99**, 12963–12968.

[14] Greshock J, Naylor TL, Margolin A, Diskin S, Cleaver SH, Futreal PA, deJong PJ, Zhao S, Liebman M, and Weber BL (2004). 1-Mb resolution array-based comparative genomic hybridization using a BAC clone set optimized for cancer gene analysis. *Genome Res* **14**, 179–187.

[15] Mitelman F, Johansson B, and Mertens F (2010). *Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer*. Cancer Genome Anatomy Project. Available at: http://cgap.nci.nih.gov/Chromosomes/Mitelman. Accessed March 2012.

[16] Hyatt DC and Ceresa BP (2008). Cellular localization of the activated EGFR determines its effect on cell growth in MDA-MB-468 cells. *Exp Cell Res* **314**, 3415–3425.

[17] Barlund M, Monni O, Kononen J, Cornelison R, Torhorst J, Sauter G, Kallioniemi O-P, and Kallioniemi A (2000). Multiple genes at 17q23 undergo amplification and overexpression in breast cancer. *Cancer Res* **60**, 5340–5344.

[18] Couch FJ, Wang XY, Wu GJ, Qian J, Jenkins RB, and James CD (1999). Localization of PS6K to chromosomal region 17q23 and determination of its amplification in breast cancer. *Cancer Res* **59**, 1408–1411.

[19] Monni O, Barlund M, Mousses S, Kononen J, Sauter G, Heiskanen M, Paavola P, Avela K, Chen Y, Bittner ML, et al. (2001). Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. *Proc Natl Acad Sci USA* **98**, 5711–5716.

[20] Sinclair CS, Rowley M, Naderi A, and Couch FJ (2003). The 17q23 amplicon and breast cancer. *Breast Cancer Res Treat* **78**, 313–322.

[21] Bartholomeusz C, Yamasaki F, Saso H, Kurisu K, Hortobagyi GN, and Ueno NT (2011). Gemcitabine overcomes erlotinib resistance in EGFR-overexpressing cancer cells through downregulation of Akt. *J Cancer* **2**, 435–442.

[22] Maiello MR, D'Alessio A, De Luca A, Carotenuto A, Rachiglio AM, Napolitano M, Cito L, Guzzo A, and Normanno N (2007). AZD3409 inhibits the growth of breast cancer cells with intrinsic resistance to the EGFR tyrosine kinase inhibitor gefitinib. *Breast Cancer Res Treat* **102**, 275–282.

[23] Yamnik RL, Digilova A, Davis DC, Brodt ZN, Murphy CJ, and Holz MK (2009). S6 kinase 1 regulates estrogen receptor alpha in control of breast cancer cell proliferation. *J Biol Chem* **284**, 6361–6369.

[24] Heinonen H, Nieminen A, Saarela M, Kallioniemi A, Klefstrom J, Hautaniemi S, and Monni O (2008). Deciphering downstream gene targets of PI3K/mTOR/p70S6K pathway in breast cancer. *BMC Genomics* **9**, 348.

Table W1. Primer Sequences and siRNA/shRNA Clone Details.

Gene Symbol	Clone ID
EGFR	LU-003114-00-0002
ERBB2	SHCLNV-NM_004448
RPS6KB1	SHCLNV-NM_003161

Primer	Sequence
EGFR-f1	GGGCCAGGTCTTGAAGGCTGT
EGFR-r1	ATCCCCAGGGCCACCACCAG
EGFR-f2	ACACCCTGGTCTGGAAGTACGCA
EGFR-r2	AGTGGGAGACTAAAGTCAGACAGTGAA
EGFR-f3	CCGAGGCAGGGAATGCGTGG
EGFR-r3	TGGCCTGAGGCAGGCACTCT
ERBB2-f1	TGCGCAGGCAGTGATGAGAGT
ERBB2-r1	TCTCGGGACTGGCAGGGAGC
ERBB2-f2	TCCTCCTCGCCCTCTTGCCC
ERBB2-r2	TCTCGGGACTGGCAGGGAGC
RPS6KB1-f1	TGCTGACTGGAGCACCCCCA
RPS6KB1-r1	GCTTCTTGTGTGAGGTAGGGAGGC
GAPDH-f1	GGCTGAGAACGGGAAGCTTGTCA
GAPDH-r1	TCTCCATGGTGGTGAAGACGCCA
MAST2_f1	GAAGTGAGTGAGGATGGCTGCCTT
MAST2_r1	GAGCCGCTCCATGCTGCTGTAC
MAST2_f2	ATTGAGGGCCATGGGGCATCT
MAST2_r2	CCCCATAGGCGCCATTGCTGATG

Table W2. List of Gene Fusions Identified in 14 Breast Cancer Cell Lines, along with Their Copy Number Status.

Sample Name	5′ Gene	3′ Gene	Type	Sequencing Platform	No. Reads	Validation	Chromosomal Location		aCGH Data (5′ and 3′)				Amplicon Status
						Fusion QPCR	5′ Gene	3′ Gene	No. Probe	Average Log Ratio	No. Probe	Average Log Ratio	
BT-474	*RPS6KB1*	SNF8	Intra	GA II	92	Y	chr17:55325224-55382568	chr17:44362457-44377153	5	2.890	2	3.557	Yes
BT-474	*STX16*	RAE1	Intra	GA II	79	Y	chr20:56659733-56687988	chr20:55360024-55386926	4	2.910	4	2.910	Yes
BT-474	*ZMYND8*	CEP250	Intra	GA II	77	Y	chr20:45271787-45418881	chr20:33506636-33563217	15	3.650	5	1.876	Yes
BT-474	*TRPC4AP*	MRPL45	Inter	GA II	30		chr20:33053867-33144279	chr17:33706516-33732628	11	3.290	4	3.452	Yes
BT-474	*MED1*	STXBP4	Intra	GA II	28	Y	chr17:34814063-34861053	chr17:50401124-50596448	4	4.029	21	2.507	Yes
BT-474	*TOB1*	AP1GBP1	Intra	GA II	16		chr17:46294585-46296412	chr17:32949013-33043559	1	2.787	10	2.556	Yes
BT-474	*ACACA*	STAC2	Intra	GA II	15		chr17:32516039-32841015	chr17:34620314-34635566	35	2.556	3	4.029	Yes
BT-474	*MED13*	BCAS3	Intra	GA II	13	Y	chr17:57374747-57497425	chr17:56109953-56824981	13	1.012	73	1.934	Yes
BT-474	*VAPB*	IKZF3	Inter	GA II	13	Y	chr20:56397580-56459562	chr17:35174724-35273967	7	3.404	10	3.701	Yes
BT-474	*RAB22A*	MYO9B	Inter	GA II	9	Y	chr20:56318176-56375969	chr19:17047590-17185104	6	3.404	13	2.122	Yes
BT-474	*GLB1*	CMTM7	Intra	GA II	7		chr3:33013103-33113698	chr3:32408166-32471337	11	−0.425	6	0.428	
BT-474	*NCOA2*	ZNF704	Intra	GA II	7	Y	chr8:71186820-71478574	chr8:81703240-81949571	35	0.916	26	0.640	Yes
BT-474	*BCAS3*	MED13	Intra	GA II	6		chr17:56109953-56824981	chr17:57374747-57497425	73	1.934	13	1.012	Yes
BT-474	*PIP4K2B*	RAD51C	Intra	GA II	6		chr17:34175469-34209684	chr17:54124961-54166691	6	4.813	5	1.700	Yes
BT-474	*PPP1R12A*	MGAT4C	Intra	GA II	6		chr12:78691473-78853366	chr12:84897167-85756812	19	1.218	90	−0.397	Yes
BT-474	*STARD3*	DOCK5	Inter	GA II	6		chr17:35046858-35073980	chr8:25098203-25326536	5	4.821	27	0.076	Yes
BT-474	*TRIM37*	MYO19	Intra	GA II	6		chr17:54414781-54539048	chr17:31925711-31965418	14	2.244	6	2.344	Yes
BT-483	*SMARCB1*	MARK3	Inter	GA II	7	Y	chr22:22459149-22506705	chr14:102921453-103039919	8	1.170	17	0.381	Yes
BT-549	*CLTC*	TMEM49	Intra	GA II	18	Y	chr17:55051831-55129099	chr17:55139644-55272734	9	−0.283	18	−1.185	
HCC1143	*C18orf45*	HM13	Inter	GA II	25	Y	chr18:19129977-19271923	chr20:29565901-29591257	18	1.280	2	1.403	Yes
HCC1143	*C2ORF48*	RRM2	Intra	GA II	23	Y	chr2:10198959-10269307	chr2:10180145-10188997	8	0.134	2	0.134	
HCC1187	*PUM1*	TRERF1	Inter	GA II	38	Y	chr1:31176939-31311151	chr6:42300646-42527761	14	1.648	27	0.336	Yes
HCC1187	*SEC22B*	NOTCH2	Intra	GA II	30	Y	chr1:143807763-143828279	chr1:120255698-120413799	2	1.557	11	0.253	Yes
HCC1187	*CTAGE5*	SIP1	Intra	GA II	15		chr14:38806079-38890148	chr14:38653238-38675928	9	0.940	4	0.235	Yes
HCC1187	*MCPH1*	AGPAT5	Intra	GA II	11		chr8:6251520-6488548	chr8:6553285-6606429	29	0.495	5	0.738	
HCC1187	*KLK5*	CDH23	Inter	GA II	5		chr19:56138370-56148156	chr10:73225533-73245710	3	0.888	1	0.953	Yes
HCC1187	*BC041478*	EXOSC10	Inter	GA II	3		chr19:42434668-42446354	chr1:11049262-11082525	1	0.816	4	0.156	
HCC1395	*EIF3K*	CYP39A1	Inter	GA II	13	Y	chr19:43801561-43819435	chr6:46625403-46728482	2	0.852	11	0.611	
HCC1395	*HNRNPUL2*	AHNAK	Intra	GA II	13	Y	chr11:62238795-62251397	chr11:62039949-62070908	2	0.629	5	1.172	Yes
HCC1395	*RAB7A*	LRCH3	Intra	GA II	6		chr3:129927668-130016331	chr3:199002541-199082853	10	0.755	11	−0.615	
HCC1395	*ERO1L*	FERMT2	Intra	GA II	5		chr14:52178354-52232169	chr14:52395955-52487565	7	0.934	14	0.934	Yes
HCC1395	*FOSL2*	BRE	Intra	GA II	5		chr2:28469282-28491020	chr2:27966985-28415271	3	0.480	51	0.480	
HCC1395	*BCAR3*	ABCA4	Intra	GA II	4		chr1:93799936-93919973	chr1:94230981-94359293	13	0.849	13	0.849	
HCC1954	*C6orf106*	SPDEF	Intra	GA II	24	Y	chr6:34663048-34772603	chr6:34613557-34632069	13	0.036	3	0.374	
HCC1954	*INTS1*	PRKAR1B	Intra	GA II	22	Y	chr7:1476438-1510544	chr7:555359-718687	4	1.034			Yes
HCC1954	*GALNT7*	ORC4L	Inter	GA II	9		chr4:174326478-174481693	chr2:148408201-148494933	15	0.409	7	0.504	
HCC2218	*SEC16A*	NOTCH1	Intra	GA II	14	Y	chr9:138454368-138497328	chr9:138508716-138560059	6	0.000	7	−0.967	
HCC2218	*POLDIP2*	BRIP1	Intra	GA II	8		chr17:23697785-23708730	chr17:57111328-57295702	3	1.113	19	3.925	Yes
HCC2218	*INTS2*	ZNF652	Intra	GA II	7		chr17:57297509-57360159	chr17:44721566-44794834	9	3.925	6	2.649	Yes
HCC2218	*INTS2*	TMEM49	Intra	GA II	5		chr17:57297509-57360159	chr17:55139644-55272734	9	3.925	18	3.202	Yes
HCC2218	*LRRC59*	NEUROD2	Intra	GA II	5		chr17:45813592-45829913	chr17:35013546-35017701	3	2.649	1	3.451	Yes
HCC2218	*PERLD1*	PPM1D	Intra	GA II	4	Y	chr17:35082579-35097833	chr17:56032335-56096818	2	3.451	7	3.340	Yes
MCF7	*BCAS4*	BCAS3	Inter	GA II	2788		chr20:48844873-48927121	chr17:56109953-56824981	7	2.107	73	2.653	Yes
MCF7	*ARFGEF2*	SULF2	Intra	GA II	305	Y	chr20:46971681-47086637	chr20:45719556-45848215	11	0.823	13	3.398	Yes
MCF7	*RPS6KB1*	TMEM49	Intra	GA II	78	Y	chr17:55325224-55382568	chr17:55139644-55272734	5	3.412	18	2.197	Yes
MCF7	*STK11*	MIDN	Intra	GA II	25		chr19:1156797-1179434	chr19:1199551-1210142	4	−1.367	2	−0.279	
MCF7	*PAPOLA*	AK7	Intra	GA II	16	Y	chr14:96038472-96103201	chr14:95928200-96024865	7	0.343	13	0.343	
MCF7	*AHCYL1*	RAD51C	Inter	GA II	12	Y	chr1:110328830-110367887	chr17:54124961-54166691	4	−0.063	5	2.788	Yes
MCF7	*EIF3H*	FAM65C	Inter	GA II	11		chr8:117726235-117837243	chr20:48636052-48686833	12	0.456	5	1.554	Yes
MCF7	*BC017255*	TMEM49	Intra	GA II	10		chr17:54538741-54550409	chr17:55139644-55272734	1	3.515	18	2.197	Yes
MCF7	*ADAMTS19*	SLC27A6	Intra	GA II	9		chr5:128824001-129102275	chr5:128329108-128397234	30	0.051	8	0.051	
MCF7	*ARHGAP19*	DRG1	Inter	GA II	8	Y	chr10:98971919-99042403	chr22:30125538-30160172	8	0.387	5	−0.420	

Table W2. (*continued*)

Sample Name	5′ Gene	3′ Gene	Type	Sequencing Platform	No. Reads	Validation	Chromosomal Location		aCGH Data (5′ and 3′)				Amplicon Status
						Fusion QPCR	5′ Gene	3′ Gene	No. Probe	Average Log Ratio	No. Probe	Average Log Ratio	
MCF7	*MYO9B*	FCHO1	Intra	GA II	8	Y	chr19:17047590-17185104	chr19:17719526-17760377	13	−1.126	4	−0.529	
MCF7	*HSPE1*	PREI3	Intra	GA II	6	Y	chr2:198072965-198076432	chr2:198089016-198125760	1	−0.361	4	−0.361	
MCF7	*PARD6G*	C18ORF1	Intra	GA II	6		chr18:76016105-76106388	chr18:13601664-13642753	10	−0.674	5	−0.407	
MCF7	*TRIM37*	TMEM49	Intra	GA II	6	Y	chr17:54414781-54539048	chr17:55139644-55272734	14	3.515	18	2.197	Yes
MCF7	*SMARCA4*	CARM1	Intra	GA II	5	Y	chr19:10955827-11033958	chr19:10843252-10894448	8	0.041	6	0.041	
MCF7	*BCAS4*	ZMYND8	Intra	GA II	4	Y	chr20:48844873-48927121	chr20:45271787-45418881	7	2.107	15	3.860	Yes
MCF7	*PVT1 (BC041065)*	MYC	Intra	GA II	4	Y	chr8:128875961-129182681	chr8:128817496-128822862	27	1.186	3	1.186	Yes
MCF7	*TRIM37*	RNFT1	Intra	GA II	3		chr17:54414781-54539048	chr17:55384504-55396899	14	3.515	2	3.412	Yes
MDA-MB-361	*TMEM104*	CRKRS	Intra	GA II	18	Y	chr17:70284216-70347517	chr17:34871265-34944326	9	2.327	7	1.529	Yes
MDA-MB-361	*TANC1*	MTMR4	Inter	GA II	12	Y	chr2:159533391-159797416	chr17:53921891-53950250	27	0.000	6	1.658	Yes
MDA-MB-361	*TOX3*	GNAO1	Intra	GA II	7		chr16:51029418-51139215	chr16:54782751-54939612	10	−0.157	19	0.281	
MDA-MB-453	*MECP2*	TMLHE	Intra	GA II	8		chrX:152948879-153016382	chrX:154375389-154495816	8	1.611	11	1.602	Yes
MDA-MB-453	*MYO15B*	MAP3K3	Intra	GA II	4		chr17:71095733-71134522	chr17:59053532-59127402	3	0.543	10	0.494	
MDA-MB-468	*UBR5*	SLC25A32	Intra	GA II	8		chr8:103334744-103493671	chr8:104480041-104496644	18	0.070	4	0.927	Yes
MDA-MB-468	*ARID1A*	MAST2	Intra	GA II	5	Y	chr1:26895108-26981188	chr1:46041871-46274383	10	0.266	23	0.818	
MDA-MB-468	*EGFR*	POLD1	Inter	GA II	5		chr7:55054218-55203822	chr19:55579404-55613083	17	4.944	4	0.732	Yes
MDA-MB-468	*RDH13*	FBXO3	Inter	GA II	3		chr19:60247503-60266397	chr11:33724866-33752647	2	0.853	3	1.507	Yes
UACC-893	*FBXL20*	CRKRS	Intra	GA II	31	Y	chr17:34662422-34811435	chr17:34871265-34944326	17	2.069	7	4.175	Yes
UACC-893	*CCDC6*	ANK3	Intra	GA II	27	Y	chr10:61218511-61336420	chr10:61458164-61570752	17	0.890	13	0.890	
UACC-893	*grb7V*	PPP1R1B	Intra	GA II	23	Y	chr17:35152031-35157064	chr17:35038278-35046404	1	4.843	2	4.843	Yes
UACC-893	*MED1*	IKZF3	Intra	GA II	9	Y	chr17:34814063-34861053	chr17:35174724-35273967	4	3.908	10	4.843	Yes
UACC-893	*EIF2AK3*	PRKD3	Intra	GA II	5		chr2:88637373-88708209	chr2:37331149-37397726	8	1.213	8	1.278	Yes
ZR-75-1	*FOXJ3*	CAMTA1	Intra	GA II	10		chr1:42414796-42573490	chr1:6767970-6854694	17	−0.380	10	−0.089	
ZR-75-1	*GPATCH3*	CAMTA1	Intra	GA II	10		chr1:27089565-27099549	chr1:6767970-6854694	3	−0.225	10	−0.089	
ZR-75-1	*C1ORF151*	RCC2	Intra	GA II	9		chr1:19796057-19828901	chr1:17605837-17637644	4	−0.013	4	−0.225	

Fusions with a recurrent partner are highlighted in yellow.

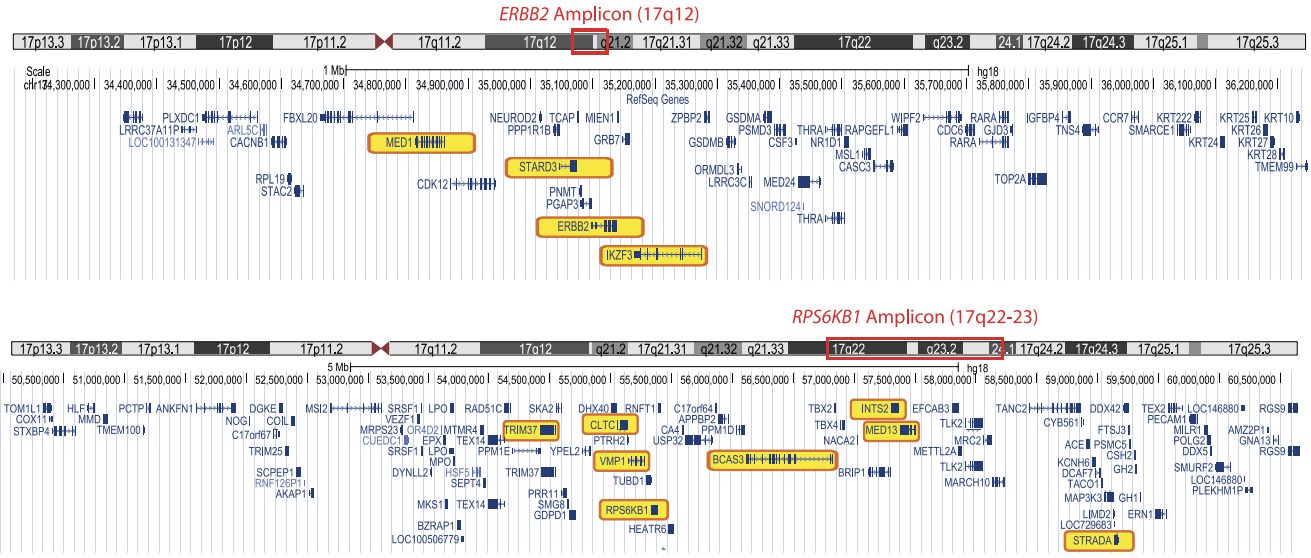


Figure W1. UCSC tracks displaying the *ERRB2* and *RPS6KB1* amplicons, with fusion genes highlighted in yellow.

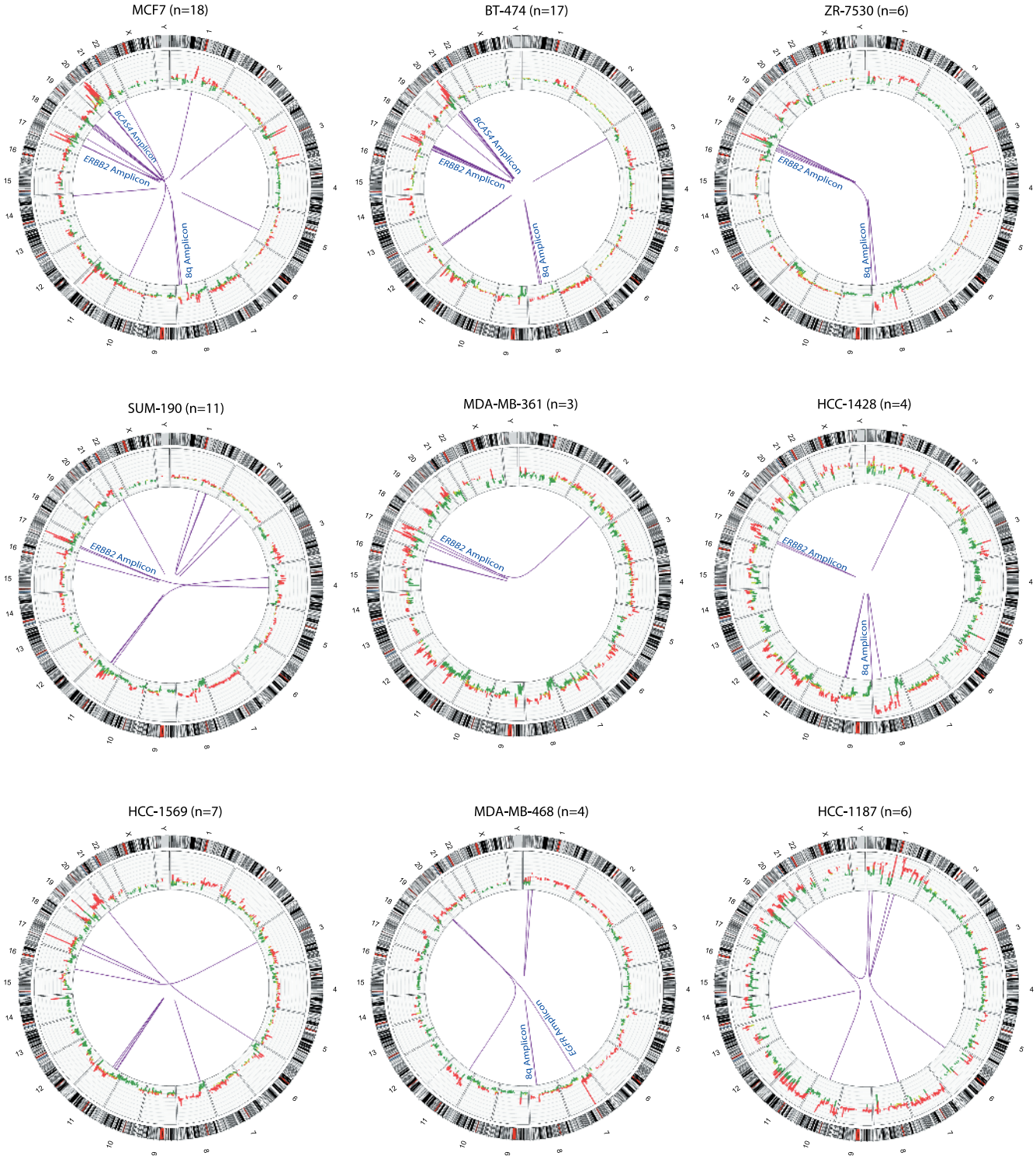


Figure W2. Graphical representation of integrative analysis of gene fusions with copy number analysis. Circos plots of the genome-wide distribution of gene fusions along with status of copy number alterations. Red and green peaks represent amplifications and deletions; purple line represents the fusions associated with amplicons and nonamplicons, respectively. "*n*" refers to the total number of fusions identified.

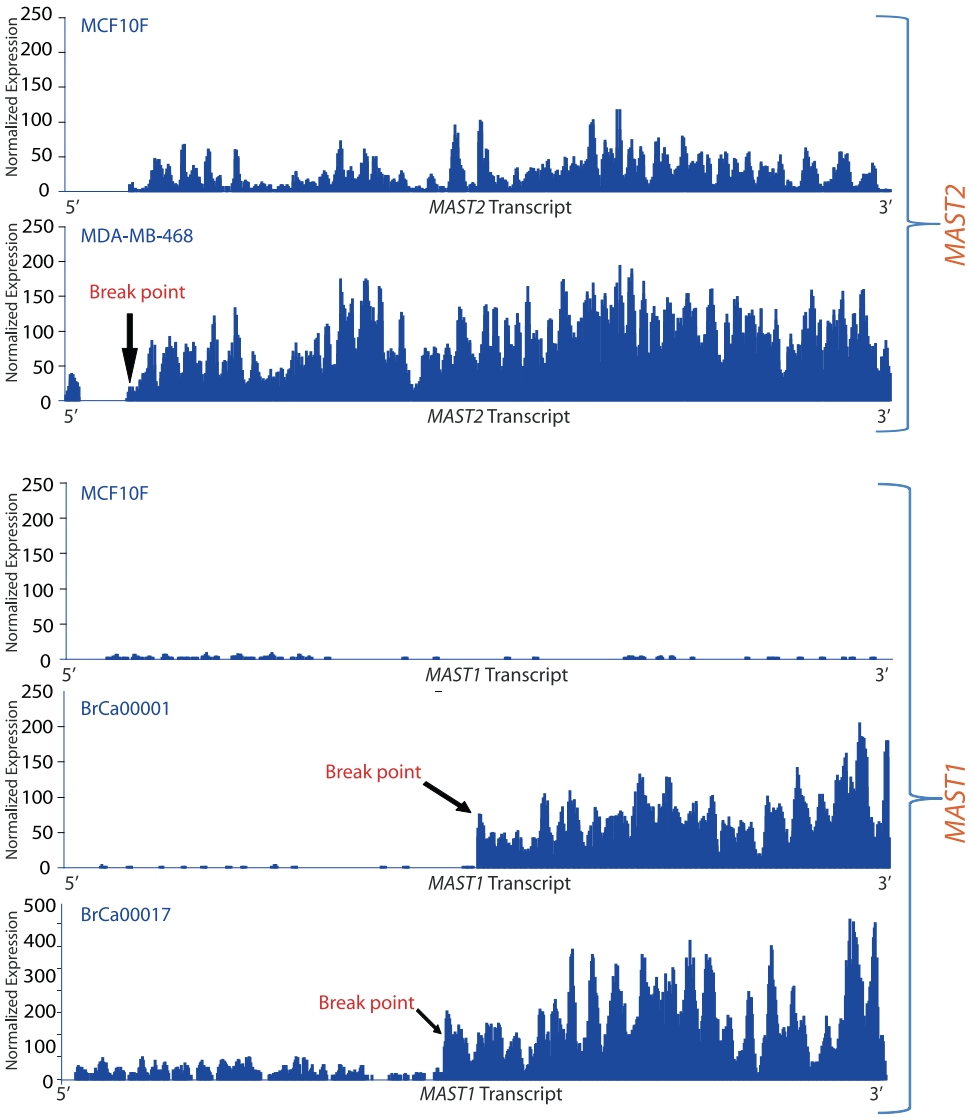


Figure W3. Plot of normalized coverage of *MAST1* and *MAST2* transcripts in *MAST* fusion-positive samples (breakpoint indicated by arrow).

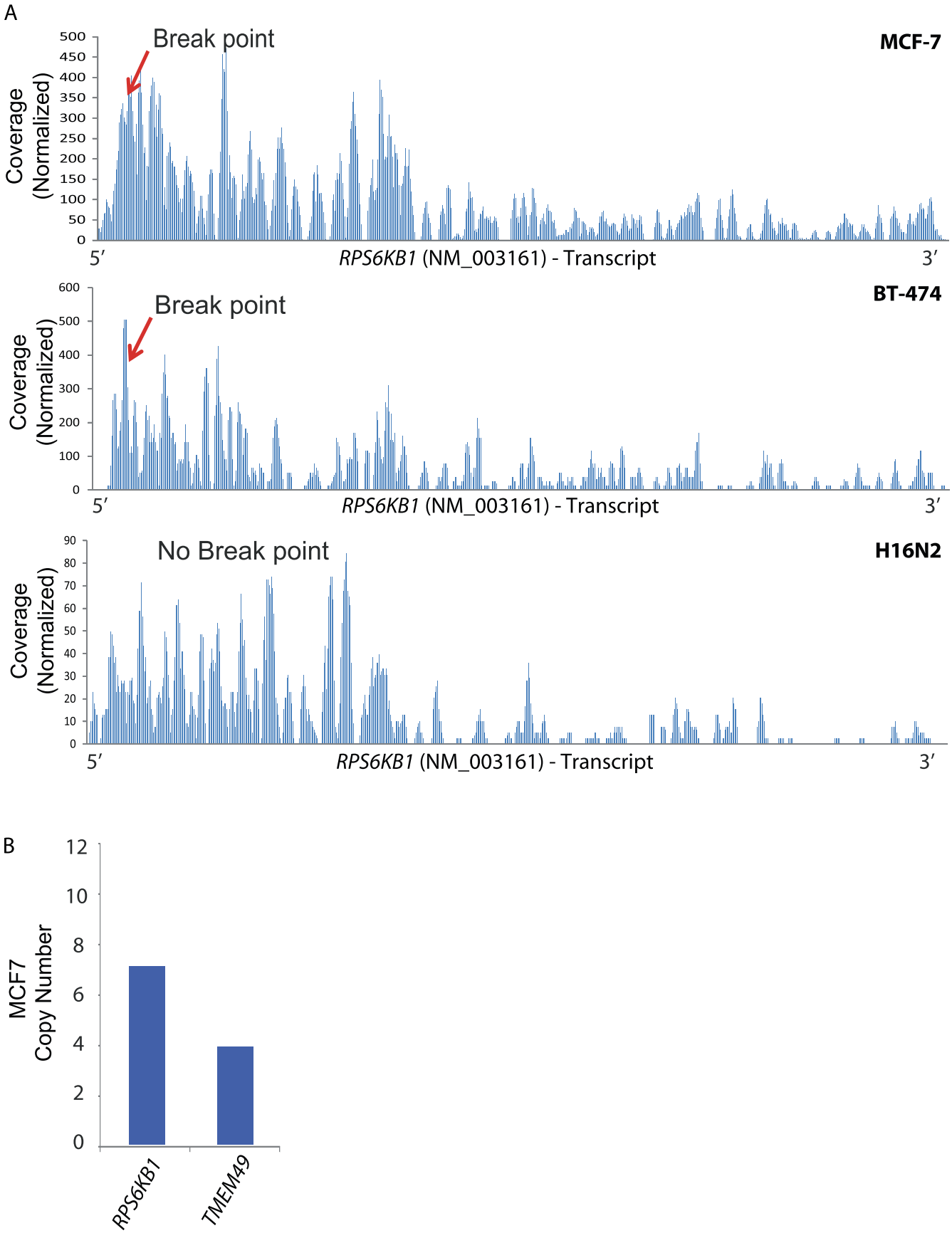


Figure W4. (A) Plot of normalized coverage of *RPS6KB1* transcript in BT-474, MCF7, and H16N2 cell lines. (B) Bar graph representing the copy number of *RPS6KB1* and *TMEM49* in MCF7.