DOI:10.1158/1541-7786.MCR-08-0107

Molecular Cancer Research

Genetic Alterations and Oncogenic Pathways Associated with Breast Cancer Subtypes

Xiaolan Hu, Howard M. Stern, Lin Ge, et al.

Mol Cancer Res 2009;7:511-522. Published online April 16, 2009.

DOI:10.1158/1541-7786.MCR-08-0107

Genetic Alterations and Oncogenic Pathways Associated with Breast Cancer Subtypes

Xiaolan Hu,[1] Howard M. Stern,[4] Lin Ge,[3] Carol O'Brien,[2] Lauren Haydu,[2] Cynthia D. Honchell,[1] Peter M. Haverty,[1] Brock A. Peters,[3] Thomas D. Wu,[1] Lukas C. Amler,[2] John Chant,[3] David Stokoe,[3] Mark R. Lackner,[2] and Guy Cavet[1]

Departments of [1]Bioinformatics, [2]Oncology Diagnostics, [3]Molecular Biology, and [4]Research Pathology, Genentech, Inc., South San Francisco, California

Abstract

Breast cancers can be divided into subtypes with important implications for prognosis and treatment. We set out to characterize the genetic alterations observed in different breast cancer subtypes and to identify specific candidate genes and pathways associated with subtype biology. mRNA expression levels of estrogen receptor, progesterone receptor, and HER2 were shown to predict marker status determined by immunohistochemistry and to be effective at assigning samples to subtypes. HER2$^+$ cancers were shown to have the greatest frequency of high-level amplification (independent of the *ERBB2* amplicon itself), but triple-negative cancers had the highest overall frequencies of copy gain. Triple-negative cancers also were shown to have more frequent loss of phosphatase and tensin homologue and mutation of *RB1*, which may contribute to genomic instability. We identified and validated seven regions of copy number alteration associated with different subtypes, and used integrative bioinformatics analysis to identify candidate oncogenes and tumor suppressors, including *ERBB2, GRB7, MYST2, PPM1D, CCND1, HDAC2, FOXA1*, and *RASA1*. We tested the candidate oncogene *MYST2* and showed that it enhances the anchorage-independent growth of breast cancer cells. The genome-wide and region-specific differences between subtypes suggest the differential activation of oncogenic pathways. (Mol Cancer Res 2009;7(4):511–22)

Introduction

Identification of the genetic and epigenetic abnormalities that contribute to cancer development can lead to new therapeutic targets and strategies. Technological advances are beginning to enable researchers to profile cancers comprehensively for these abnormalities and to elucidate their functional significance. At the same time, there is an increasing understanding of the importance of cancer subtypes for prognosis and treatment. Cancer subtypes are distinct forms of a disease that have different origins, molecular characteristics, or clinical features. They often also respond differently to therapy, so higher-resolution definitions of subtypes and improved understanding of their biological differences are important for improving cancer treatment.

Breast cancer has relatively well-established and studied subtypes. Newly diagnosed cancers are routinely assessed for the expression of estrogen receptor (ER) or progesterone receptor (PR) and for overexpression or amplification of the *ERBB2* gene (encoding the HER2 protein). In general, patients with *ERBB2*/HER2–positive (HER2$^+$) tumors are candidates for treatment with the anti-HER2 monoclonal antibody trastuzumab (1). Patients with ER- or PR-positive tumors are candidates for treatment with hormonal therapy, including selective ER modulators such as tamoxifen for premenopausal women or aromatase inhibitors for postmenopausal women (2-4). Patients whose tumors are neither HER2$^+$ nor ER$^+$/PR$^+$ (so-called triple-negative tumors) currently have no targeted therapy available and have relatively poor prognosis (5-7).

In recent years, the application of molecular profiling technologies, in particular DNA microarrays for measuring mRNA expression, has added to our knowledge of breast cancer subtypes. Hierarchical clustering with gene expression data has been shown to organize tumor samples into groups with distinct gene expression patterns and significantly different clinical outcomes (8-10). The three major subgroups defined by gene expression profiling are similar to those defined in clinical practice: a HER2$^+$ group, a luminal group (which is predominantly ER$^+$/PR$^+$), and a basal-like group (which predominantly lacks HER2 overexpression, ER, and PR). The terminal duct-lobular unit of the breast, the structure from which the majority of breast cancers arise, is composed of two types of epithelial cells. The inner or luminal cells are potential milk-secreting cells and are surrounded by an outer basal layer of contractile myoepithelial cells. Basal-like breast tumors are so called because they express genes, such as cytokeratins 5, 6, and 17, that are typically expressed in normal myoepithelial cells, whereas luminal tumors are characterized by expression of more epithelial-like genes such as E-cadherin and cytokeratins 8 and 18 (7-13). Clustering also suggests higher-resolution subtypes with biologically interesting features. For example, several studies have reported

Received 2/22/08; revised 11/21/08; accepted 12/15/08; published online 4/16/09.
Grant support: Genentech, Inc.
The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
Note: Supplementary data for this article are available at Molecular Cancer Research Online (http://mcr.aacrjournals.org/).
Current address for X. Hu: Pfizer Global Research and Development, Groton, Connecticut. Current address for L. Haydu: Faculty of Medicine, The University of Sydney, Sydney, Australia.
Requests for reprints: Guy Cavet, Crescendo Bioscience, 400 East Jamie Court, South San Francisco, CA 94010. Phone: 650-351-1602; Fax: 650-592-3664. E-mail: glcavet@gmail.com

doi:10.1158/1541-7786.MCR-08-0107

DOI:10.1158/1541-7786.MCR-08-0107

subdivisions of the luminal type dubbed luminal A and luminal B (the latter has lower levels of ER transcripts and less favorable clinical outcomes). The presence of other subtypes varies more from study to study, presumably due to methodologic differences, variation between sample sets, and the instability of hierarchical clustering results. A statistically stringent meta-analysis of data from multiple gene expression studies identified three robust subtypes that closely resemble the $HER2^+$, ER^+/PR^+, and triple-negative types defined in clinical practice (14).

In this study, we explore the hypothesis that breast cancers of different subtypes are both characterized and caused by different sets of genetic lesions. Beyond the associations of HER2 and hormone receptor status with the $HER2^+$ and luminal subtypes, there are indications that distinct oncogenic alterations and pathways are associated with different breast cancer subtypes. For example, epidermal growth factor receptor overexpression and phosphatase and tensin homologue (PTEN) loss are most common in basal-like cancers (6, 7, 15-19). Amplification of *CCND1* and activating mutation of *PIK3CA* are associated with ER and PR positivity (17, 20). In addition, the availability of genome-wide copy number profiling has made it possible to identify region-specific copy number alterations associated with clinical characteristics such as hormone receptor status, gene mutation status, disease subtype, and clinical outcome (21-25). We use high-resolution genome-wide copy number profiling and cancer gene resequencing to identify copy number alterations and somatic mutations that are associated with breast cancer subtypes in both tumors and cancer cell lines. We also characterize subtype differences in overall levels of genome alteration. We use bioinformatics methods to prioritize candidate causative genes in subtype-associated regions of copy number gain, and show that these methods can help to identify oncogenes. Finally, we relate our results to distinct patterns of oncogenic pathway activation in different subtypes.

Results

ER, PR, and HER2 mRNA Expression Levels Predict Breast Cancer Subtypes

We characterized genetic alterations in 51 breast tumors and 46 breast cancer cell lines. We set out to study the three major subtypes defined by current clinical practice: Samples with amplified or overexpressed HER2 were assigned to the $HER2^+$ subtype; HER2 negative samples expressing ER or PR were assigned to the ER^+/PR^+ subtype; and other samples were assigned to the triple-negative subtype. These subtypes are determined using a simple algorithm with a clear biological basis, and their relevance to prognosis and treatment is well established.

ER, PR, and HER2 status is generally determined in clinical practice by immunohistochemistry or fluorescence *in situ* hybridization, but results from these approaches were available for only a fraction of the tumor samples in this study (see Materials and Methods and Supplementary Table S1). We therefore developed logistic regression classifiers for ER and PR status using microarray data for the *ESR1* and *PGR* genes. The microarray data for these genes predicted pathologist-determined marker status very well (Supplementary Fig. S1) with error rates in the training data set of 12 of 143 for ER and 32 of 142 for PR. To infer HER2 status, we identified samples with copy gains at the ERBB2 locus or mRNA overexpression (see Materials and Methods and Supplementary Fig. S1). These procedures were used to assign 51 breast tumors to three subtypes (8 $HER2^+$, 26 ER^+/PR^+, 17 triple negative). Assigning the tumors to the five main subtypes defined by gene expression data (26) gave similar results (Supplementary Table S1). None of our samples were assigned to the normal-like gene expression subtype and only one was assigned to the luminal B subtype. Our sample set is therefore insufficient to investigate these additional subtypes.

Application of the mRNA- and copy number–based classifiers to breast cancer cell lines assigned 15 lines as $HER2^+$, 12

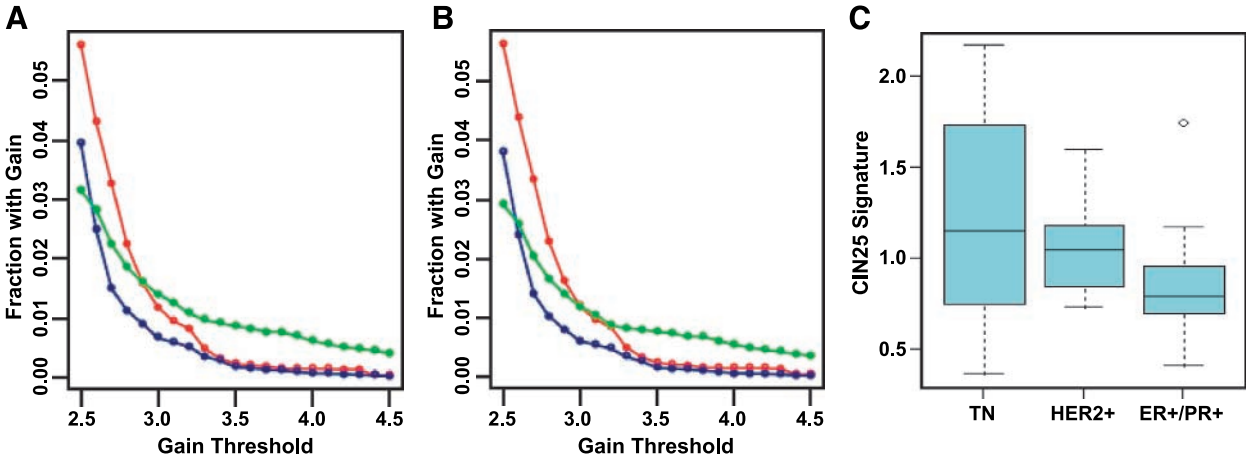

FIGURE 1. Associations between subtypes and genome-wide frequencies of genetic alterations. **A** and **B.** The mean fraction of the genome exhibiting gains, including and excluding chromosome 17, respectively, as different thresholds for copy gain are considered in triple-negative (*red*), $HER2^+$ (*green*), and ER^+/PR^+ (*blue*) samples. **C.** Box plot of the chromosomal instability signature (CIN25) by subtype. Carter et al. (28) estimated the aneuploidy of individual cancer samples by quantifying the extent to which genes in the same chromosomal region have coordinated mRNA expression levels. The CIN25 signature is made up of genes whose expression is correlated with this measure of aneuploidy. High CIN25 values are associated with poor outcome in several tumor types.

DOI:10.1158/1541-7786.MCR-08-0107

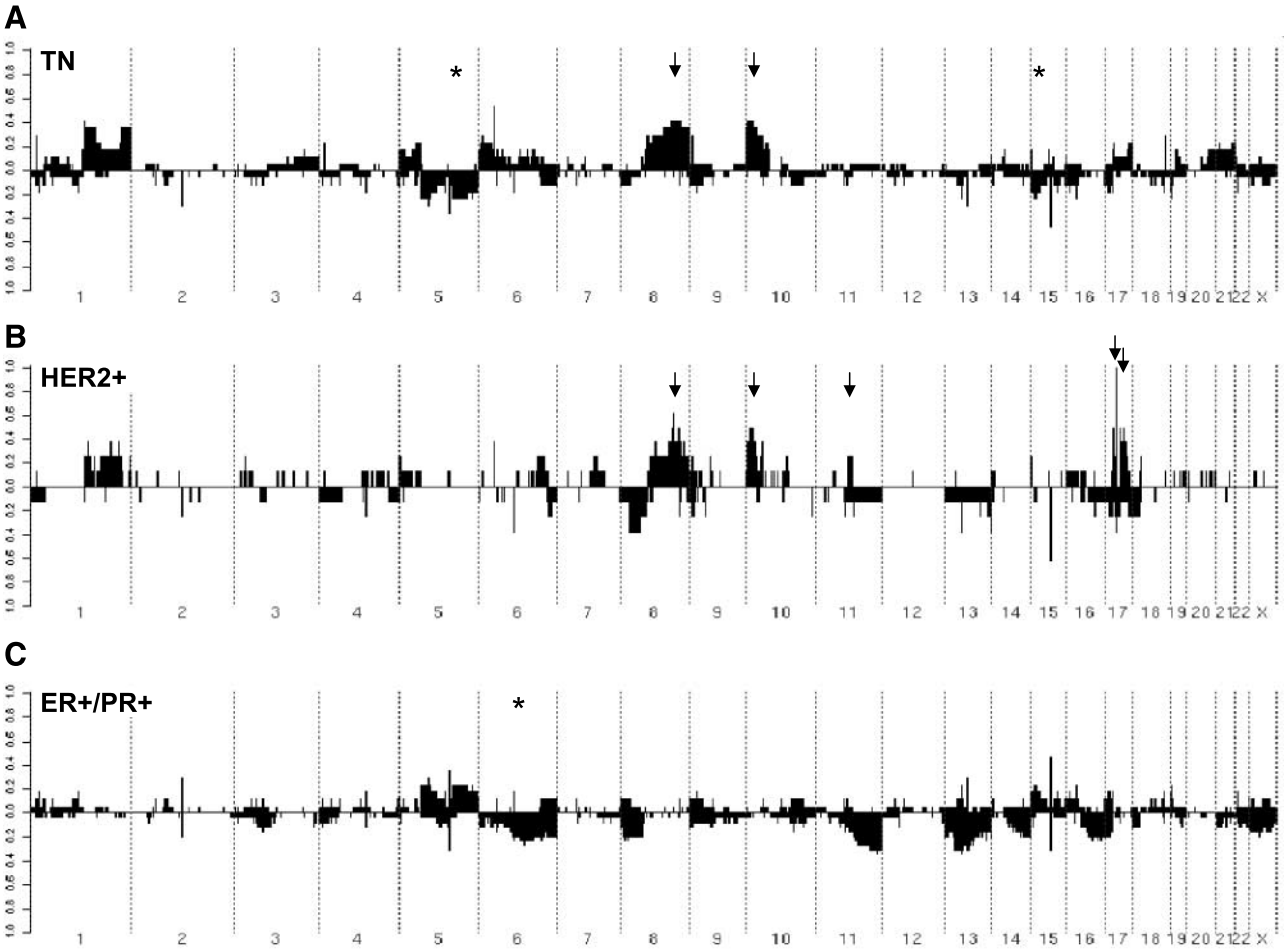

FIGURE 2. Genome-wide gains and losses of breast tumors by subtype. Graphs show gains above the *x* axis (frequency times magnitude) and losses below the *x* axis (frequency) in triple negative (**A**), HER2$^+$ (**B**), and ER$^+$/PR$^+$ (**C**) subtypes. Arrows, regions of subtype-associated gains; asterisks, regions of subtype-associated losses. These regions were selected based on the tumor data only, without regard to replication in cell lines.

as ER$^+$/PR$^+$, and 19 as triple negative. Our assignments are in good agreement with the recent study of Neve et al. (27), which classified all of our triple-negative cell lines as basal-like and all but one of our ER$^+$/PR$^+$ cell lines as luminal. The results from both breast tumors and cell lines indicated that the expression levels of *ESR1* and *PGR* are good surrogates for ER and PR status.

HER2$^+$ and Triple-Negative Breast Tumors Exhibit Higher Copy Number Instability

To understand whether overall genomic instability differs between breast cancer subtypes, we first examined the overall frequencies of copy number alterations in tumors (Fig. 1A). The fractions of the genome exhibiting copy gains differ significantly between the subtypes ($P < 0.05$, when any threshold for copy number gain ≥2.8 copies is considered). Triple-negative tumors have the highest frequencies of modest gains. However, HER2$^+$ tumors have the highest frequencies of high-level gains, which include focal amplifications. This trend is not explained simply by the *ERBB2* amplicon on chromosome 17 because the same differences between the subtypes are observed when chromosome 17 is excluded from consideration (Fig. 1B). The same subtype differences were observed in cell lines (data not shown). We did not find a statistically significant difference in frequencies of copy number loss between the subtypes.

To further characterize the differences in genomic instability among the tumor subtypes, we used a gene expression signature that reflects chromosomal instability and is associated with poor outcome in several cancer types (28). We tested whether the signature is the same in all three subtypes. We found that triple-negative and HER2$^+$ tumors have higher expression of the instability signature than ER$^+$/PR$^+$ tumors ($P = 0.005$; Fig. 1C). This is consistent with the subtype differences in frequencies of copy number gains and also with the worse prognosis associated with HER2$^+$ and triple-negative cancers (8-10).

Copy Number Changes Associated with Breast Cancer Subtypes

In addition to global patterns of copy number alteration, we set out to identify specific regions of copy number alteration associated with subtypes and the functionally important genes

DOI:10.1158/1541-7786.MCR-08-0107

Table 1. Regions of Copy Number Alteration Associated with Subtype in Two Independent Sample Sets

Major Alteration	Type	Location	Coordinates	Tumors		Cell Lines	Candidate Genes
				P	*q*	*P*	
LOSS	TN, $HER2^+$	5q13-14	71990072-87561269	2.10e−04	0.011	8.12e−03	*RASA1*
LOSS	ER^+/PR^+	6q22	115073292-116966464	1.02e−04	0.011	3.97e−02	*HDAC2, NCOA7, ARHGAP18*
GAIN	ER^+/PR^+, HER2	11q13	69105561-70039903	6.99e−03	0.067	2.11e−02	*ORAOV1, CCND1, FADD, PPFIA1*
GAIN	ER^+/PR^+	14q13	35518338-38046131	2.27e−03	0.042	8.50e−03	*FOXA1*
GAIN	$HER2^+$	17q12	34786442-35236409	7.90e−06	0.014	2.23e−04	*HER2, GRB7*
GAIN	ER^+/PR^+, HER2	17q21	44402113-46074589	4.59e−03	0.063	1.05e−03	*MYST2*
GAIN	ER^+/PR^+, HER2	17q23	55155170-56050216	1.03e−02	0.070	8.39e−05	*RPS6KB1, PPM1D, TMEM49*

NOTE: Genomic intervals consistently associated with subtype in both tumors and cell lines were identified. Physically adjacent intervals with the same pattern of subtype association were merged together to form the seven regions reported. Region boundaries are from tumor data.
Abbreviation: TN, triple negative.

within those regions. Figure 2 shows the frequencies of gains and losses for the different subtypes as measured in tumors with Affymetrix 500k single-nucleotide polymorphism (SNP) arrays. We observed some differences in frequencies between subtypes similar to those described in recent studies (e.g., losses on 5q in triple negative, gains on 10p in triple negative and $HER2^+$, gains on 11q13 in ER^+/PR^+; refs. 22-24). We also observed some additional differences (e.g., losses on 6q in ER^+/PR^+, gains on 17q21 and 17q23 in ER^+/PR^+ and $HER2^+$, losses on 15p in triple negative).

Different frequencies of copy number alteration within a region may occur by chance and may be characteristic of one sample set but not observed in others. We sought to identify regions that were both statistically significantly associated with subtypes and observed in two independent sample sets. We first identified chromosomal regions with recurrent gains or losses (>2.5 copies or <1.6 copies in at least 10% of tumor samples) with differences in gain or loss frequency between subtypes in tumors. We then validated these regions in data from an independent set of 47 breast cancer cell lines. A total of seven chromosomal regions showed consistent association in both sample sets (Table 1; Fig. 3).

Candidate Driver Genes for Subtype-Associated Regions

A recurrent region of copy number alteration typically extends over multiple genes. Identifying the driver genes (genes functionally involved in tumorigenesis) with confidence generally requires extensive experimentation. We therefore developed a computational strategy to identify candidate driver genes for experimental follow-up based on three hypotheses: first, that driver genes are found in the part of a region of recurrent copy number alteration that is most frequently and most significantly gained or lost; second, that driver genes

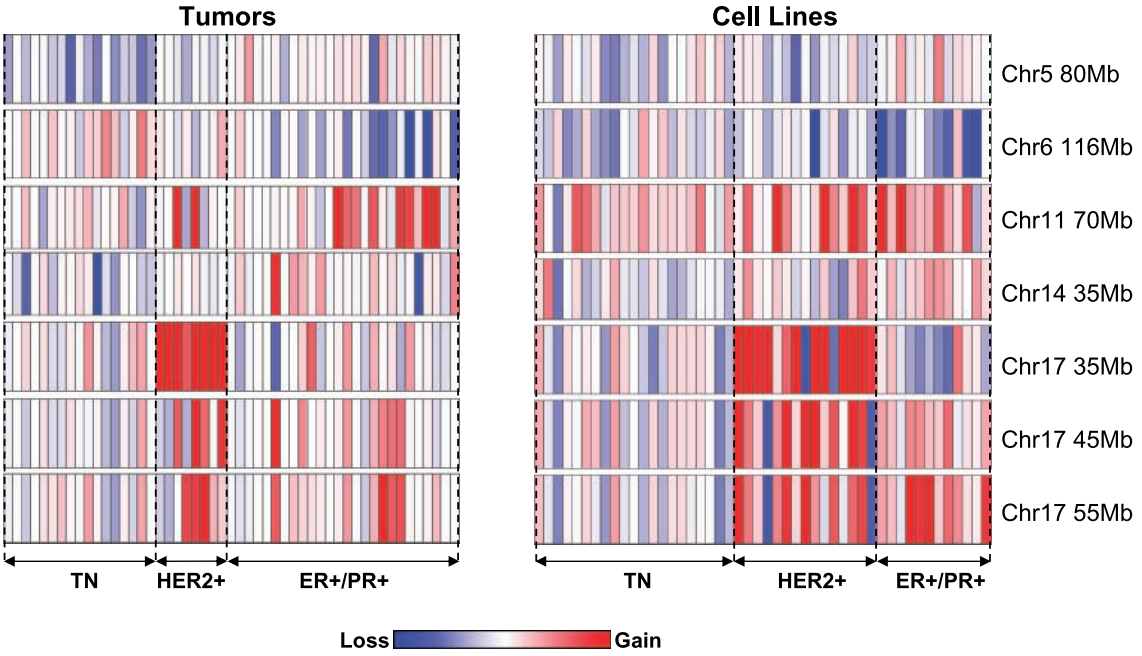

FIGURE 3. Copy number in regions associated with subtype. Blue, losses; red, gains. The genomic intervals consistently associated with subtype in both tumors and cell lines were identified. The physically adjacent intervals with the same pattern of subtype association were merged together to form the seven regions shown. The average copy number of the merged intervals is indicated by color: blue, losses; red, gains.

DOI:10.1158/1541-7786.MCR-08-0107

undergo marked alterations in gene expression as a consequence of copy number change; and third, that changes in expression of driver genes may be associated with poor clinical outcome. Not all of these three hypotheses may hold for every driver gene, but combining the evidence from the three can identify candidates.

In the HER2 amplicon, for example, *ERBB2* is found at the most amplified point, is overexpressed on amplification, and is associated with increased occurrence of distant metastasis in the study of van't Veer et al. (ref. 29; Fig. 4A). One other gene in the amplicon, *GRB7*, also meets these criteria. This may simply be because of its proximity to *ERBB2*, but RNA interference studies suggest that *GRB7* may also contribute to proliferation of breast cancer cells (30).

Applying the same procedure to other regions of copy gain identifies other candidate oncogenes. In the amplicon around 45 Mbp on chromosome 17q21 (Fig. 4B), which is associated with the ER^+/PR^+ and $HER2^+$ subtypes, there is a single gene that fulfills all three criteria: *MYST2*, which encodes a histone acetyltransferase (HBO1) that has been implicated in the regulation of DNA synthesis and in steroid hormone receptor signaling (31-34). In the amplicon around 55 Mbp (17q23), *PPM1D* is the only gene meeting all three criteria, although it is not strongly supported by any individual line of evidence. The genes *TMEM49* and *RPS6KB1* (encoding p70 ribosomal S6 kinase) are also located under the peak of this amplification, and their expression levels are strongly associated with copy number alteration in the region (Fig. 4C). The peak of amplification around 70 Mbp on chromosome 11q13, also associated with ER^+/PR^+ and $HER2^+$ subtypes, contains no genes associated with recurrence (Fig. 4D). Driver gene status is often attributed to *CCND1* (encoding cyclin D1), which has significantly increased expression in amplified samples, along with the neighboring genes *ORAOV1, FADD*, and *PPFIA1*. In the region around 35 Mbp on chromosome 14q13, which is associated with the ER^+/PR^+ subtype, only *FOXA1* has elevated expression in tumors with increased copy number (Supplementary Fig. S2A).

For regions of deletion/loss, the amplitude or the frequency of the alteration does not provide as much information as in regions of copy gain because there are at most two chromosomes to be lost and deletions are usually broad. The identification of candidates therefore relies primarily on the association of gene expression with copy number loss and clinical outcome (Supplementary Fig. S2B and C). Both these lines of evidence support the candidates *ARHGAP18, HDAC2*, and *NCOA7* in deletions around 116 Mbp on chromosome 6q (associated with the ER^+/PR^+ subtype). The deletion located around 80 Mb on chromosome 5q13-14 (associated with $HER2^+$ and triple-negative subtypes) contains a single gene supported by two lines of evidence: *RASA1*, which encodes the RAS GTPase activating protein p120GAP.

MYST2 *Drives the Growth of Breast Cancer Cells*

We selected the candidate driver gene *MYST2* (also known as HBO1) for further study because it has particularly strong supporting evidence in our analysis. *MYST2* is required for growth in 293T cells (31) but is not an established oncogene. Overexpression of *MYST2* (Supplementary Fig. S4) dramatically enhanced the anchorage-independent growth of both MCF7 (Fig. 5A and B) and SKBR3 breast cancer cells (Fig. 5C and D). Previous studies in MCF7 cells, which have a modest copy number gain in this region, have also shown that siRNA-mediated knockdown of *MYST2* substantially impairs cell proliferation and blocks progression through the S phase of the cell cycle compared with control siRNA treatment (31). Taken together, the observations that *MYST2* knockdown blocks proliferation and that *MYST2* overexpression can shift cell lines to a more transformed state support the hypothesis that it is the oncogene driving amplification around 45 Mbp on chromosome 17.

Association of PTEN Loss and Somatic Mutations with Breast Cancer Subtypes

We sequenced a panel of genes with well-characterized cancer-causing mutations in the cell line collection, and we evaluated both the mutations we found and published mutation data for association with breast cancer subtypes (Table 2). We observed the following overall mutation frequencies: 73% (*TP53*), 34% (*PIK3CA*), 11%(*RB1*), 9% (*PTEN*), 7% (*CDKN2A*), 5% (*BRAF*), 2% (*KRAS*), and 2% (*HRAS*). These are generally very similar to those reported in the COSMIC database (35). TP53, however, has a lower mutation rate in COSMIC (54%), so our cell line collection may have different characteristics from the 80 breast carcinoma samples that currently have data in COSMIC. We did not see evidence that mutations in *TP53, PTEN*, or *CDKN2A* are associated with specific breast cancer subtypes. *PIK3CA* mutations were enriched in the ER^+/PR^+ and $HER2^+$ subtypes, which is consistent with previous reports (17), although this enrichment was not statistically significant.

All four mutations in *BRAF, KRAS*, and *HRAS* were in triple-negative samples. This suggests an association between mutations in the RAS/RAF/mitogen-activated protein kinase/extracellular signal-regulated kinase kinase/extracellular signal–regulated kinase pathway and the triple-negative subtype ($P = 0.05$). An additional triple-negative cell line (CAL-51) has also been reported to harbor an activating oncogenic mutation in a RAS family member (36, 37).

Loss of PTEN has previously been reported to be associated with ER and PR negative status (15, 17, 38). Mutation and deletion at the *PTEN* locus are imperfect surrogates for loss of PTEN protein (15, 27), so we evaluated the association between PTEN and subtypes in cell lines by Western blot. We observed PTEN loss in 59% (10 of 17) of triple-negative samples but only in 17% (2 of 12) of ER^+/PR^+ samples and 8% (1 of 13) of $HER2^+$ samples ($P = 0.002$). The chromosomal instability signature was higher in samples with PTEN loss than in those without, although this difference was not statistically significant ($P > 0.05$; data not shown).

All four mutations in *RB1* were found in triple-negative samples, and we found that loss of retinoblastoma protein (RB) function is associated with the triple-negative subtype ($P = 0.006$). The inactivation of RB was further examined in the 51 breast tumor samples using a 59-gene expression signature reflecting RB dysregulation (39, 40). This signature was significantly higher in the triple-negative subtype than in $HER2^+$ and ER^+/PR^+ subtypes ($P = 0.002$, Fig. 6). In addition,

DOI:10.1158/1541-7786.MCR-08-0107

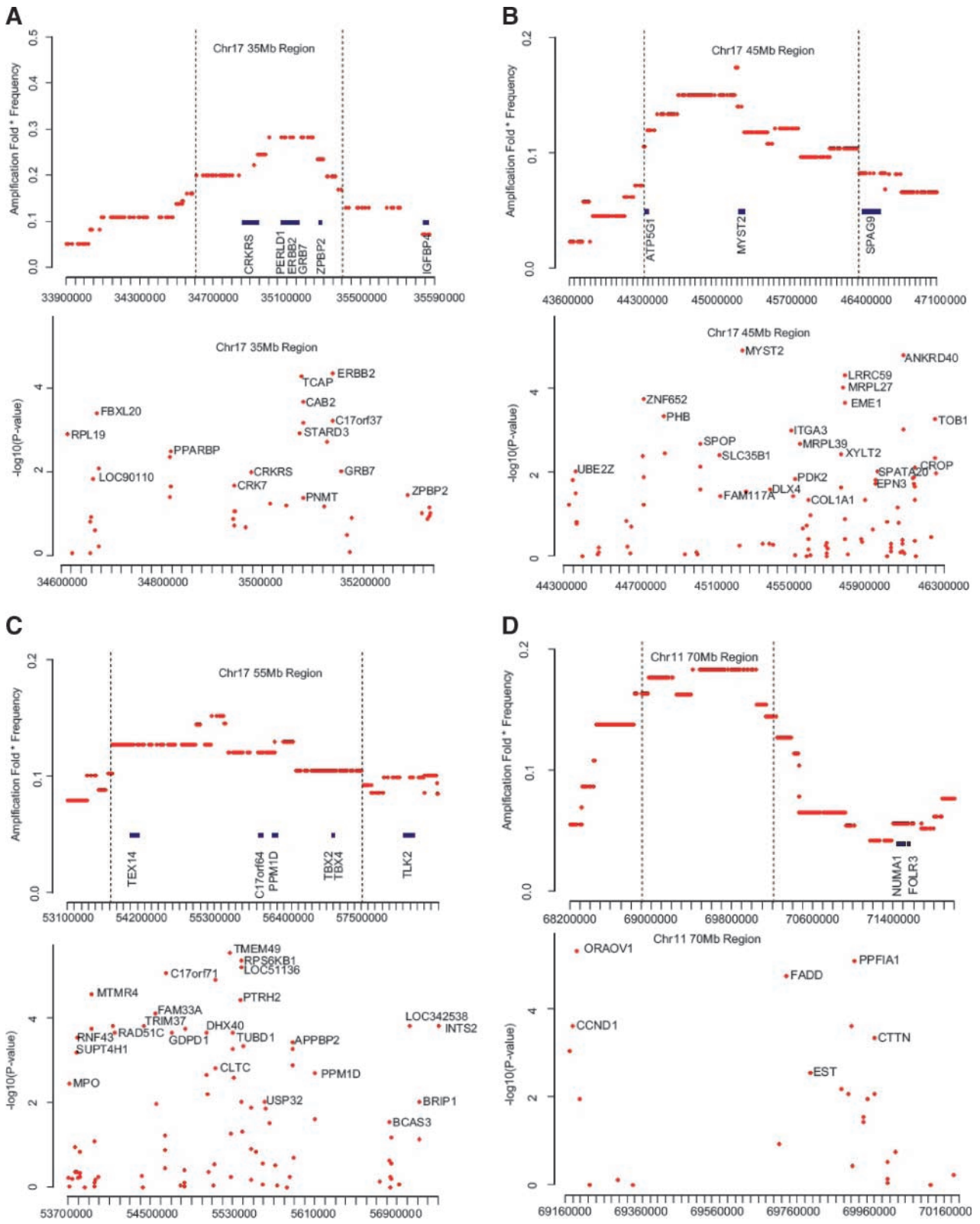

FIGURE 4. Identification of candidate driver genes. **A.** Chromosome 17 35-Mbp region. **B.** Chromosome 17 45-Mb region. **C.** Chromosome 17 55 Mb region. **D.** Chromosome 11 70 Mb region. Top, summarized copy gain (*red*) and the locations of genes associated with distant metastasis at $P < 0.05$ (*blue*). Dotted lines define the focal region within which copy number was associated with subtype and within which driver genes were sought. Bottom, $-\log_{10}$ transformed *P* values reflecting increased gene expression on amplification. Genes passing the *P* value threshold (0.05) are labeled.

DOI:10.1158/1541-7786.MCR-08-0107

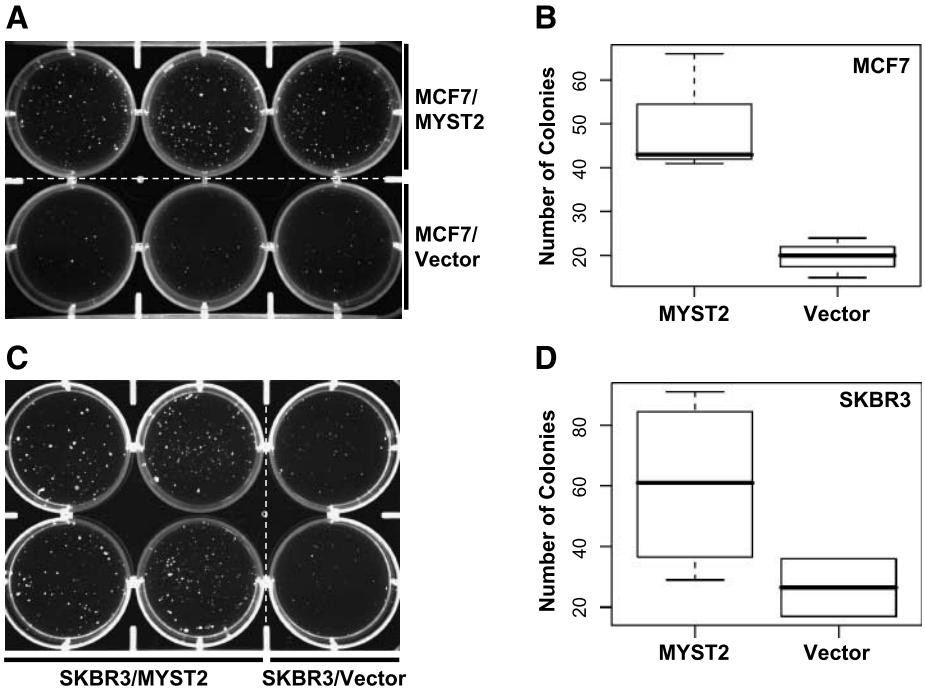

FIGURE 5. *MYST2* overexpression enhances colony formation in soft agar. **A.** MCF7 colonies. Top row, cells transfected with MYST2 expression vector; bottom row, cells transfected with control vector. **C.** SKBR3 colonies. Left and center columns, cells transfected with MYST2 expression vector; right column, cells transfected with control vector. **B** and **D.** Box plots representing the numbers of colonies in MCF7 and SKBR3, respectively.

the chromosomal instability signature was higher in samples with RB1 mutation than in those without ($P = 0.01$; data not shown).

Discussion

In this study, we have characterized genetic alterations associated with the three major and most reproducible breast cancer subtypes, defined by tumor ER, PR, and HER2 status. The widespread adoption of consistent and robust analysis methods and larger sample sets should make it possible to carry out a similar analysis of higher-resolution subtypes such as luminal A and luminal B.

In clinical practice, ER and PR status is generally established by immunohistochemistry, and HER2 status by immunohistochemistry or fluorescence *in situ* hybridization. We found that levels of *ESR1* and *PGR* mRNAs, even when measured by microarray, can accurately predict immunohistochemistry status and be used in determining subtype. Similar conclusions were reached for HER2 and ER in a recent study (41), although the authors did not examine PR.

In this study, we used both tumor and cell line samples. Analysis of copy number was carried out primarily in tumors, with cell lines used for confirmation. Point mutations were studied in cell lines. The copy number alterations observed in cell lines have been shown to be very similar to those observed in tumors (27), but the differences that do exist have implications for our study. To select copy number alterations that are reproducibly and robustly associated with breast cancer subtypes, we required that regions discovered in tumors also be associated with subtypes in cell lines. This protects us from finding alterations induced by cell culture, but it almost certainly contributed to our false negative rate. Some of the most significant findings in tumors were not reproduced in cell lines, for example, the association of gains around 7 Mbp on 10p14-15.1 in the triple-negative and $HER2^+$ subtypes. This may be due in part to some copy number alterations being selected against *in vitro* and in part to lack of statistical power. Another caveat is that cell lines believed to represent breast cancer may actually be derived from other cancers, as in the case of MDA-MB-435, which represents melanoma (42, 43) and which was not used in this study. We note that the mutations in *BRAF* and *APC* (44) observed in DU4475 are more typical of colorectal than breast cancer, and the gene expression profile of this line is not very similar to other cancers of either the large intestine or breast (data not shown).

In the seven regions of copy number alteration that we found to be reproducibly associated with subtype, we used three bioinformatics-based criteria to prioritize candidate driver genes. Such approaches can only generate hypotheses requiring further studies, but the results on regions with known driver genes are encouraging. First, we assumed that the driver gene would be found in the most recurrently and profoundly altered location in a region of alteration. It may be that local features such as repeat sequences or chromatin structure bias the exact boundaries of these regions. However, our data and other studies suggest that these boundaries vary widely between individual tumors (45-49) so that, over an adequate number of samples, the location with most copy number alteration can be used to pinpoint candidate driver genes. Our second assumption, that the expression of a driver gene should be significantly affected by copy number, is limited primarily by the finding that alterations in copy number affect the expression of a significant fraction of genes. The final assumption that the expression of a driver gene should be associated with variation in clinical outcome may be more useful for identifying very

DOI:10.1158/1541-7786.MCR-08-0107

strong candidates than for excluding genes from consideration. The driver gene by definition is (or has been) beneficial to the tumor, but this benefit may not always translate into outcome, particularly because outcome may be evaluated in different ways. However, when alteration in gene expression is found to be associated with poor outcome, this provides additional reason to prioritize a gene for follow-up.

Further studies may also be informed by other features of the candidate driver genes and regions we identified. Several candidates have links to steroid hormone receptor signaling, which may contribute to the association of some regions of gain with the ER^+/PR^+ subtype. However, this does not offer a complete explanation because these gains are sometimes observed in ER- and PR-negative cancers. *FOXA1* is the only gene significantly up-regulated upon the amplification on 14q20 associated with the ER^+/PR^+ subtype. It encodes a forkhead transcription factor, which is required for ER-mediated transcriptional activation of many transcriptional targets (50, 51). Within the amplicon on 17q23 (associated with ER^+/PR^+ and $HER2^+$ cancers) there are several genes that have evidence of oncogenic activity (52-55). The only gene to meet all three criteria in our prioritization is *PPM1D*. PPM1D activity has been shown to not only lead to inactivation of mitogen-activated protein kinase and p53 but also to stimulate ER and PR activity (53), again suggesting a relationship with hormone receptor–dependent cancers. However, the 17q23 amplification also includes other candidate protein-coding oncogenes and mir-21, a microRNA gene with oncogenic properties known to be

Table 2. Amino Acid Mutations and PTEN Protein Loss in Breast Cancer Cell Lines

Cell Line	Subtype	PTEN Protein	PTEN Mutation	PIK3CA	RB1	BRAF	KRAS	HRAS	CDKN2A	TP53
BT549	TN	PTEN (−)	WT	WT	S82del	WT	WT	WT	WT	R249S
CAL-120	TN	WT	WT	WT	WT	WT	WT	WT	WT	WT
CAL-148	TN	PTEN (−)	I252N, Q297X	H1047R	WT	WT	WT	WT	G124E	E224K
CAL-51	TN	PTEN (−)	WT	E542K	WT	WT	WT	WT	WT	WT
CAL-85-1	TN	WT	WT	WT	21fs	WT	WT	WT	WT	K132E
DU4475	TN	PTEN (−)	WT	WT	WT	V600E	WT	WT	WT	WT
HCC1143	TN	WT	WT	WT	WT	WT	WT	WT	WT	WT
HCC1395	TN	PTEN (−)	WT	WT	WT	WT	WT	WT	WT	R175H
HCC1806	TN	ND	ND	ND	ND	ND	ND	ND	ND	ND
HCC1937	TN	PTEN (−)	WT	WT	K715X	WT	WT	WT	WT	R306X
HCC38	TN	PTEN (−)	WT	WT	WT	WT	WT	WT	ND	R273L
HCC70	TN	PTEN (−)	WT	WT	N480del	WT	WT	WT	WT	WT
HDQ-P1	TN	WT	WT	WT	WT	WT	WT	WT	WT	R213X
HS578T	TN	WT	WT	WT	WT	WT	WT	G12D	WT	V157F
MDA-MB-231	TN	WT	WT	WT	WT	G464V	G13D	WT	WT	R280K
MDA-MB-436	TN	PTEN (−)	WT	WT	WT	WT	WT	WT	WT	202fs
MDA-MB-468	TN	PTEN (−)	WT	WT	R876C	WT	WT	WT	WT	R273H
MFM223	TN	WT	WT	H1047R	WT	WT	WT	WT	WT	K132R
SW527	TN	ND	ND	ND	ND	ND	ND	ND	ND	ND
AU565	$HER2^+$	WT	WT	WT	WT	WT	WT	WT	WT	R175H
BT-474	$HER2^+$	WT	WT	WT	WT	WT	WT	WT	WT	E285K
EFM-192A	$HER2^+$	WT	WT	C420R	WT	WT	WT	WT	WT	270fs
HCC1419	$HER2^+$	WT	WT	WT	WT	WT	WT	WT	WT	81fs, Y220C
HCC1569	$HER2^+$	PTEN (−)	266fs	WT	WT	WT	WT	WT	WT	227fs, E294X
HCC1954	$HER2^+$	WT	WT	H1047R	WT	WT	WT	WT	WT	Y163C
HCC202	$HER2^+$	ND	WT	E545K	WT	WT	WT	WT	WT	283>FS
HCC2218	$HER2^+$	WT	WT	WT	WT	WT	WT	WT	WT	R283C
JIMT-1	$HER2^+$	WT	WT	C420R	WT	WT	WT	WT	WT	R248W
MDA-MB-361	$HER2^+$	WT	WT	E545K	WT	WT	WT	WT	D108H	E56X
MDA-MB-453	$HER2^+$	WT	E306K	H1047R	WT	WT	WT	WT	WT	WT
SK-BR-3	$HER2^+$	WT	WT	WT	WT	WT	WT	WT	WT	R175H
UACC-812	$HER2^+$	WT	WT	WT	WT	WT	WT	WT	WT	WT
UACC-893	$HER2^+$	ND	WT	H1047R	WT	WT	WT	WT	WT	R342X
ZR-75-30	$HER2^+$	WT	WT	WT	WT	WT	WT	WT	WT	WT
BT-483	ER^+/PR^+	WT	WT	E542K	WT	WT	WT	WT	WT	246I
CAMA-1	ER^+/PR^+	WT	WT	WT	WT	WT	WT	WT	WT	R280T
EFM-19	ER^+/PR^+	WT	WT	H1047L	WT	WT	WT	WT	ND	H193R
EVSA-T	ER^+/PR^+	PTEN (−)	317fs	WT	WT	WT	WT	WT	WT	S241C
HCC1428	ER^+/PR^+	WT	WT	WT	WT	WT	WT	WT	WT	WT
HCC1500	ER^+/PR^+	WT	WT	WT	WT	WT	WT	WT	WT	WT
KPL-1	ER^+/PR^+	WT	WT	E545K	WT	WT	WT	WT	ND	WT
MCF7	ER^+/PR^+	WT	WT	E545K	WT	WT	WT	WT	G6E	WT
MDA-MB-175VII	ER^+/PR^+	WT	WT	WT	WT	WT	WT	WT	WT	WT
MDAMB415	ER^+/PR^+	WT	WT	WT	WT	WT	WT	WT	WT	Y236C
T-47D	ER^+/PR^+	WT	WT	H1047R	WT	WT	WT	WT	WT	L194F
ZR-75-1	ER^+/PR^+	PTEN (−)	WT	WT	WT	WT	WT	WT	WT	WT
TN Frequency (%)		59*	6	18	29*	12	6	6	6	71
$HER2^+$ Frequency (%)		8	13	47	0	0	0	0	7	80
ER^+/PR^+ Frequency (%)		17	8	42	0	0	0	0	10	50

Abbreviations: WT, wild-type; ND, not done; fs, frameshift; X, stop codon.
*Mutation frequencies differ significantly between subtypes ($P < 0.01$).

DOI:10.1158/1541-7786.MCR-08-0107

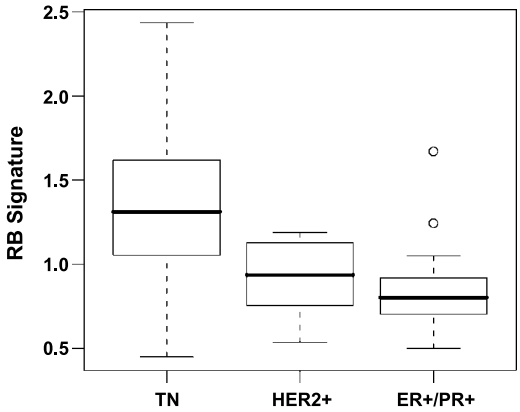

FIGURE 6. RB pathway dysregulation is associated with the triple-negative subtype. The RB signature score, reflecting pathway dysregulation, is summarized in a box plot by subtype.

up-regulated in breast cancer (56-58). *MYST2*, the only gene in the amplicon on 17q21 to meet all three selection criteria, encodes the HBO1 histone acetyltransferase implicated in the regulation of DNA synthesis (31, 34, 59, 60). HBO1 has also been shown to enhance transcription mediated by steroid receptors including ER and PR (32, 33, 61). Our finding that overexpression of *MYST2* enhances colony formation in soft agar by SKBR3 cells, which are $HER2^+$, suggests that *MYST2* is a bona fide oncogene that acts independently of ER and PR. Further studies are required to confirm and elucidate the relationship between *MYST2* and oncogenesis.

Another candidate driver gene that fits into an emerging pattern of subtype-specific pathway activation is *RASA1*, which encodes the RAS GTPase activating protein p120GAP. Somatic mutations in *RASA1* have been observed in basal cell carcinoma (62) and the gene is located in a region of frequent loss of heterozygosity in *BRCA1* mutant breast cancers (63), which have gene expression profiles similar to those of triple-negative cancers. *RASA1* is the only gene meeting all three selection criteria in the region of loss on 5q13 associated with the triple-negative and $HER2^+$ subtypes. p120GAP activates the intrinsic GTPase activity of wild-type RAS, shifting it into its inactive GDP-bound form. Triple-negative cancers also harbor all the KRAS and BRAF activating mutations we observed and have been reported by others to be characterized by activation of epidermal growth factor receptor signaling (6, 7, 18, 19). These several lines of evidence are consistent with triple-negative cancers being more frequently dependent on activation of mitogen-activated protein kinase signaling, suggesting they may be more susceptible to therapeutic intervention targeting this pathway.

Another interesting feature of triple-negative cancers that may have therapeutic implications is their increased genomic instability, as evidenced by frequent copy gains and the chromosomal instability expression signature. We cannot exclude the possibility that the differences we observed in genome instability between the subtypes are due to the representation of different stages of development within our tumor samples. However, we observed similar differences in instability between subtypes in cell lines. Triple-negative cancers also have more frequent alterations affecting PTEN and RB1, both of which are involved in maintaining genome integrity (64-66). Defects in cell cycle checkpoints may make triple-negative breast cancers more sensitive to cytotoxic therapy (67), particularly DNA damaging agents (39), and to other therapies that exploit defects in DNA repair (68).

Continuing advances in molecular profiling technologies are making it possible to identify the genetic and epigenetic alterations characteristic of cancer subtypes. Through integrative analysis of different data types, we can relate these alterations to the activity of oncogenic pathways and to therapeutic opportunities.

Materials and Methods

Breast Samples

Fresh frozen breast tumors with >70% tumor content were obtained from 51 patients by Gene Logic, Inc. Tumor samples were acquired from several centers and were accompanied by varying amounts of clinical data. Fixed material was not available to the authors of this study to directly evaluate HER2, ER, or PR status. The cell lines AU565, BT-474, BT-483, BT-549, CAMA-1, DU4475, HCC1143, HCC1395, HCC1419, HCC1428, HCC1500, HCC1569, HCC1806, HCC1954, HCC202, HCC2218, HCC38, HCC70, Hs578T, KPL-1, MCF7, MDA-MB-175VII, MDA-MB-231, MDA-MB-361, MDA-MB-415, MDA-MB-436, MDA-MB-453, MDA-MB-468, SK-BR-3, SW527, T-47D, UACC-812, UACC-893, ZR-75-1, and ZR-75-30 were obtained from American Type Culture Collection. The cell lines CAL-120, CAL-148, CAL-51, CAL-85-1, EFM-19, EFM-192A, EVSA-T, HCC1937, HDQ-P1, JIMT-1, and MFM-223 were obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH. All cell lines were maintained in RPMI 1640 or DMEM supplemented with 10% fetal bovine serum (Sigma) and 2 mmol/L L-glutamine. Cell line copy number analysis (69) was carried out on genomic DNA extracted from subconfluent cell cultures using Qiagen DNAeasy kits.

Microarray Profiling

All microarrays were purchased from Affymetrix. Gene expression data were generated for all 51 tumor and 46 cell line samples with HG-U133 Plus 2.0 arrays according to the manufacturer's protocols and processed using Microarray Suite 5.0 software (Affymetrix). Where multiple probe sets were available for a gene, the probe set with the highest mean intensity was used. Tumor DNA was run on Mapping 500k SNP arrays and cell line DNA was run on Mapping 100k SNP arrays according to the manufacturer's protocols. Signal intensities were extracted using dChip 2005 (70), median centered, converted to log 2 ratios, and segmented using the GLAD R package (71). Copy numbers were calculated as $2^{(\text{inferred log 2 ratio} + 1)}$. Other analyses were carried out in R[5] and Spotfire (TIBCO). Microarray data are available in the National Center for Biotechnology Information Gene Expression Omnibus repository under accession nos. GSE7545, GSE12777, and GSE13696.

[5] http://www.r-project.org

DOI:10.1158/1541-7786.MCR-08-0107

Breast Cancer Subtype Assignment

ER and PR classifiers were developed using gene expression data from 143 tumors (ER) and 142 tumors (PR), which had immunohistochemistry status supplied by Gene Logic. Samples for which any level of positivity was recorded were considered positive. The probe set with the smallest residual variance was selected for each gene (ER: 205225_at, PR: 208305_at) and used to train a univariate logistic regression model. For HER2, our intent was to mimic the criteria used to direct trastuzumab therapy, but the HER2 status information available from Gene Logic was not detailed enough. We assigned the HER2-positive status to samples with *ERBB2* mRNA expression (probe set 216836_s_at) or copy number (by SNP array) at least 2 SD above the mean (Supplementary Fig. S1). These classifiers were used to determine subtype for 10 cell lines and all 51 breast tumors. HER2-positive samples were assigned to the HER2 subtype; HER2-negative samples that were positive for ER or PR were assigned to the ER^+/PR^+ subtype; and HER2-, ER-, and PR-negative samples were assigned to the triple-negative subtype. Subtype assignment with 370 "intrinsic" genes (10) mapped to the Affymetrix microarray was by hierarchical clustering with Pearson correlation. Samples were also assigned to the higher-resolution subtype (26) for which the gene expression centroid had the highest Spearman correlation.

Analysis of Genome-Wide Copy Number Alterations

The May 2004 human genome release (72) was divided into nonoverlapping 100kb segments. The representative copy number for each of 27,820 segments with at least one SNP was calculated by taking the median GLAD-segmented copy number for the SNPs found therein. The fraction of segments with gains or losses was then counted for a range of gain or loss thresholds. Similar results were observed when the entire genome was divided into 862 cytobands according to the boundaries available in release hg17 of the UCSC genome browser (ref. 73; data not shown). Kruskal-Wallis tests were used to evaluate the association between gain or loss frequency and subtype.

Gene Expression Signatures

Chromosomal instability (CIN25) and RB signatures were calculated as the mean MAS5 expression level for the relevant genes (28, 40) and scaled to have a mean value of 1. The Wilcoxon rank sum test was used to evaluate subtype differences.

Association between Copy Number and Subtypes

The unique set of copy number segment boundaries from all tumor samples was used to define the boundaries of 4,556 genomic intervals. The segmented copy number value within each interval is constant in each tumor sample. For further analysis, 2,714 intervals including at least three SNPs and with absolute log 2 copy number ratio >0.3 in at least five tumors were selected. Association between the copy number alteration in each interval and subtype was evaluated by two-sample *t* test. False discovery rates were estimated using the *q* value procedure (74). Intervals with *q* values <7% were tested by the same analysis procedure in 47 breast cancer cell lines, and the intervals showing consistent association ($P < 0.05$) were selected.

Identification of Candidate Driver Genes

The regions of most notable copy number gain were identified by multiplying mean gain amplitude by gain frequency (75). Expression levels in samples with and without copy number alteration were compared by two-sided Wilcoxon rank sum test. A log 2 copy number ratio <−0.3 was used to identify samples with losses and a ratio >0.3 was used to identify samples with gains. Association between expression and clinical outcome (time to distant metastasis) was evaluated using the data of van't Veer et al. (29). Their microarray probes were assigned to genes by mapping to the genome with BLAT (76) and finding overlaps with alignments of RefSeq mRNA sequences from UCSC (ref. 73; hg17 release). Within each region the relationship between expression level and time to distant metastasis was estimated by a Cox proportional hazards model and the significance level was calculated by a likelihood ratio test. *P* value thresholds of 0.05 were used without adjustments for multiple testing. The results of this analysis are limited to prioritizing candidates for follow-up and do not provide statistical evidence that any individual gene is a driver.

Somatic Mutation Data in Breast Cell Lines

Exonic regions were amplified by PCR of genomic DNA and sequenced in both directions using Big Dye Terminator Kit reagents and ABI PRISM 3730xl sequencing machines (Applied Biosystems). Sequence variants were identified using Mutation Surveyor software (SoftGenetics) and confirmed by manual trace review and a second round of sequencing. Matched nontumor samples were not available to confirm that the variants are somatic. Oncogenes tend to be subject to a relatively small number of recurrent activating mutations, and for these genes we only analyzed mutations that have been shown to be somatic in other studies (35). For tumor suppressors, a much wider variety of mutations can impair function, so we considered all variants that do not correspond to any record in dbSNP (77) and that change the amino acid sequence. It is possible that a few of these variants represent novel polymorphisms or functionally insignificant mutations. However, the vast majority correspond to known somatic mutations (35), and most of the rest introduce frameshifts and are likely to result in protein inactivation. Associations between mutation status and subtype were evaluated by two-sided Fisher's exact test.

Western Blotting

Equal amounts of protein were separated by electrophoresis through NuPage Bis-Tris 4% to 12% gradient gels (Invitrogen). Proteins were transferred onto polyvinylidene difluoride membranes using the Criterion system from Bio-Rad. Specific antigen-antibody interaction was detected with a horseradish peroxidase–conjugated secondary antibody IgG using enhanced chemiluminescence Western blotting detection reagents (Amersham Biosciences). The anti-PTEN antibody (Santa Cruz Biotechnology) was used with an anti–glyceraldehyde-3-phosphate dehydrogenase antibody (Abcam) as loading control. The anti-MYST2 antibody (Santa Cruz Biotechnology) was used with an anti-actin antibody (Santa Cruz Biotechnology) as loading

DOI:10.1158/1541-7786.MCR-08-0107

control. Western blot images were scanned and MYST2 and actin bands were quantitated using Image J software.[6]

Transient Transfection and Colony Formation Assay

Cells were transfected with pCMV-XL5 expressing MYST2 (Origene) using DharmaFECT 1 (Dharmacon). Seven thousand five hundred cells in growth medium containing 0.4% agarose were seeded in each well of a six-well plate on medium containing 0.8% agarose. Colonies were stained after 3 to 4 wk with 5 mmol/L Calcein AM (Invitrogen) and counted with the ImageXpress Micro system (Molecular Devices).

[6] http://rsbweb.nih.gov/ij/

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank Jeffrey Eastham-Anderson for imaging assistance, C. Bone for inspiration, and Rich Neve, Zemin Zhang, and anonymous reviewers for helpful comments and discussions.

References

1. Hudis CA. Trastuzumab—mechanism of action and use in clinical practice. N Engl J Med 2007;357:39–51.

2. Bush NJ. Advances in hormonal therapy for breast cancer. Semin Oncol Nurs 2007;23:46–54.

3. Jordan VC. SERMs: meeting the promise of multifunctional medicines. J Natl Cancer Inst 2007;99:350–6.

4. Ponzone R, Biglia N, Jacomuzzi ME, Mariani L, Dominguez A, Sismondi P. Antihormones in prevention and treatment of breast cancer. Ann N Y Acad Sci 2006;1089:143–58.

5. Cleator S, Heller W, Coombes RC. Triple-negative breast cancer: therapeutic options. Lancet Oncol 2007;8:235–44.

6. Nielsen TO, Hsu FD, Jensen K, et al. Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. Clin Cancer Res 2004;10:5367–74.

7. Turner NC, Reis-Filho JS. Basal-like breast cancer and the BRCA1 phenotype. Oncogene 2006;25:5846–53.

8. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. Nature 2000;406:747–52.

9. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98:10869–74.

10. Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 2003;100:8418–23.

11. Abd El-Rehim DM, Pinder SE, Paish CE, et al. Expression of luminal and basal cytokeratins in human breast carcinoma. J Pathol 2004;203:661–71.

12. Gusterson BA, Ross DT, Heath VJ, Stein T. Basal cytokeratins and their relationship to the cellular origin and functional classification of breast cancer. Breast Cancer Res 2005;7:143–8.

13. Lakhani SR, Reis-Filho JS, Fulford L, et al. Prediction of BRCA1 status in patients with breast cancer using estrogen receptor and basal phenotype. Clin Cancer Res 2005;11:5175–80.

14. Kapp AV, Jeffrey SS, Langerod A, et al. Discovery and validation of breast cancer subtypes. BMC Genomics 2006;7:231.

15. Perren A, Weng LP, Boag AH, et al. Immunohistochemical evidence of loss of PTEN expression in primary ductal adenocarcinomas of the breast. Am J Pathol 1999;155:1253–60.

16. Saal LH, Gruvberger-Saal SK, Persson C, et al. Recurrent gross mutations of the PTEN tumor suppressor gene in breast cancers with deficient DSB repair. Nat Genet 2008;40:102–7.

17. Saal LH, Holm K, Maurer M, et al. PIK3CA mutations correlate with hormone receptors, node metastasis, and ERBB2, and are mutually exclusive with PTEN loss in human breast carcinoma. Cancer Res 2005;65:2554–9.

18. Sainsbury JR, Farndon JR, Sherbet GV, Harris AL. Epidermal-growth-factor receptors and oestrogen receptors in human breast cancer. Lancet 1985;1:364–6.

19. van der Groep P, Bouter A, van der Zanden R, et al. Re: Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. J Natl Cancer Inst 2004;96:712–3; author reply 4.

20. Reis-Filho JS, Savage K, Lambros MB, et al. Cyclin D1 protein overexpression and CCND1 amplification in breast carcinomas: an immunohistochemical and chromogenic *in situ* hybridisation analysis. Mod Pathol 2006;19:999–1009.

21. Al-Kuraya K, Schraml P, Torhorst J, et al. Prognostic relevance of gene amplifications and coamplifications in breast cancer. Cancer Res 2004;64:8534–40.

22. Bergamaschi A, Kim YH, Wang P, et al. Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer. Genes Chromosomes Cancer 2006;45:1033–40.

23. Chin K, DeVries S, Fridlyand J, et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell 2006;10:529–41.

24. Fridlyand J, Snijders AM, Ylstra B, et al. Breast tumor copy number aberration phenotypes and genomic instability. BMC Cancer 2006;6:96.

25. Loo LW, Grove DI, Williams EM, et al. Array comparative genomic hybridization analysis of genomic alterations in breast cancer subtypes. Cancer Res 2004;64:8541–9.

26. Hu Z, Fan C, Oh DS, et al. The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics 2006;7:96.

27. Neve RM, Chin K, Fridlyand J, et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell 2006;10:515–27.

28. Carter SL, Eklund AC, Kohane IS, Harris LN, Szallasi Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. Nat Genet 2006;38:1043–8.

29. van't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415:530–6.

30. Kao J, Pollack JR. RNA interference-based functional dissection of the 17q12 amplicon in breast cancer reveals contribution of coamplified genes. Genes Chromosomes Cancer 2006;45:761–9.

31. Doyon Y, Cayrou C, Ullah M, et al. ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. Mol Cell 2006;21:51–64.

32. Georgiakaki M, Chabbert-Buffet N, Dasen B, et al. Ligand-controlled interaction of histone acetyltransferase binding to ORC-1 (HBO1) with the N-terminal transactivating domain of progesterone receptor induces steroid receptor coactivator 1-dependent coactivation of transcription. Mol Endocrinol 2006;20:2122–40.

33. Sharma M, Zarnegar M, Li X, Lim B, Sun Z. Androgen receptor interacts with a novel MYST protein, HBO1. J Biol Chem 2000;275:35200–8.

34. Thomas T, Voss AK. The diverse biological roles of MYST histone acetyltransferase family proteins. Cell Cycle 2007;6:696–704.

35. Forbes S, Clements J, Dawson E, et al. Cosmic 2005. Br J Cancer 2006;94:318–22.

36. Barker KT, Crompton MR. Ras-related TC21 is activated by mutation in a breast cancer cell line, but infrequently in breast carcinomas *in vivo*. Br J Cancer 1998;78:296–300.

37. Chan AM, McGovern ES, Catalano G, Fleming TP, Miki T. Expression cDNA cloning of a novel oncogene with sequence similarity to regulators of small GTP-binding proteins. Oncogene 1994;9:1057–63.

38. Depowski PL, Rosenthal SI, Ross JS. Loss of expression of the PTEN gene protein product is associated with poor outcome in breast cancer. Mod Pathol 2001;14:672–6.

39. Bosco EE, Wang Y, Xu H, et al. The retinoblastoma tumor suppressor modifies the therapeutic response of breast cancer. J Clin Invest 2007;117:218–28.

40. Markey MP, Bergseid J, Bosco EE, et al. Loss of the retinoblastoma tumor suppressor: differential action on transcriptional programs related to cell cycle control and immune function. Oncogene 2007.

41. Gong Y, Yan K, Lin F, et al. Determination of oestrogen-receptor status and ERBB2 status of breast carcinoma: a gene-expression profiling study. Lancet Oncol 2007;8:203–11.

42. Rae JM, Creighton CJ, Meck JM, Haddad BR, Johnson MD. MDA-MB-435 cells are derived from M14 Melanoma cells-a loss for breast cancer, but a boon for melanoma research. Breast Cancer Res Treat 2007;104:13–9.

43. Ross DT, Scherf U, Eisen MB, et al. Systematic variation in gene expression patterns in human cancer cell lines. Nat Genet 2000;24:227–35.

DOI:10.1158/1541-7786.MCR-08-0107

44. Schlosshauer PW, Brown SA, Eisinger K, et al. APC truncation and increased β-catenin levels in a human breast cancer cell line. Carcinogenesis 2000;21:1453–6.

45. Albertson DG. Gene amplification in cancer. Trends Genet 2006;22:447–55.

46. Cheng KW, Lahad JP, Kuo WL, et al. The RAB25 small GTPase determines aggressiveness of ovarian and breast cancers. Nat Med 2004;10:1251–6.

47. Garcia MJ, Pole JC, Chin SF, et al. A 1 Mb minimal amplicon at 8p11-12 in breast cancer identifies new candidate oncogenes. Oncogene 2005;24:5235–45.

48. Garraway LA, Widlund HR, Rubin MA, et al. Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. Nature 2005;436:117–22.

49. Kauraniemi P, Kallioniemi A. Activation of multiple cancer-associated genes at the ERBB2 amplicon in breast cancer. Endocr Relat Cancer 2006;13:39–49.

50. Carroll JS, Liu XS, Brodsky AS, et al. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell 2005;122:33–43.

51. Laganiere J, Deblois G, Lefebvre C, Bataille AR, Robert F, Giguere V. From the Cover: location analysis of estrogen receptor α target promoters reveals that FOXA1 defines a domain of the estrogen response. Proc Natl Acad Sci U S A 2005;102:11651–6.

52. Barlund M, Forozan F, Kononen J, et al. Detecting activation of ribosomal protein S6 kinase by complementary DNA and tissue microarray analysis. J Natl Cancer Inst 2000;92:1252–9.

53. Proia DA, Nannenga BW, Donehower LA, Weigel NL. Dual roles for the phosphatase PPM1D in regulating progesterone receptor function. J Biol Chem 2006;281:7089–101.

54. Sinclair CS, Rowley M, Naderi A, Couch FJ. The 17q23 amplicon and breast cancer. Breast Cancer Res Treat 2003;78:313–22.

55. Vance KW, Carreira S, Brosch G, Goding CR. Tbx2 is overexpressed and plays an important role in maintaining proliferation and suppression of senescence in melanomas. Cancer Res 2005;65:2260–8.

56. Iorio MV, Ferracin M, Liu CG, et al. MicroRNA gene expression deregulation in human breast cancer. Cancer Res 2005;65:7065–70.

57. Si ML, Zhu S, Wu H, Lu Z, Wu F, Mo Y Y. miR-21-mediated tumor growth. Oncogene 2007;26:2799–803.

58. Zhu S, Si ML, Wu H, Mo YY. MicroRNA-21 targets the tumor suppressor gene tropomyosin 1 (TPM1). J Biol Chem 2007;282:14328–36.

59. Iizuka M, Stillman B. Histone acetyltransferase HBO1 interacts with the ORC1 subunit of the human initiator protein. J Biol Chem 1999;274:23027–34.

60. Neal KC, Pannuti A, Smith ER, Lucchesi JC. A new human member of the MYST family of histone acetyl transferases with high sequence similarity to *Drosophila* MOF. Biochim Biophys Acta 2000;1490:170–4.

61. Sharma D, Fondell JD. Ordered recruitment of histone acetyltransferases and the TRAP/Mediator complex to thyroid hormone-responsive promoters *in vivo*. Proc Natl Acad Sci U S A 2002;99:7934–9.

62. Friedman E, Gejman PV, Martin GA, McCormick F. Nonsense mutations in the C-terminal SH2 region of the GTPase activating protein (GAP) gene in human tumours. Nat Genet 1993;5:242–7.

63. Johannsdottir HK, Jonsson G, Johannesdottir G, et al. Chromosome 5 imbalance mapping in breast tumors from BRCA1 and BRCA2 mutation carriers and sporadic breast tumors. Int J Cancer 2006;119:1052–60.

64. Hernando E, Nahle Z, Juan G, et al. Rb inactivation promotes genomic instability by uncoupling cell cycle progression from mitotic control. Nature 2004;430:797–802.

65. Puc J, Keniry M, Li HS, et al. Lack of PTEN sequesters CHK1 and initiates genetic instability. Cancer Cell 2005;7:193–204.

66. Shen WH, Balajee AS, Wang J, et al. Essential role for nuclear PTEN in maintaining chromosomal integrity. Cell 2007;128:157–70.

67. Rouzier R, Perou CM, Symmans WF, et al. Breast cancer molecular subtypes respond differently to preoperative chemotherapy. Clin Cancer Res 2005;11:5678–85.

68. McCabe N, Turner NC, Lord CJ, et al. Deficiency in the repair of DNA damage by homologous recombination and sensitivity to poly(ADP-ribose) polymerase inhibition. Cancer Res 2006;66:8109–15.

69. Huang J, Wei W, Zhang J, et al. Whole genome DNA copy number changes identified by high density oligonucleotide arrays. Human genomics 2004;1:287–99.

70. Lin M, Wei LJ, Sellers WR, Lieberfarb M, Wong WH, Li C. dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. Bioinformatics 2004;20:1233–40.

71. Hupe P, Stransky N, Thiery JP, Radvanyi F, Barillot E. Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. Bioinformatics 2004;20:3413–22.

72. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. Nature 2001;409:860–921.

73. Karolchik D, Baertsch R, Diekhans M, et al. The UCSC Genome Browser Database. Nucleic Acids Res 2003;31:51–4.

74. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 2003;100:9440–5.

75. Takeyama K, Monti S, Manis JP, et al. Integrative analysis reveals 53BP1 copy loss and decreased expression in a subset of human diffuse large B-cell lymphomas. Oncogene 2007.

76. Kent WJ. BLAT—the BLAST-like alignment tool. Genome Res 2002;12:656–64.

77. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001;29:308–11.