Cancer Cell, Volume *15*

Supplemental Data

Article

MTDH Activation by 8q22 Genomic Gain Promotes Chemoresistance and Metastasis of Poor-Prognosis Breast Cancer

Guohong Hu, Robert A. Chong, Qifeng Yang, Yong Wei, Mario A. Blanco, Feng Li, Michael Reiss, Jessie L.-S. Au, Bruce G. Haffty, and Yibin Kang

Supplemental Experimental Procedures

Validation of ACE in Various Data Sets

To validate the ACE approach, we analyzed several published expression data sets with the corresponding information of genomic alterations or long-range epigenetic regulation. For each expression data set with probe detection flags available, the genes that were flagged as "absent" in more than 90% of the samples were removed from further analysis. Duplicate probes mapped to the same transcripts were collapsed and the average expression intensities were used. Expression data were normalized for each study so that each hybridization had equal median intensity across the entire array. Student's t-test was used to score the gene expression prior to NS calculation. To avoid possible bias, we did not perform data set-specific optimization of ACE analysis, but used a uniform set of pre-defined analysis parameters.

We first analyzed the gene expression data of the Ts1Cje mouse (Amano et al., 2004), which is the animal model for human Down Syndrome and hosts a partial trisomic region from gene *Sod1* to *Znf295* on chromosome 16. Affymetrix microarray expression data of the Ts1Cje and normal mouse brain tissues were downloaded from the NCBI GEO database (accession number GSE1294). Genes were scored by the expression differences between trisomic and normal mice followed by NS calculation. ACE detected only one region of gain and no regions of loss in trisomic mice. The significant region overlapped precisely to the expected area (Fig. S1A). The first p-distal boundary in the detected region corresponds to the gene *Mylc2b,* which is immediately adjacent to *Sod1* on the chromosome. The second expected boundary gene *Znf295* is located q-distal of all the probes available on the microarray and ACE consistently defined the region to the end of the q arm.

Next we used ACE to analyze the gene expression of taxane-resistant ovarian cancer cells compared to the parental lines (Wang et al., 2006). Results were validated with the aCGH data for the same samples. The expression and CGH data of 6 human ovarian cancer cell lines and their taxane-resistant derivatives were obtained from the Stanford Microarray Database (http://genome-www5.stanford.edu). ES were scored according to the expression difference of each gene between the parental and drug-resistant lines prior to NS calculation. To avoid bias, we used the same method as in the original paper, circular binary segmentation (CBS) (Olshen et

al., 2004), to analyze the aCGH data. CBS analysis detected 3 regions on chromosome 7 with increased copy numbers in the drug-resistant lines, which was consistent with the previous finding (Wang et al., 2006). ACE detected the same areas as the only significant regions (Fig. S1B). In addition to these significant regions, high concordance was observed between the NS and the CBS copy number data throughout the genome. The overall correlation between the NS and aCGH data was 0.55 (Pearson's correlation coefficient), whereas the correlation was only 0.16 if the original expression scores were used, suggesting that NS can significantly help uncover the correlation between gene dosage and expression. From the correlation data, we could deduce that approximately 30% of all variation observed in NS could be directly explained by the underlying variations in genetic copy number (Pollack et al., 2002).

We further tested ACE's performance in more complicated data using the MDA-MB-231 sublines with different degrees of breast-to-bone metastatic activities (Kang et al., 2003). We compared the expression profiles of 5 highly metastatic lines (2268, 2269, 2271, 2274, 2287, 1833) and 5 weakly metastatic lines (2297, 1834, 2293, 2295, ATCC) using ACE and detected 5 CNA events, including gain at 2p, 6p, 12q, 19q and loss at 7q, in metastasis. CGH analysis was performed as previously described (Kang et al., 2003) on these cell lines to validate our computational analysis. Four out of these 5 genetic events had been directly observed in our cytogenetic analysis. For example, consistent with the ACE prediction, aCGH data indicated a loss at the q arm of chromosome 7 in highly metastatic cells (Fig. S1C).

We are aware of the fact that long-range epigenetic alteration could also contribute to the regional gene deregulation. To test ACE's capability to detect such changes, we analyzed a data set of 57 bladder tumors (Stransky et al., 2006), and detected 22 regions with genes underexpressed in tumor tissues as compared to normal samples. By analyzing the aCGH data, we found that 15 of these regions were lost in more than 10% of the tumor tissues, but gained in significantly fewer tumors (binomial $p < 0.05$), indicating that genomic loss of these regions was associated with bladder carcinomas. Furthermore, 4 of the remaining regions were proven or suggested by Stransky *et al.* as regions under epigenetic control (Stransky et al., 2006) (Table S8). For example, a region at 3p22.3 was shown to be regulated by histone H3 trimethyl modification in tumor samples (Stransky et al., 2006) (Figure S1D).

ACE Analysis of the Three Published Data Sets

In search of metastasis-associated CNAs in breast cancer, we analyzed three published breast cancer data sets (van 't Veer et al., 2002; van de Vijver et al., 2002; Wang et al., 2005). The microarray data and patient records of the tumor samples were obtained from GEO (Wang, GSE2034), and Rosetta websites (van't Veer, http://www.rii.com/publications/2002/vantveer.html; van de Vijver, http://www.rii.com/publications/2002/nejm.html). Some of the samples in the van de Vijver study had been previously used in the van't Veer data set and thus were removed from the van de Vijver data set in our analysis to avoid bias. Gene expression data was compared between the patients who developed metastasis within 5 years and those free of metastasis for more than 5 years. Metastasis-specific CNA regions were identified in each data set and the SRO regions that were identified in more than one data set were defined as the consensus poor-prognosis CNAs. To analyze the prognostic power of the copy number at each SRO region, the NS of the center locus were calculated for each sample using the z-score like expression scores. All the samples were classified into two groups using an NS cutoff so that the number of samples in the high NS

group was equal to the number of samples with 5-year relapses. The clinical outcomes were analyzed by comparing the samples in the two groups (Table S2).

Comparison of ACE Algorithm to Previous Approaches

Several approaches have been previously reported for CNA prediction based on expression microarray data (Crawley and Furge, 2002; Hertzberg et al., 2007; Liu et al., 2006; Myers et al., 2004; Stransky et al., 2006; Zhou et al., 2004). The majority of these approaches utilized an intuitive "odd-ratio" like method, in which the individual genes were first defined as significant or not significant by a cutoff of the expression correlation with the phenotype, and the densities of the significant genes were analyzed for each region with a pre-chosen width. The regions with aberrantly high densities were predicted as regions of gain or loss. Our analyses with several expression data sets have shown that the "odd-ratio" approach with different significance cutoffs and window sizes generated quite inconsistent results, and therefore was not suitable for large-scale analysis of multiple data sets. ACE can be distinguished from these previous approaches by several features including: 1) A quantitative expression score, instead of the binary significant/non-significant flag of each gene is used for the regional analysis, which evades the problem associated with the arbitrary significance cutoff; 2) A position-dependent weight is employed for each neighboring gene of the locus in consideration, which more accurately reflects the fact that linkage strengths decrease with physical distances; and 3) All the genes on the chromosome, instead of those within an arbitrarily pre-chosen window size, were analyzed for each genomic locus. These features increase the sensitivity and the robustness of the algorithm.

Laser Capture Microdissection (LCM) and DNA/RNA Extraction

To quantify DNA copy number and expression of genes at 8q22 in clinical breast tumor specimens, laser captured microdissection (LCM) was performed to isolate tumor cells from each tissue specimen. A panel of 50 snap-frozen breast tumors from anonymous patients was used in this study. These samples were examined by H&E staining and only those with approximately >50% tumor cells were selected for LCM followed by DNA and RNA extraction. The quality of DNA/RNA preparation was monitored by O.D. reading, leaving 36 high-quality samples for analysis. Another panel of 50 formalin-fixed paraffin-embedded tissues from anonymous patients was examined for MTDH expression by immunohistochemistry, and 20 samples were selected with strong or negative MTDH staining for microdissection followed by DNA extraction. Two of these samples failed in the DNA preparation step and thus 18 samples were used in the analysis.

For each sample, sequential sections of 15 μm thickness were prepared for LCM. The sections were mounted on the glass PEN-membrane slides (Leica) and stained using the Histogene staining solution (Arcturus) following the manufacturer's protocol. Slides were then immediately transferred for microdissection using a Leica AS LMD microscope. Approximately 10,000 tumor cells were prepared for DNA purification for each sample. For the fresh tumors, a separate sample of ~10,000 tumors cells was collected in 20 μl of RNAlater stabilization reagent (Qiagen) for RNA extraction.

DNA extraction was performed as previously described (Frank et al., 1996) with or without the paraffin-dissolving step for archived and fresh tumors, respectively. The RNeasy mini kit (Qiagen) was used to extract RNA from the tumor samples according to the manufacturer’s instructions.

Real-Time PCR and Data Evaluation

To analyze the DNA copy numbers, primer pairs were designed using the intron sequences of genes at chromosome 8q, including *CA2* (8q21), *LAPTM4β* and *MTDH* (8q22), and *EIF3S6* (8q23). Real-time PCR and data analysis were performed essentially as previously described (Glockner et al., 2001; Lehmann et al., 2000). Briefly, primers were designed using the software PrimerExpress (Applied Biosystems). PCR was performed using CyberGreen Universal PCR Master Mix (Applied Biosystems) with the ABI Prism 7900HT thermocycler (Applied Biosystems) according to the manufacturer's protocol. The absolute DNA copy number of each sample was analyzed with SDS 2.0 software (Applied Biosystems) using standard curves of known concentrations. The gene *APP*, located at 21q21, for which no amplifications in breast cancer have been reported, was used as the internal reference locus (Glockner et al., 2001; Lehmann et al., 2000). The copy numbers of the samples were normalized by healthy human tissue DNA. The previously used copy number ratio threshold 1.8 was applied to define a genomic gain (Glockner et al., 2001; Lehmann et al., 2000).

qRT-PCR was performed to analyze the RNA level of genes at 8q22, including *MTDH*, *LAPTM4β* and *PTDSS1* in fresh tumors following reverse transcription using the SuperScript first-strand synthesis kit (Invitrogen). The β-actin control kit (Applied Biosystems) was used for normalization. Primer sequences are listed in Table S9.

FISH (Fluorescence In Situ Hybridization)

For metaphase FISH, the cells were cultured to 80% confluence and treated with 0.05 μg/ml colcemid for 2 h. Then the cells were harvested by trypsinization and treated with 75 mM KCl for 20 min at room temperature, followed by fixation in fresh 3:1 methanol : acetic acid. The cells were washed in fixative for two more times and then dropped to precleaned slides. Metaphase FISH and tissue FISH was performed by the Dana-Farber Cancer Institute Cytogenetic Core Facility. One microgram of DNA from the BAC clone RP11-662P7 (Children's Hospital Oakland Research Institute), which covers the *MTDH* locus and other areas at 8q22 was labeled using the Nick Translation kit and SpectrumOrange dUTP (Vysis) following the manufacturer's protocol. Chromosome enumeration probe CEP8 labeled with SpectrumGreen (Vysis) was used for centromere 8 hybridization. Metaphase slides were pretreated with 0.1 mg/ml RNase, washed with 2x SSC and dehydrated in ethanol. Paraffin-embedded tissue slides were pretreated with xylene, dehydrated and digested with Digest-All 3 (Zymed). The slides were then washed in 1x PBS, fixed in formalin, and dehydrated in ethanol. Probes were added onto the slides and denatured at 94°C for three minutes. Hybridization was performed at 37°C in a humidified chamber. Forty-eight hours later the slides were washed in 2x SSC at 72°C and phosphate buffered tween-20 solution at room temperature, and counterstained with DAPI. Hybridization signals were viewed on a fluorescence Olympus BX-51 microscope system. For each tissue sample 50-100 nuclei were analyzed and the average 8q22 copy numbers were calculated. Eighty-two of the 170 samples on the tissue microarray with successful hybridization were analyzed and scored by the staff of the Cytogenetic Core at Dana Farber Cancer Institute.

Tissue Array Immunostaining

A breast cancer tissue microarray composed of 170 primary tumors was used in our clinical study. At the time of tumor resection, these patients were at an age of 25 to 49 years (median =

40 yrs, SD = 4.7 yrs). All patients in the study were treated with breast conserving surgery followed by radiation therapy to the intact breast. Systemic therapy was administered as clinically indicated in accordance with standard clinical practice. Local or regional relapses were defined as clinically and histologically documented relapses in the ipsilateral breast or regional nodes. Distant metastases were defined as clinical evidence of distant disease based on clinical and/or radiographic findings (Table S4).

Immunostaining was performed at the immunohistochemistry core facility of the Cancer Institute of New Jersey (CINJ) with a rabbit monoclonal anti-MTDH antibody (Invitrogen) and a rabbit polyclonal anti-CCNE2 antibody (Imgenex). A BLAST search of the antigen sequence used to raise the CCNE2 antibody was performed to ensure it does not cross-react with other cyclin E family members. Out of the 170 samples, 117 samples were stained successfully for MTDH and 133 samples for CCNE2. Each sample was scored as negative (0), low (1), medium (2), or high (3) according to staining intensities. A Kaplan-Meier curve was used to compare the survival rates of patients with low (scores 0 and 1) and high (scores 2 and 3) levels of MTDH or CCNE2. Log rank and Wilcoxon tests were used to compare the differences between curves using the SAS statistical software package. To assess whether the MTDH prognosis significance was associated with the other clinicopathological factors, Cox analysis of MTDH stratified with the expression status of ER, PR, HER or p53 (negative or positive), or the primary tumor sizes (smaller or larger than 2 cm) was performed. Multivariate Cox analysis with all the parameters in assessment was also undertaken to analyze the dependence of MTDH significance on other parameters.

Generation of Knockdown and Overexpression Cells

MTDH, ALDH3A1, and *MET* knockdown were achieved with the pSuper-Retro system with puromycin or hygromycin selection markers (OligoEngine) using the following sequences: 5'-GGCAGGTATCTTTGTAACTA-3' (*MTDH* KD1), 5'-GCTGACTGATTCTGGTTCAT-3' (*MTDH* KD2), 5'-CGCTACTTATGTGAACGTAA-3' (*MET*) and 5'-GGTTCGACCATATCCTGTA-3' (*ALDH3A1*). shRNA retroviral vectors were transfected into the amphotropic Phoenix packaging cell line and viruses were collected, filtered and used to infect target cells in the presence of 5 μg/ml polybrene 48 h after transfection. The infected cells were selected with 0.5 μg/ml puromycin or/and 0.4 mg/ml hygromycin. Double knockdown of *MET* and *ALDH3A1* was achieved by simultaneous infection of *MET* and *ALDH3A1* targeting viruses with different drug selection markers. Vectors expressing a non-targeting shRNA sequence were used for establishing control cell lines. *MTDH, LAPTM4b, PTDSS1, SDC2, TSPYL5,* and *UQCRB* overexpression was achieved using the retroviral expression vector pBabe-hygro. Viruses were generated and used to infect target cells as above and the infected cells were selected with 0.4 mg/ml hygromycin. For combinational overexpression of genes at 8q22, the viruses generated from the expression vector pBabe-puro containing each of the four genes were concentrated by ultracentrifugation and pooled for infection. Northern blots, qRT-PCR, and/or Western blots were performed to validate the knockdown or overexpression of target genes.

To generate an inducible knockdown of *ALDH3A1*, a retroviral vector expressing the Tet repressor (TetR) was constructed by cloning the TetR coding sequence from the pcDNA6/TR plasmid (Invitrogen) to pQCXIH (Clontech). LM2 cell line with stable expression of TetR was generated by transduction with retroviruses produced from pQCXIH-TetR. The cell line was then infected with retroviruses generated from the pRSMX vector (Ngo et al., 2006) containing the *ALDH3A1*-targeting shRNA sequence. The expression of a shRNA against *ALDH3A1* is

under the control of the histone H1 promoter and two adjacent tetracycline operators (TetOs). The bacterial Tet repressor (TetR) is constitutively expressed from the integrated pQCXIH-TetR in this cell line and suppresses the expression of shRNA by binding to TetOs. In the presence of 1 μg/ml doxycycline in the media, TetR is released from the TetOs and allows the transcription of *ALDH3A1* shRNA and thus the repression of *ALDH3A1* expression. The pMSCV-hygro vector was used to overexpress *ALDH3A1* and *MET*.

In Vivo Chemoresistance Assay

MTDH-knockdown or control LM2 cells (1×10^6 cells/0.1 ml in a 50:50 solution of PBS and Matrigel) were injected subcutaneously into each flank of nude mice. The mice were treated with chemotherapeutic drugs (20 mg/kg paclitaxel or 5 mg/kg doxorubicin) or the corresponding drug vehicles (Cremophor for paclitaxel and saline for doxorubicin) twice a week by intravenous delivery a week after the tumor xenografting. Six mice (12 tumors) were used for each group. Tumor growth was monitored twice a week by size measurement. Both maximum (L) and minimum (W) diameters of the tumor were measured using a slide caliper, and the tumor volume was calculated as $\pi LW^2/6$. Tumor growth was normalized to that before drug treatment.

Lung Histology

Mice were sacrificed and lungs were harvested followed by fixation in 10% neural buffered formalin overnight, washing with PBS and dehydration in 70% ethanol. Tissue paraffin-embedding, sectioning and H&E staining were performed by Histoserv, Inc. (Germantown, MD).

Wound Healing Assay

Cancer cells were grown in 10 cm culture dishes to confluence. A "wounding" line was scratched into the cell monolayer using a sterile pipet tip and its width was measured under microscope. The width was measured again at the same place after 3 h of culturing. The migration distance was defined as half of the difference between the scratch widths before and after the culturing period. Six measurements of each cell line were made and a Student's t-test was performed to compare the migration capacity of different cell lines.

Two-Chamber Migration Assay

10^5 luciferase-labeled cancer cells in serum-free medium were seeded into the upper chamber of the insert membranes with a 3 μm pore size (BD Bioscience) in a 24-well plate. Serum-containing medium was used in the bottom chamber as the attractant. After 12 h of culturing the cells in the upper chamber were removed using a cotton swab. The insert membrane with trans-well cells was cut off with a blade and added into a tube with cell lysis buffer. The cell numbers were quantified using a luciferase assay and the luminescence intensities of each line were normalized to that of 10^5 cells. A luciferase signal standard curve of each line with 10^2 to 10^5 cells was generated for quantification.

Matrigel Invasion Assay

Invasion assays were performed essentially as the above migration assay procedure except that the insert membrane was coated with a Matrigel (BD Bioscience) monolayer before cell seeding. Invasion index of each cell line was calculated as the fraction of trans-well cell number divided by that obtained in the migration assay.

Endothelial Adhesion Assay
To test the adhesion of cancer cells to the endothelial cells, different endothelial cell lines (HBMEC-60, from bone marrow; HUVEC, from umbilical vein; HMVEC-L, from lung microvascules) and a control fibroblast cell line WI-38 were grown to confluence in a 24-well plate. 10^5 luciferase-labeled cancer cells were seeded onto the endothelial monolayer. After 3 h of culturing, the unbound cells in the supernatant were removed by washing 3 times with PBS and the attached cancer cells were harvested by trypsinization. The cell number was quantified by luciferase assay as described above.

Chemoresistance Clonogenic Assay
Cancer cells with genetic modification of *MTDH* expression and the vector control were seeded into a 48-well plate (10^4 cells/well). After 24 h, the cells were treated with apoptosis-inducing chemicals for the indicated time (20, 50 or 100 nM paclitaxel, EMD Biosciences, 24 h; 50, 100 or 200 μM doxorubicin, EMD Biosciences, 24 h; 40 μM cisplatin, EMD Biosciences, 2 h; 10 or 80 ug/ml 4-hydroxycylcophosphamide, Toronto Research Chemical, 0.5 h; 200 or 500 μM H_2O_2, Fisher Scientific, 2 h) or 10 mJ/cm^2 UV irradiation. After culturing in drug-free DMEM medium for another 48 h, the surviving cells were quantified by clonogenic assay with the standard procedure for long-term recovery. Briefly, an aliquot of the harvested cell population was seeded onto a 10 cm dish. Crystal violet staining was used to count the colonies after 10 day culture in DMEM medium. The colony numbers from untreated cells of the same line were used to normalize the experimental data. In the HMVEC-L coculture assays, HMVEC-L cells were grown to confluence in the 48-well plates with supplemented EGM-2 medium (Lonza) before seeding of cancer cells. Because HMVEC-L cells could not form colonies in the DMEM medium (data not shown), the rest of the assay was performed following the standard procedure.

Drug Uptake and Retention Analysis of Paclitaxel and Doxorubicin
Cells were seeded into 12-well plates at densities of 3×10^5 per well in 1 ml of culture medium. One day after seeding, the medium was replaced with 1 ml of medium containing 50 nM [H^3]-Paclitaxel (Moravek, 2 Ci/mmol) or 100 nM [C^{14}]-Doxorubicin (GE HealthCare, 56 mCi/mmol). A pilot study showed biphasic kinetics in the uptake and retention of paclitaxel and doxorubicin in the parent LM2 cells. Based on this data, we selected 4 and 24 h for comparison of drug uptake and 4 and 12 h for comparison of retention, in all derivative cell lines. For the uptake study, cells were harvested immediately after incubation with drug-containing medium. For the retention study, cells were incubated with drug-containing medium for 4 h, followed by incubation in 2 ml drug-free medium and then harvested. After washing with cold PBS, the pelleted cells were lysed with 200 ul of 0.1 N NaOH. An aliquot (5 μl) was used to determine the protein concentration by Bradford assay (Sigma-Aldrich) with BSA as standards. The remaining cell lysates were transferred to scintillation count vials and mixed with 4 ml ECoScint scintillation fluid (National Diagnostics) and the radioactivity was measured by liquid scintillation counting. A standard curve was established and used to calculate the amount of cell-associated drug.

Endothelial Coculture, FACS, and Microarray Analysis
HMVEC-L cells were grown to confluence in 150 mm culture dishes and washed once with PBS before SNARF-1 labeling. The cells were cultured in serum-free EGM-2 medium containing 10 μM SNARF-1 (Molecular Probes) at 37°C for 30 min followed by washing with PBS twice. 2 ×

10^6 GFP-labeled LM2 control or KD1 cells were seeded into the plate in serum-containing DMEM medium. Cell sorting was performed in the Princeton Flow Cytometry Core Facility to purify the GFP^+ LM2 cells by using a FACSVantage SE cell sorter (BD Biosceinces) 48 h later (Fig. 5B). Cells were collected in RNAlater solution (Qiagen) and RNA extracted with RNeasy mini kit (Qiagen). The quality of purified RNA samples was monitored using a 2100 bioanalyzer (Agilent) before expression profiling.

To identify genes regulated by *MTDH* knockdown, RNA samples of LM2 control and *MTDH*-KD cells with or without HMVEC-L coculture were analyzed with the Agilent Whole Human Genome 4x44k arrays. RNA samples were labeled with Cy5 using the Agilent Low RNA Input Linear Amplification Kit and were hybridized with the Cy3-labeled Human Universal Reference RNA (Stratagene). Triplicate arrays were performed for each sample. Arrays were scanned with an Agilent G2565BA scanner and analyzed with the Agilent Feature Extraction v9.5 software. The Cy5/Cy3 ratios were calculated using the feature medium signal and normalized by the array median. Probes with >2.5 fold changes and Student's t-test p values <0.05 in both culturing conditions were identified as the MTDH regulated genes. Several significant genes, including *ALDH3A1*, *MET* and *HMOX1* were randomly selected for qRT-PCR confirmation with the RNA samples used for microarray analysis. RNA samples prepared from cells after the same FACS procedure but without HMVEC-L coculture were also analyzed by qRT-PCR to rule out the possibility that the expression differences were an artifact of the sorting procedure.

Pharmacologic Data Analysis

The pharmacological data set was downloaded from the NCI website http://dtp.nci.nih.gov, where the $-\log GI_{50}$ of 42,796 small molecules and natural products, as well the SNP microarray data were available for 58 human tumor cell lines (Garraway et al., 2005). GI_{50} was defined as the drug concentration necessary to inhibit cell growth by 50%. The SNP genotyping data were analyzed with the CBS algorithm (Olshen et al., 2004). A segment mean value of 0.4 was used as the threshold to define regional gain at the 8q22 region. Fifteen (26%) out of the 58 cell lines were classified as having a gain. Multiple $-\log GI_{50}$ entries of each compound were filtered as described (Garraway et al., 2005). The compounds were further filtered to exclude those with GI_{50} data in less than 50 cell lines. This yielded a final total of 24,642 compounds for further analysis. The $\log GI_{50}$ mean difference of each compound in the cells with and without 8q22 or 8q24 gain was calculated, and the significance of this difference was estimated by 1,000 permutations of the 8q22 or 8q24 status in the cell lines. The numbers of compounds with higher GI_{50} associated with 8q22 or 8q24 gain were counted by applying a significance threshold (0.05, 0.01, or 0.001, etc.) of GI_{50} difference and were compared to the permutations. A similar analysis with the mRNA expression microarray data of the genes on chromosome 8 was performed by dividing the cell lines into two groups with equal sizes, but with high and low levels of expression for each gene. We assessed the association of 8q22 copy number with gene expression by calculating a NS from the expression of genes in this region for each cell line as described earlier. A Pearson's correlation coefficient was calculated between the NS and the copy number.

Supplemental References

Amano, K., Sago, H., Uchikawa, C., Suzuki, T., Kotliarova, S.E., Nukina, N., Epstein, C.J., and Yamakawa, K. (2004). Dosage-dependent over-expression of genes in the trisomic region of Ts1Cje mouse model for Down syndrome. Hum Mol Genet *13*, 1333-1340.

Crawley, J.J., and Furge, K.A. (2002). Identification of frequent cytogenetic aberrations in hepatocellular carcinoma using gene-expression microarray data. Genome Biol *3*, RESEARCH0075.

Frank, T.S., Svoboda-Newman, S.M., and Hsi, E.D. (1996). Comparison of methods for extracting DNA from formalin-fixed paraffin sections for nonisotopic PCR. Diagn Mol Pathol *5*, 220-224.

Garraway, L.A., Widlund, H.R., Rubin, M.A., Getz, G., Berger, A.J., Ramaswamy, S., Beroukhim, R., Milner, D.A., Granter, S.R., Du, J.*, et al.* (2005). Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. Nature *436*, 117-122.

Glockner, S., Lehmann, U., Wilke, N., Kleeberger, W., Langer, F., and Kreipe, H. (2001). Amplification of growth regulatory genes in intraductal breast cancer is associated with higher nuclear grade but not with the progression to invasiveness. Lab Invest *81*, 565-571.

Hertzberg, L., Betts, D.R., Raimondi, S.C., Schafer, B.W., Notterman, D.A., Domany, E., and Izraeli, S. (2007). Prediction of chromosomal aneuploidy from gene expression data. Genes Chromosomes Cancer *46*, 75-86.

Kang, Y., Siegel, P.M., Shu, W., Drobnjak, M., Kakonen, S.M., Cordon-Cardo, C., Guise, T.A., and Massague, J. (2003). A multigenic program mediating breast cancer metastasis to bone. Cancer Cell *3*, 537-549.

Krump-Konvalinkova, V., Bittinger, F., Unger, R.E., Peters, K., Lehr, H.A., and Kirkpatrick, C.J. (2001). Generation of human pulmonary microvascular endothelial cell lines. Lab Invest *81*, 1717-1727.

Lehmann, U., Glockner, S., Kleeberger, W., von Wasielewski, H.F., and Kreipe, H. (2000). Detection of gene amplification in archival breast cancer specimens by laser-assisted microdissection and quantitative real-time polymerase chain reaction. Am J Pathol *156*, 1855-1864.

Liu, F., Park, P.J., Lai, W., Maher, E., Chakravarti, A., Durso, L., Jiang, X., Yu, Y., Brosius, A., Thomas, M.*, et al.* (2006). A genome-wide screen reveals functional gene clusters in the cancer genome and identifies EphA2 as a mitogen in glioblastoma. Cancer research *66*, 10815-10823.

Myers, C.L., Dunham, M.J., Kung, S.Y., and Troyanskaya, O.G. (2004). Accurate detection of aneuploidies in array CGH and gene expression microarray data. Bioinformatics *20*, 3533-3543.

Ngo, V.N., Davis, R.E., Lamy, L., Yu, X., Zhao, H., Lenz, G., Lam, L.T., Dave, S., Yang, L., Powell, J.*, et al.* (2006). A loss-of-function RNA interference screen for molecular targets in cancer. Nature *441*, 106-110.

Olshen, A.B., Venkatraman, E.S., Lucito, R., and Wigler, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics *5*, 557-572.

Pollack, J.R., Sorlie, T., Perou, C.M., Rees, C.A., Jeffrey, S.S., Lonning, P.E., Tibshirani, R., Botstein, D., Borresen-Dale, A.L., and Brown, P.O. (2002). Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. Proceedings of the National Academy of Sciences of the United States of America *99*, 12963-12968.

Stransky, N., Vallot, C., Reyal, F., Bernard-Pierrot, I., de Medina, S.G., Segraves, R., de Rycke, Y., Elvin, P., Cassidy, A., Spraggon, C.*, et al.* (2006). Regional copy number-independent deregulation of transcription in cancer. Nat Genet *38*, 1386-1396.

van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T.*, et al.* (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature *415*, 530-536.

van de Vijver, M.J., He, Y.D., van't Veer, L.J., Dai, H., Hart, A.A., Voskuil, D.W., Schreiber, G.J., Peterse, J.L., Roberts, C., Marton, M.J.*, et al.* (2002). A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med *347*, 1999-2009.

Wang, Y., Klijn, J.G., Zhang, Y., Sieuwerts, A.M., Look, M.P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M.E., Yu, J.*, et al.* (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet *365*, 671-679.

Wang, Y.C., Juric, D., Francisco, B., Yu, R.X., Duran, G.E., Chen, G.K., Chen, X., and Sikic, B.I. (2006). Regional activation of chromosomal arm 7q with and without gene amplification in taxane-selected human ovarian cancer cell lines. Genes Chromosomes Cancer *45*, 365-374.

Zhou, X., Cole, S.W., Hu, S., and Wong, D.T. (2004). Detection of DNA copy number abnormality by microarray expression analysis. Hum Genet *114*, 464-467.

Table S1. Poor Prognosis-Associated CNA Regions in Breast Cancer

	Region No.	Genomic Location	Data Set
Regions of Gain	1	1p32.2-q32.3	van't Veer
	2	1q41-42	Wang
	3	3q26-27	Wang, van't Veer
	4	5q13	Wang
	5	5p15	van de Vijver
	6	6p22.2	van de Vijver
	7	7p22.3-p22.2	van de Vijver
	8	8p11	Wang
	9	8q22	all
	10	8q24.3	Wang, Vijver
	11	11q13.1-13.2	van de Vijver
	12	12p13.1	van de Vijver
	13	15q15.1	van de Vijver
	14	16p11.2	van de Vijver
	15	16q22.1-q24	van de Vijver
	16	17q11.2	van de Vijver
	17	17q23.3-q25	Wang, Vijver
	18	20q11.21-q11.23	van de Vijver
	19	20q13.3	Wang, Vijver
	20	Xp11.22-11.23	van de Vijver
Regions of Loss	1	1p22.1-p21.3	van de Vijver
	2	1q2.1	van de Vijver
	3	2p11	Wang
	4	3p22.2-p13	van de Vijver
	5	5q11.2	van de Vijver
	6	6p21	Wang
	7	7q21.2-q21.3	van de Vijver
	8	8p21.2-p12	van de Vijver
	9	14q32	Wang
	10	17q21.2-q21.31	van't Veer
	11	19p13.2-p13.13	van't Veer
	12	19q13.43	van't Veer

Table S2. Cox Hazard Ratios for Relapse Based on NS of the Recurrent Regions of Gain in the Three Analyzed Data Sets

Region	Data Set	Sample #	HR	95% CI	p
8q22	van't Veer	78	2.52	1.26-5.04	0.0092
	van de Vijver	295	1.88	1.20-2.96	0.0062
	Wang	286	1.43	0.97-2.10	0.0700
3q26.33-q27.1	van't Veer	78	2.65	1.32-5.32	0.0060
	van de Vijver	295	1.52	1.01-2.28	0.0430
	Wang	286	1.67	1.14-2.44	0.0089
8q24.3	van't Veer	78	0.97	0.49-1.92	0.9370
	van de Vijver	295	0.76	1.18-2.62	0.0049
	Wang	286	1.29	0.88-1.91	0.1970
17q23.3-q25	van't Veer	78	1.00	0.61-1.98	0.9920
	van de Vijver	295	1.71	1.15-2.55	0.0080
	Wang	286	1.22	0.82-1.79	0.3240
20q13.3	van't Veer	78	0.78	0.39-1.56	0.4798
	van de Vijver	295	1.67	1.11-2.49	0.0120
	Wang	286	1.81	1.23-2.65	0.0021

Table S3. Genomic and cDNA qPCR Data of Genes at 8q22 in Human Breast Tumors

Sample Label	DNA Copy Numbers				RNA or Protein Expression			Note
	CA2 (8q21)	LAPTM4B	MTDH	EIF3S6 (8q23)	LAPTM4B	PTDSS1	MTDH	
F1	2.04	2.72	2.22	1.50	0.44	0.00	0.09	
F2	1.45	1.56	1.73	1.30	0.17	3.44	1.71	
F3	6.26	8.14	8.71	8.35	0.11	0.05	0.28	
F4	1.83	4.15	3.82	3.63	0.29	5.15	1.66	
F5	2.39	2.63	1.44	2.60	0.03	1.19	0.11	
F6	1.69	2.15	1.55	3.00	0.00	0.02	0.10	
F7	1.63	0.82	2.52	3.20	0.06	1.85	1.60	
F8	3.26	1.31	1.59	1.04	0.00	0.06	0.05	
F9	3.08	1.99	2.66	2.78	0.06	0.00	0.07	
F10	1.71	4.50	5.51	6.66	0.12	5.15	1.15	
F11	1.72	5.09	6.15	3.60	0.21	1.54	1.08	
F12	4.09	2.80	3.13	3.60	0.09	2.04	0.67	
F13	1.50	2.99	1.88	2.09	0.00	0.03	0.05	
F14	3.66	4.07	5.36	4.62	0.14	2.31	1.26	
F15	3.72	2.79	9.60	8.60	0.04	0.71	1.45	
F16	2.50	2.63	2.49	1.80	0.22	2.25	0.14	
F17	1.79	2.16	1.41	1.72	0.08	0.01	0.06	Fresh Frozen Tumor Samples, RNA expression detected by RT-qPCR
F18	2.51	2.44	2.45	3.60	0.13	2.25	0.07	
F19	2.95	0.37	1.44	3.80	0.00	0.00	0.00	
F20	3.52	4.36	5.85	3.70	0.19	10.56	1.30	
F21	2.69	6.51	8.34	3.02	0.04	0.36	1.18	
F22	2.55	1.76	1.48	3.20	0.04	0.00	0.09	
F23	1.69	2.49	3.05	1.92	0.02	0.00	0.00	
F24	3.02	6.01	7.69	5.37	0.17	3.44	1.11	
F25	1.50	1.20	1.61	1.19	0.00	0.01	0.13	
F26	2.10	2.92	2.41	1.94	0.04	0.02	2.34	
F27	1.89	1.95	1.52	2.20	0.01	0.11	0.76	
F28	2.72	2.99	2.44	2.80	0.00	0.02	1.04	
F29	1.96	1.63	2.44	2.16	0.06	1.85	0.00	
F30	2.52	1.30	1.67	2.79	0.12	5.15	0.40	
F31	2.72	3.42	2.96	1.62	0.14	2.31	0.59	
F32	4.60	3.19	3.02	1.67	0.22	4.25	0.01	
F33	3.85	4.46	3.91	3.80	0.09	0.68	1.67	
F34	2.37	2.19	1.88	2.00	0.03	0.00	0.02	
F35	2.50	2.45	3.13	2.34	0.00	0.00	0.00	
F36	2.72	2.74	1.59	1.97	0.09	0.16	0.00	
P1	1.69	-	1.50	5.14	-	-	Low	Paraffinized samples. Protein levels of MTDH detected by immunohisto-chemistry
P2	3.11	-	1.76	2.65	-	-	High	
P3	3.42	-	2.32	1.67	-	-	High	
P4	7.03	-	3.19	2.05	-	-	High	
P5	1.97	-	1.44	1.57	-	-	Low	
P6	1.75	-	1.43	3.41	-	-	Low	
P7	1.07	-	1.85	1.27	-	-	Low	
P8	1.70	-	1.23	1.74	-	-	Low	
P9	1.63	-	1.97	1.53	-	-	High	
P10	1.95	-	2.61	0.36	-	-	Low	
P11	15.0	-	15.4	16.7	-	-	High	
P12	3.93	-	1.88	3.27	-	-	Low	
P13	4.72	-	3.39	1.55	-	-	Low	
P14	6.38	-	9.63	5.96	-	-	High	
P15	7.81	-	7.40	6.97	-	-	High	
P16	13.67	-	11.78	6.44	-	-	High	
P17	5.62	-	3.16	3.83	-	-	High	
P18	7.34	-	3.27	1.04	-	-	Low	

Table S4. Patient Records of Tumors Used in the Breast Cancer Tissue Array

Patient ID	Age	MFS-time	CSS-time	Metastasis	Death	8q22	MTDH	CCNE2	ER	PR	HER	p53	P_size	CK5/6
p1	33	11.11	11.11	1	1	3.3	2	1	0	0	0	1		1
p10	37	19.59	19.59	0	0	1.8	1	3	1	1	0	0	2	0
p100	37	6.52	6.60	0	0		0	3	0	0	0	0	1.8	0
p101	38	2.32	2.32	1	1	2.9	2	0	0	0	0	1	4	1
p102	31	3.87	3.87	0	0	1.8	3	3	0	0	0	1	1.7	0
p103	40	7.12	7.12	0	0			0	0	0	1	0	1	
p104	42	7.68	7.68	0	0					0	0	0	1	
p105					1		0	0						
p106	39	12.54	12.54	1	0	2.1	2	2	1	1	1	0	1.3	0
p107	45	11.94	11.94	1	0				0	0	0	0	2	
p108	44	4.99	4.99	0	0		1	3	0	0	0	0	2.5	1
p109	36	21.79	21.79	1	1	1.9	0	0	0	0	0	0	2	0
p11	36	17.72	17.72	0	0			1					1.5	1
p110	45	5.14	5.14	1	0	3.0	3	1	0	0	0	0	1	
p111	41	1.58	1.58	0	0	1.9	1	1	1	1	0	0	1	0
p112	41	8.17	8.17	0	0	2.0	3	3	0	0	0	1	2.7	1
p113	45	3.41	3.41	0	0		1	1	0	0	0	1	1.7	1
p114	42	1.73	7.23	0	0			0					1	
p115	44	7.30	7.30	0	0	1.9	0	0	0	0	0	0	2	0
p116	41	3.94	4.10	0	0		0	0	0	0	0	0	1.4	0
p117	42	3.42	3.42	0	0		0	0	0	0	0	0	1.5	
p118	41	5.19	5.19	0	0		1	2	0	0	0	1	1.5	
p119	42	0.98	0.98	0	0		1	2	1	1	0	0	1.5	0
p12	33	10.37	10.37	0	0	1.7	0	0	0	0	0	0	1.5	
p120	44	3.00	3.00	1	1	3.4	0	1	0	0	0	0		0
p121	44	6.18	6.18	1	0	2.5	2	1	0	0	1	1	2.6	
p122	45	3.73	3.73	0	0			0	0	0	0	0	2.5	
p123	39	1.00	1.00	0	0		0	3	0	0	0	0	2	0
p124	39	1.21	1.21	1	0								2.5	
p125	43	2.24	2.24	0	0		0	0	0	0	0	0	1.8	0
p126	49	0.71	0.71	1	0								2.7	
p127	43	5.68	5.68	0	0		0	0	0	1	0	0	0.5	
p128	45	4.32	6.65	1	0			3	1	0	0	0	2	
p129	37	4.26	4.26	1	0	2.0	0	2	1	1	1	0	3.2	1
p13	40	19.01	19.01	1	0			3	1	1	1	0	2	0
p130	42	4.27	4.61	0	0	2.1	2	0	0	0	0	0	2	0
p131	39	5.95	5.95	0	0		0	3	1	1	0	0	1.2	1
p132	36	4.96	4.96	0	0		0	0	1	1	0	0	0.6	1
p133	49	6.19	6.19	0	0	2.6	1	0	0	0	0	1	2	
p134	44	0.74	0.74	0	0			1	1	0	0	0	0.3	
p135	42	3.51	3.51	0	0		0	2	1	0	0	0	2	
p136	40	4.42	4.42	0	0			0	1	1	0	0	1.6	
p137	44	4.78	4.78	0	0	1.7	0	0	0	0	1	0	1.5	0
p138	33	2.62	2.62	0	0	1.7	0	2	1	0	0	0	2.4	0
p139	45	4.81	4.81	0	0		0	3	1	1	0	0	1	0

p14	49	14.10	22.26	0	0	9.6	3	2	1	0	0	1	2	0
p140					1									
p141					1									
p142					1									
p143	40	7.60	7.60	0	1				0	0	0	0		
p144	40	7.60	7.60	0	1				0	0	0	0		
p145	41	1.94	1.94	0	0	3.2	3	3	0	1	0	0	2	0
p146	44	3.56	3.56	0	0	2.7	3	0	1	0	1	0	1.7	
p147	40	4.45	4.45	0	0	1.5	2	3	1	1	0	0	1.7	0
p148					1		2	2						
p149	34	0.33	0.33	0	0	1.8	1	0	0	0	0	0	2.4	
p15	44	13.22	13.22	0	0								1.5	
p150	45	4.41	4.41	0	0	5.1	2	1					1.5	0
p151	31	5.22	5.22	1	1	3.4	2	3	0	0	0	0	3.8	0
p152	43	4.10	4.10	0	0	1.9	0	2	1	1	0	0	1.5	0
p153	35	8.68	8.68	1	0			2	0	0	0	1	3.3	0
p154	44	6.34	6.34	0	0			0	0	0	0	1	1.7	
p155	43	3.85	3.85	0	0	1.5	0	3	0	0	0	0	1.1	0
p156	32	5.68	5.68	0	0				0	0	1	0	0.9	
p157	33	1.93	1.93	0	0		0	1	0	0	0	1	5	
p158	33	1.93	1.93	0	0		0	1	0	0	0	1	5	
p159	40	0.99	0.99	0	0	3.3	3	3	1	0	0	0	1.5	0
p16	43	7.89	7.89	1	1	2.0	2	1	1	1	0	1	1.5	0
p160	40	0.65	0.65	0	0	3.5	3	3	0	1	0	0	2.4	0
p161	40	1.65	1.65	0	0			3	1	1	0	0	1.8	
p162	46	2.73	2.73	0	0		2	1	0	0	0	1	2.5	0
p163					1									
p164	41	3.00	3.00	0	0	1.7	2	1	0	1	0	0	2.1	1
p165	31	1.70	1.70	0	0		3	3	1	1	0	0	4.1	0
p166	45	0.76	0.76	0	0	1.5	2	3	0	1	0	0	2.5	1
p167	45	1.93	1.93	0	0	2.0	2	2	1	1	0	0	3.5	0
p168	39	1.14	1.14	0	0	1.8	1	1	1	0	0	0	3.2	
p169	45	0.86	0.86	0	0	1.7	0	2	1	1	0	0	1.2	1
p17	45	16.51	16.51	0	0			0					4	
p170	38	4.29	4.70	0	0		0	3	1	1	0	0	2.5	0
p18	42	10.15	10.15	0	1									
p19	32	16.73	16.73	0	0	2.0	2	2	0	0	0	0	2.8	1
p2	32	28.61	28.61	0	0								1.8	
p20	41	17.78	17.78	0	0	2.0	0				0	0	1.5	
p21	37	8.57	8.57	0	0		0	3	0	1	0	0	1.6	0
p22	42	15.87	15.87	0	0		1	3	1	1	0	0	2	0
p23	41	16.55	16.55	0	0	4.3	3	2	0	0	0	1	1.5	0
p24	35	6.27	6.27	1	1			1	1	1	0	0	1.5	
p25	42	20.48	20.65	0	0			0	0	0	0	0	1.3	
p26					1		2	1						
p27	42	0.54	0.54	1	1		2	3	0	0	0	0	3.5	1
p28	31	6.67	6.67	1	1	1.8	2	2	1	1	0	0	2.5	
p29	31	14.40	14.40	0	0	3.0	3	3	1	1	0	0	3	0
p3	43	7.67	7.67	0	0	1.8	0	3	0	1	0	0		0
p30	33	14.74	14.74	0	0		0	0	0	0	0	0	2	0

p31	37	18.26	20.81	0	0	1.6	1	1	0	0	0	0	2	
p32	35	3.35	3.35	0	1	2.0	2	3	0	0	1	1	2.5	0
p33	45	6.08	6.08	0	0		1	1	0	0	0	0	5	0
p34	33	15.13	15.13	0	0			0	0	1	0	0	2	
p35					1		0							
p36	42	14.49	14.49	0	0	1.4	2	3	0	1	0	0		0
p37	35	4.23	17.97	1	0	4.8	0	3	0	0	0	0	3.8	
p38	35	14.74	14.74	0	0	2.0	3	0	0	0	0	1	2.1	1
p39					1		1	1						
p4	45	26.92	26.92	0	0			1	0	0	0	0	1	
p40	31	14.01	14.01	0	0	1.8	2	0	0	0	0	0	1.8	0
p41	42	11.12	11.12	0	0			3	1	1	0	0	1.5	0
p42	40	12.56	12.56	0	1								1	
p43	40	12.88	12.88	0	0	1.8	0		0	0	0	0	0.1	0
p44	43	11.47	11.59	0	0	1.8	1	1	0	0	0	1	1.5	
p45	41	13.35	13.35	0	0	7.8	2	1	0	0	1	0	1.75	0
p46	41	7.73	7.73	1	1									
p47	37	2.33	3.60	1	1	3.3	2	3	1	0	0	0	1.9	
p48	43	2.14	2.14	1	1	1.5	2		0	0	0	0	4	
p49	34	1.47	1.47	1	1	4.0	2	0	0	0	0	0	1.5	
p5	43	1.00	3.12	1	1	2.0	2	1	0	0	1	0	1.5	
p50	48	0.19	0.19	0	1								1.5	
p51	34	4.43	11.24	0	0	3.1	3	3	1	1	0	0	2.5	0
p52	40	2.73	2.73	0	1	1.7	3			0	0	0	0.5	1
p53	44	16.65	16.65	0	0	1.9	2	2	0	0	0	1	1.3	1
p54	35	11.69	11.69	0	0								1.5	
p55	26	9.45	9.45	0	0	1.7	3	3	0	0	0	0	1	
p56					1		2	3						
p57	35	10.62	10.62	0	0	3.2	3	2	0	0	0	1	1.9	1
p58	35	11.17	11.17	0	0	1.7	2	0	0	1	0	0	1	
p59	42	8.58	8.58	0	0	1.7	0	1	0	0	0	0		0
p6	35	7.95	7.95	1	1	1.8	2	2	1	1	0	0	1	0
p60	40	2.67	6.00	1	1	2.4	2				0		2.8	
p61	39	2.15	4.10	1	1			0	0	0	0	0	1.8	
p62	35	7.84	7.84	0	0	1.7	2	3	1	1	0	1		1
p63	40	7.60	7.60	0	0		1	3	1	1	0	0	2	0
p64	40	7.56	7.56	0	0		0	3	1	1	0	0	2	0
p65	41	7.11	7.11	0	0	2.3	2	3	0	1	0	0	1	0
p66	32	1.51	1.51	0	0	1.8	0						1.5	
p67	40	14.33	21.41	0	0	2.0	1	2	1	1	0	0		
p68	33	3.19	3.19	0	0		0	3	0	0	0	0	3	1
p69	34	10.13	10.13	0	0	2.3	2	3	0	0	0	1	1.7	0
p7	48	10.29	10.29	0	0		0	2	0	0	0	1		1
p70	42	7.68	7.68	1	0			0	0	0	0	0	1.3	
p71	41	2.27	2.27	1	1								3.5	
p72	37	10.39	10.39	0	0		0	0	0	0	0	0	1.7	
p73	45	11.97	11.97	0	0	2.0		0	0	0	0	0	1	
p74	45	6.82	6.82	0	0		0	0	0	1	0	0	2	
p75	30	4.67	4.67	1	1				0	0	0	0	2.8	
p76	39	8.53	8.53	0	0		1	0	0	0	0	1	2	0

p77	44	1.92	1.92	0	0		1		0	0	0		2	
p78	42	3.72	3.72	1	0			0	0	0	0	0	2	
p79	35	9.41	9.41	0	0	1.6	3	3	1	1	0	0	1.1	0
p8	45	4.67	4.67	1	1	5.5	2	0	0	0	1	1	2	
p80	39	4.80	4.80	0	0	3.3	2	3	0	0	1	0	2	1
p81	38	8.68	8.68	0	0			0	0	0	0	0	3	
p82	49	8.92	8.92	0	0	2.2	1	3	0	0	0	1	1.3	1
p83	40	5.56	5.56	0	0	1.8	1	0	0	0	0	0	1	0
p84	37	6.24	6.24	0	0			0	0	0	0	0	2	
p85	34	0.85	4.15	0	0				0	0	0	0	2.3	
p86	37	0.58	0.58	0	0	6.2	2	1	0	0	0	1	1.5	0
p87	42	6.75	6.75	1	0	3.2	3	3	1	1	0	0	1.5	0
p88	40	0.79	0.79	0	0		0	1	0	0	0	0	1	0
p89	39	7.95	7.95	0	0	1.5	1	2	1	1	0	0	1.2	0
p9	43	16.05	16.05	0	0	2.0	0	2	1	0	0	0	1.5	0
p90	33	10.81	10.81	1	0	3.8	1	3	0	0	0	1	4.6	1
p91	42	1.73	2.68	1	0								2.2	
p92	36	0.68	0.68	0	0			0	0	0	0	0	1.2	0
p93	43	10.62	10.62	1	0	2.0	0	2	1	1	0	0	2.5	1
p94	42	9.38	9.38	0	0		1			0	0	0	1	
p95	42	2.07	5.13	1	1								4	
p96	41	10.97	10.97	0	0	1.8	2	0	0	1	0	0	3	0
p97	36	6.91	6.91	0	0	3.5	2	3	1	0	0	0	1.3	0
p98	25	1.58	1.58	0	0	2.7	2	0	0	1	1	1	2.7	0
p99	39	1.39	1.39	0	0		1	3	1	1	0	0	1	1

Notes:

MFS-time: Distant metastasis-free survival time (years).

CSS-time: Cancer-specific survival time (years).

Metastasis: Distant metastasis event. 1 = yes; 0 = no.

Death: Cancer-specific death. 1 = yes. 0 = no.

8q22: average 8q22 copy number detected by FISH.

MTDH:MTDH immunostaining level. 0 = negative; 1 = negligible; 2 = moderate; 3 = intense.

CCNE2: CCNE2 immunostaining level. 0 = negative; 1 = negligible; 2 = moderate; 3 = intense.

ER: ER expression status. 1 = pos; 0 = neg.

PR: progesterone receptor expression status. 1 = pos; 0 = neg.

HER: HER2 expression status. 1 = pos; 0 = neg.

P53: P53 expression status. 1 = pos; 0 = neg.

P_size: Primary tumor pathology size (cm).

CK5/6:Cytokeratin 5/6 immunostaining level. 1 = pos; 0 = neg.

Table S5. Cox Hazard Ratio Prognosis Analysis of Chromosome 8 Genes Using the Three Published Data Sets

(See Excel workbook.)

Table S6. Computational Analysis of Chemoresistance for Chromosome 8 Genes Using the NCI 60 Data Set

(See Excel workbook.)

Table S7. Microarray Data of Genes with Altered Expression after *MTDH* Knockdown in LM2 Cancer Cells

	Gene	Description	Cancer Cell Alone			Coculture with HMVEC-L	
			KD1/control	KD2/control	p values	KD1/control	p values
MTDH upregulated genes	*EGR1*	Homo sapiens early growth response 1 (EGR1), mRNA [NM_001964]	0.38	0.62	0.0001	0.20	0.0046
	FOS	Homo sapiens v-fos FBJ murine osteosarcoma viral oncogene homolog (FOS), mRNA [NM_005252]	0.39	0.55	0.0127	0.25	0.0017
	ALDH3A1	Homo sapiens aldehyde dehydrogenase 3 family, memberA1 (ALDH3A1), [NM_000691]	0.30	0.37	0.0193	0.31	0.0001
	AK123083	Homo sapiens cDNA FLJ41088 fis, clone ASTRO2002459. [AK123083]	0.17	0.54	0.0019	0.20	0.0053
	BC009463	Homo sapiens, clone IMAGE:3606519, mRNA, partial cds. [BC009463]	0.16	0.40	0.0045	0.22	0.0070
	AY029066	Homo sapiens Humanin (HN1) mRNA, complete cds. [AY029066]	0.31	0.48	0.0079	0.37	0.0368
	AY358690	Homo sapiens clone DNA62876 LPPA601 (UNQ601) mRNA, [AY358690]	0.31	0.27	0.0064	0.34	0.0050
	HNRPU	Homo sapiens heterogeneous nuclear ribonucleoprotein U (scaffold attachment factor A) (HNRPU), mRNA [NM_004501]	0.29	0.49	0.0050	0.34	0.0062
	MTDH	Homo sapiens metadherin (MTDH), mRNA [NM_178812]	0.19	0.25	0.0035	0.29	0.0016
	THC2269172	Unknown	0.19	0.14	0.0063	0.39	0.0049
	ADAMTS1	Homo sapiens ADAM metallopeptidase with thrombospondin type 1 motif, 1 (ADAMTS1), mRNA [NM_006988]	0.20	0.25	0.0014	0.20	0.0063
	HBB	Homo sapiens hemoglobin, beta (HBB), mRNA [NM_000518]	0.30	0.32	0.0055	0.27	0.0002
	AK056809	Homo sapiens cDNA FLJ32247 fis, clone PROST1000120. [AK056809]	0.32	0.39	0.0133	0.39	0.0204
	HMOX1	Homo sapiens heme oxygenase (decycling) 1 (HMOX1), [NM_002133]	0.14	0.18	0.0125	0.22	0.0000
	KCNJ14	Homo sapiens potassium inwardly-rectifying channel, subfamily J, member 14 (KCNJ14), [NM_170720]	0.32	0.27	0.0001	0.31	0.0045
	MT-ND6	Homo Sapiens mitochondrially encoded NADH dehydrogenase 6 (ENST00000361681)	0.32	0.36	0.0005	0.36	0.0003
	SRRM1	Homo sapiens serine/arginine repetitive matrix 1 (SRRM1), mRNA [NM_005839]	0.31	0.45	0.0254	0.31	0.0057
	NOL8	Homo sapiens nucleolar protein 8 (NOL8), mRNA [NM_017948]	0.33	0.41	0.0096	0.38	0.0066
	HSP90AB1	Homo sapiens heat shock protein 90kDa alpha (cytosolic), class B member 1 (HSP90AB1), mRNA [NM_007355]	0.28	0.35	0.0030	0.31	0.0139
	HSP90AB3P	Homo sapiens heat shock protein 90Bc (HSP90Bc) mRNA, [AY956764]	0.30	0.36	0.0005	0.34	0.0116
	CD3EAP	Homo sapiens CD3e molecule, epsilon associated protein (CD3EAP), mRNA [NM_012099]	0.30	0.26	0.0070	0.34	0.0007
	CTGF	Homo sapiens connective tissue growth factor (CTGF), mRNA [NM_001901]	0.31	0.30	0.0005	0.23	0.0011
	MET	Homo sapiens met proto-oncogene (hepatocyte growth factor receptor) (MET), mRNA [NM_000245]	0.32	0.33	0.0014	0.23	0.0030

MTDH downregulated genes	*HLA-DRA*	Homo sapiens major histocompatibility complex, class II, DR alpha (HLA-DRA), mRNA [NM_019111]	3.29	3.28	0.0067	2.52	0.0000
	RASD1	Homo sapiens RAS, dexamethasone-induced 1 (RASD1), mRNA [NM_016084]	5.72	16.76	0.0476	2.57	0.0113
	CXCL1	Homo sapiens chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha) (CXCL1), mRNA [NM_001511]	30.79	36.60	0.0024	3.13	0.0250
	STMN3	Homo sapiens stathmin-like 3 (STMN3), mRNA [NM_015894]	4.67	6.09	0.0038	2.66	0.0011
	TIMP3	Homo sapiens TIMP metallopeptidase inhibitor 3 (Sorsby fundus dystrophy, pseudoinflammatory) (TIMP3), mRNA [NM_000362]	6.86	8.25	0.0145	3.26	0.0011
	ABCA1	Homo sapiens ATP-binding cassette, sub-family A (ABC1), member 1 (ABCA1), mRNA [NM_005502]	5.78	6.41	0.0001	2.89	0.0024
	BNIP3	Homo sapiens BCL2/adenovirus E1B 19kDa interacting protein 3 (BNIP3), nuclear gene encoding mitochondrial protein, mRNA [NM_004052]	3.67	5.07	0.0051	3.24	0.0215
	PPAP2B	Homo sapiens phosphatidic acid phosphatase type 2B (PPAP2B), transcript variant 1, mRNA [NM_003713]	3.10	2.74	0.0066	3.03	0.0079
	GPR56	Homo sapiens G protein-coupled receptor 56 (GPR56), transcript variant 3, mRNA [NM_201525]	3.91	3.17	0.0142	3.45	0.0002
	TRAIL	Homo sapiens tumor necrosis factor (ligand) superfamily, member 10 (TNFSF10), mRNA [NM_003810]	3.81	1.98	0.0142	3.37	0.0000

* Data represent analysis from three replicate experiments.

Table S8. Regions Detected by ACE in Bladder Tumors Compared to Normal Samples in the Data by Stransky et al.

Region No	chr	start pos	end pos	start-band	end-band	cytoband	No. of Tumors with CNAs detected by aCGH (total 57 tumors)		Epigenetic region reported by Stransky et al. (Region No. in parenthesis)
							Gain	Loss	
Regions of Gain									
1	1	147880413	154500530	q21	q21	q21-q22	19	0	
2	2	8740840	17430600	p25	p24	p25-p24	9	1	
3	2	73301373	74543858	p13	p13	p13	8	1	
4	2	183554951	190407593	q32	q32	q32	2	5	
5	2	233707101	236383177	q37	q37	q37	0	9	
6	4	56032174	68869767	q12	q13	q12-q13	0	6	
7	5	32134948	38926133	p13	p13	p13	11	1	
8	6	26231920	28161025	p22	p22	p22	8	1	
9	7	636514	5995824	p22	p22	p22	13	0	
10	7	92191481	97329522	q21	q21	q21	13	1	
11	7	100653396	101928102	q22	q22	q22	12	1	
12	8	49127738	72917665	q11	q13	q11-q13	16	3	
13	8	101603149	103414206	q22	q22	q22	15	2	
14	8	117941775	125562990	q24	q24	q24	18	3	
15	10	6585689	43420464	p15	q11	p15-q11	12	4	
16	10	51738231	59798667	q11	q21	q11-q21	3	7	
17	12	75786778	86967256	q21	q21	q21	0	1	
18	12	101393624	107209809	q23	q23	q23	3	3	
19	13	99880793	112939874	q32	q34	q32-q34	8	3	
20	14	18684819	20319767	q11	q11	q11	3	6	
21	16	54973658	55217871	q13	q13	q13	4	11	
22	17	53685949	58091732	q22	q23	q22-q23	13	2	
23	22	18747065	21319513	q11	q11	q11	7	5	
24	X	54486137	62630872	p11	q11	p11-q11	18	1	
25	X	129109214	134250291	q25	q26	q25-q26	16	1	
26	X	150883029	151864997	q28	q28	q28	16	9	
Regions of Loss (or Epigenetic Regulation)									
1	1	23525986	25113799	p36	p36	p36			1p36.33-p36.11, (1-1)
2	1	205147907	206375188	q32	q32	q32	4	4	
3	3	1264953	9775305	p26	p25	p26-p25	11	7	
4	3	37321511	40327566	p22	p22	p22			3p22.3, (3-2)
5	4	184666015	187813961	q35	q35	q35	1	12	
6	6	29905182	33278438	p21	p21	p21			6p21.32, (6-2)
7	6	154563456	163362757	q25	q26	q25-q26	1	6	
8	8	27305648	30079033	p21	p12	p21-p12	5	20	
9	9	70662172	72121416	q21	q21	q21	2	22	
10	9	96887301	99655278	q22	q22	q22	4	22	
11	9	129589973	135898080	q34	q34	q34	4	19	
12	10	93380076	96077103	q23	q23	q23	2	14	
13	11	389566	10284376	p15	p15	p15	3	20	
14	14	22933845	23763661	q11	q12	q11-q12	3	6	
15	14	62995492	64560904	q23	q23	q23	3	8	
16	14	73602925	74223618	q24	q24	q24	4	9	
17	16	20694008	28281818	p12	p11	p12-p11	8	7	
18	17	7085469	7430061	p13	p13	p13	4	16	
19	17	15631906	17941997	p12	p11	p12-p11	5	14	
20	19	15381262	18147806	p13	p13	p13			19p13.12, (19-3)
21	19	44570371	46529259	q13	q13	q13	7	0	
22	22	35199934	44532474	q12	q13	q12-q13	5	5	

Table S9. Primers Used in the qPCR for DNA Copy Number and Gene Expression Analysis

	Gene	5' Primer	3' Primer	Gene Position
DNA Copy Number	*MTDH*	TGGCAAATGTGGCCAACA	TATTAGGTAACCGACCCCCTCTT	8q22
	LAPTM4b	GAGTCTCACTCTGTGGCCCG	TGAGCAGAGATCATGCCATTG	8q22
	CA2	CTCAGCACGAACTGTCCCG	CCATGGATTCAAACCAGCACT	8q21
	EIF3S6	GCAATTGCAAGACGGCTGT	AAACCTTGGCTTACCCAGGAA	8q23
	APP	TCAGGTTGACGCCGCTGT	TTCGTAGCCGTTCTGCTGC	21q21, as control
RNA expression	*MTDH*	TCCGAGAAGCCCAAACCAAAT	CTTCACCCTCAGCCACTTCAA	8q22
	LAPTM4b	TTTTATTGAGTGCCCTGGCTG	GGCAATGCACATGTTGGC	8q22
	PTDSS1	AAGTGGAGGACATCACCATTG	TCATCCCTGGTAAAGGCG	8q22
	ALDH3A1	AGCTGAGTGAGAACATGGCGA	ATGGTCGAACCTCTCCTTGAGC	17p11.2
	MET	TCGCTTCATGCAGGTTGTGG	TGTCTGCAGCCCAAGCCAT	7q31

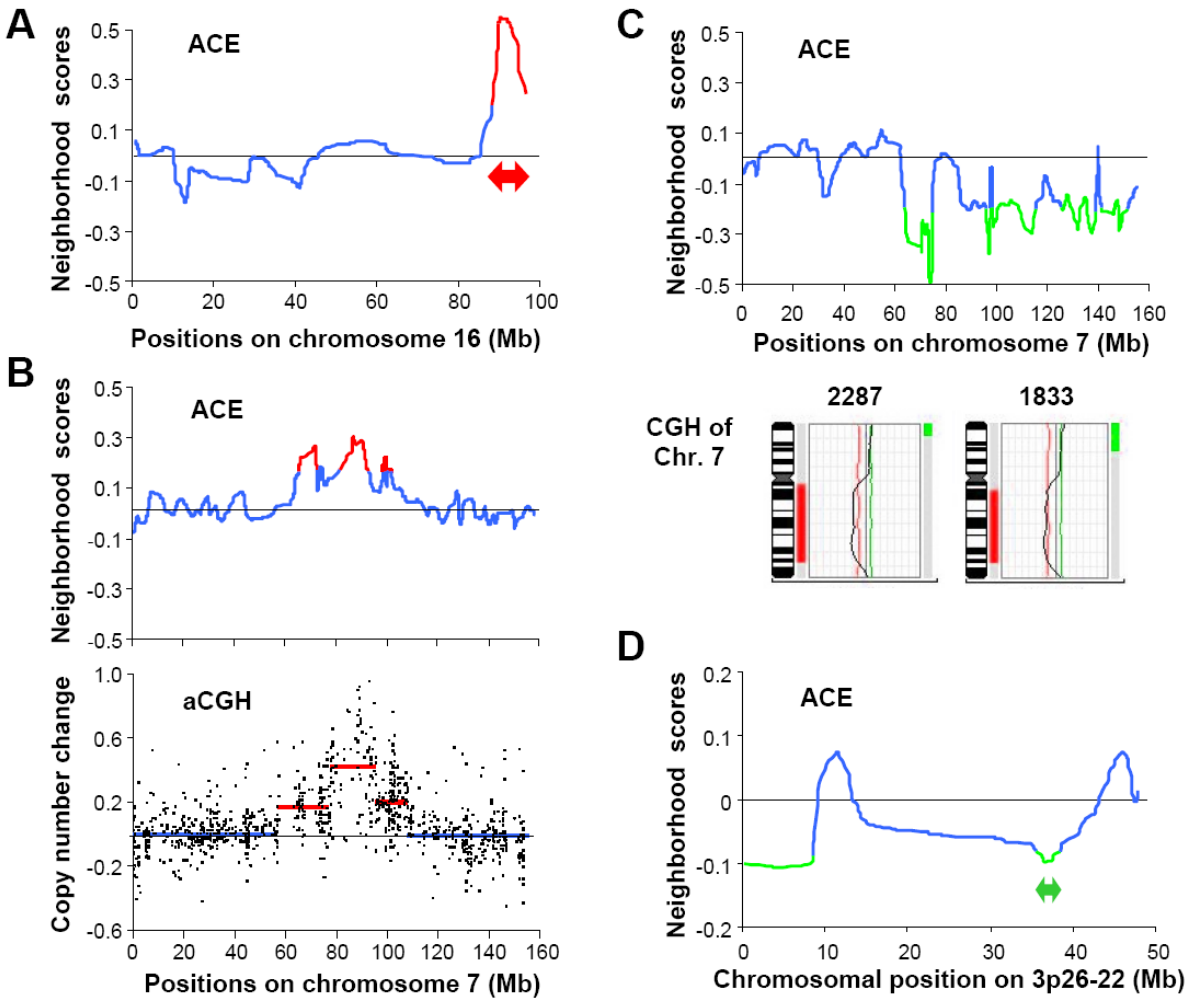

Figure S1. Validation of the ACE Algorithm Using Available Expression Data with Corresponding Genomic Alteration Data

(A) Expression microarray data of the brain tissue from Ts1Cje mice, the Down Syndrome animal model known to have a partial trisomic region on chromosome 16, were compared to that of the normal mice (Amano et al., 2004). ACE predicted a sole CNA region on chromosome 16. The NS produced by ACE is shown along this chromosome, where the red line indicates the predicted region of gain. Red double-arrow: the known trisomic region in Ts1Cje mice.

(B) Taxane-resistant cells were established by continuous exposure of 6 ovarian cancer cell lines - 1A9, ES-2, MESOV, OVCA429, OVCA433 and OVCAR-3 (Wang et al., 2006) - to docetaxel or paclitaxel. ACE detected 3 amplified regions on chromosome 7 in the taxane-resistant derivatives when compared to their parental lines (upper panel), which were highly consistent with the analysis of aCGH data (lower panel). Colored horizontal lines in the lower panel are the segment means produced by the aCGH analysis tool CBS, of which red indicates the significant regions of gain.

(C) ACE compared the expression of 10 cell lines derived from the breast cancer cell MDA-MB-231 with high or low breast-to-bone metastatic capability (Kang et al., 2003) and defined a loss

at chromosome 7q associated with bone metastasis. The upper panel shows the NS of chromosome 7 in the highly metastatic cells with green lines indicating predicted regions of loss. The lower panel displays the previously published CGH data of the same chromosome in the two highly metastatic lines 2287 and 1833, with DNA of the lowly metastatic parental line MDA-MB-231 used as a control (Kang et al., 2003). Red and green vertical bars indicate regions of genomic loss and gain, respectively.
(D) ACE was used to analyze regional epigenetic regulation using the gene expression data of bladder tumors (Stransky et al., 2006). Partial chromosome 3 is shown. Dark green double-arrow: the epigenetic regulated region that was experimentally validated in a previous study (Stransky et al., 2006). See Table S8 for all significant regions in this data set.

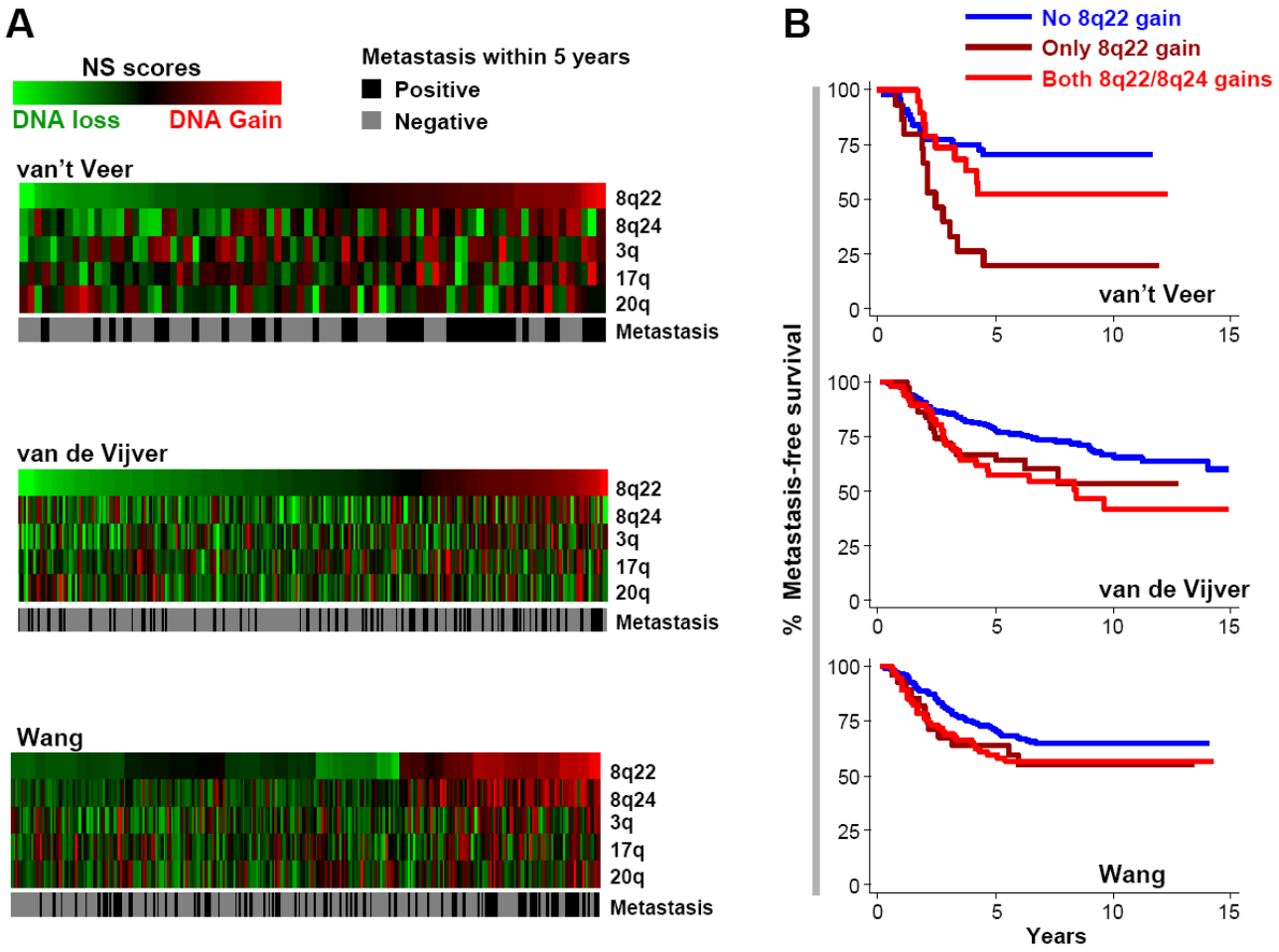

Figure S2. Lack of Interaction between Genomic Gains at 8q22 and 8q24.3 in Poor Prognosis

(A) Tumors in the three data sets were clustered by the genomic status of the 5 identified poor-prognosis regions.
(B) Kaplan-Meier analysis of the samples with genomic gains at both 8q22 and 8q24.3, with gain at only 8q22, and without 8q22 gain.

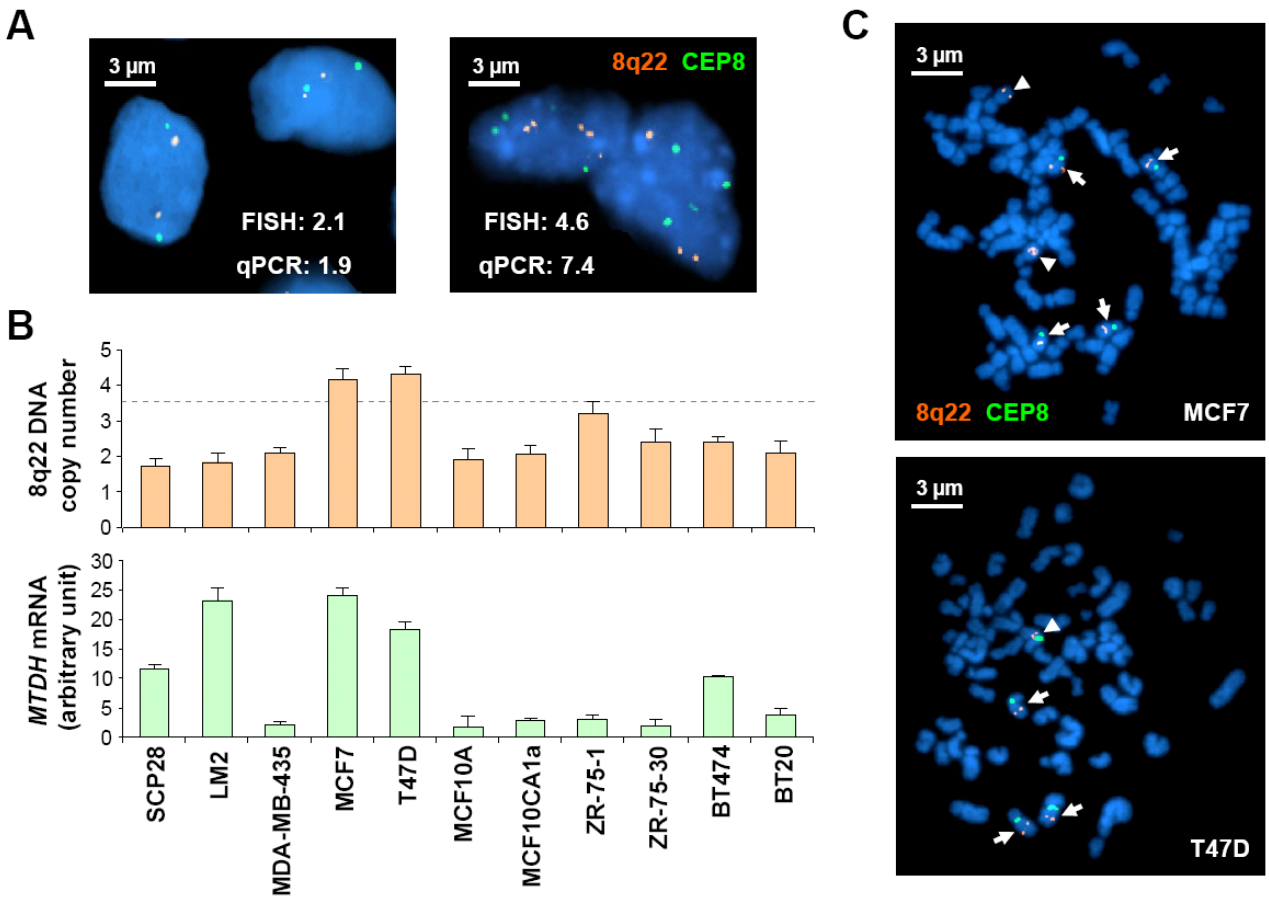

Figure S3. *MTDH* DNA Copy Number and mRNA Quantification in a Panel of Breast Cancer Cell Lines and Breast Tumors

(A) Both genomic qPCR and tissue FISH were performed to analyze breast tumors. The results from both assays were highly consistent. Shown are FISH images of 2 paraffin tissue samples. The average 8q22 copy numbers scored from at least 100 nuclei in FISH and from the genomic qPCR assay are also shown for each sample.

(B) DNA copy numbers and mRNA levels analyzed by qPCR. The horizontal line in the upper panel indicates the qPCR copy number cutoff used in our study and other previous studies (Glockner et al., 2001; Lehmann et al., 2000) to define genomic gain. Two cell lines with 8q22 genomic gain, MCF7 and T47D, also have high level of *MTDH* expression. The 4175 subline of MDA-MB-231 overexpresses *MTDH* although it does not have genomic gain at 8q22. Error bars represent ±SD.

(C) Metaphase FISH analysis of two breast cancer cell lines with *MTDH* genomic gain. Upper panel: In MCF7, 8q22 signals (orange) were found in 4 copies of apparently normal chromosome 8 (arrows) that also showed signals for chromosome 8 centromere (CEP8), in addition to two submetacentric chromosomes with only 8q22 signals (arrow heads). Lower panel: In T47D, 8q22 signals (orange) were found in 3 copies of apparently normal chromosome 8 (arrows) and one der(8) with both CEP8 and 8q22 signals (arrow head).

In (A) and (C), Green SpectrumGreen and red SpectrumOrange probes detect chromosome 8 centromere and the 8q22 region, respectively.

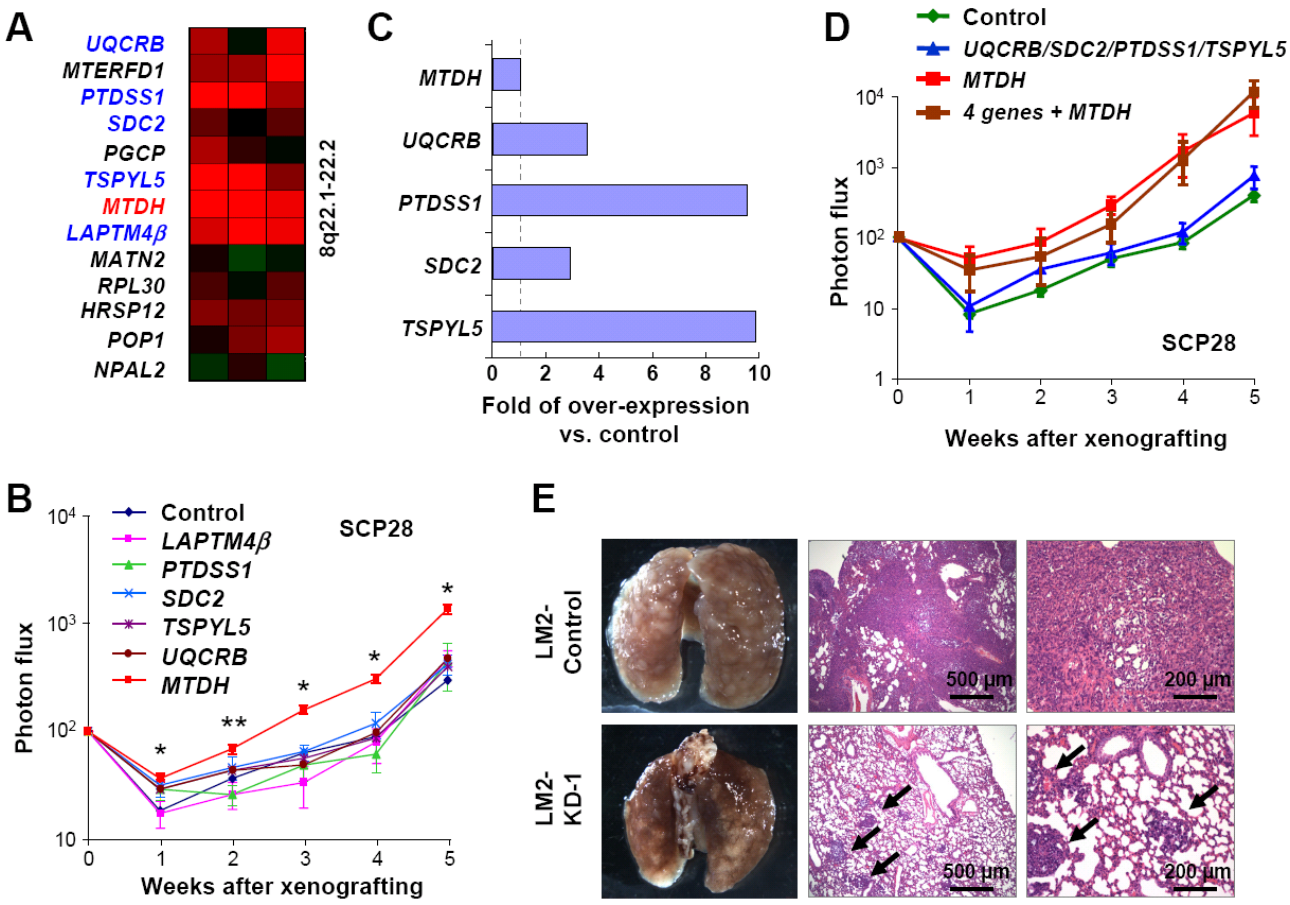

Figure S4. Functional Analysis of 8q22 Genes Identifies *MTDH* as the Target Gene of the Genomic Gain to Promote Metastasis

(A) Differential expression patterns of genes at 8q22 in patients with poor prognosis compared to those with good prognosis. To identify the target gene(s) among those, six putative candidates including *MTDH* (color highlighted) with the expression pattern most strongly correlated with prognosis or previously implicated in tumor biology were chosen and their possible roles in promoting metastasis were analyzed using the xenograft animal model.

(B) SCP28 cells with expression of each of the six genes as well as the empty vector were tested for their metastatic capability. The cells were injected into the nude mice intravenously, and the animal lung metastasis burden was monitored by bioluminescent imaging. Shown are the normalized luminescent signals from the cancer cells colonized in lung. Only *MTDH* overexpression led to significant increase of lung metastasis. *$p < 0.05$; **$p < 0.01$, two-sided Wilcoxon rank test to compare *MTDH* overexpression vs. control.

(C) Retroviruses encoding *UQCRB*, *PTDSS1*, *SDC2* and *TSPYL5* were pooled and used to infect SCP28 to achieve simultaneous overexpression of these four genes, as validated by qRT-PCR analysis.

(D) Xengraft assays of the cells with combinational overexpressed in SCP28 cells with or without *MTDH* overexpression did not show an increase of lung metastasis.

(E) Photographs and hematoxylin/eosin stain sections of representative lungs harvested at necropsy from mice injected with control and *MTDH*-knockdown LM2 cells.

In (B) and (D), error bars represent ±SEM.

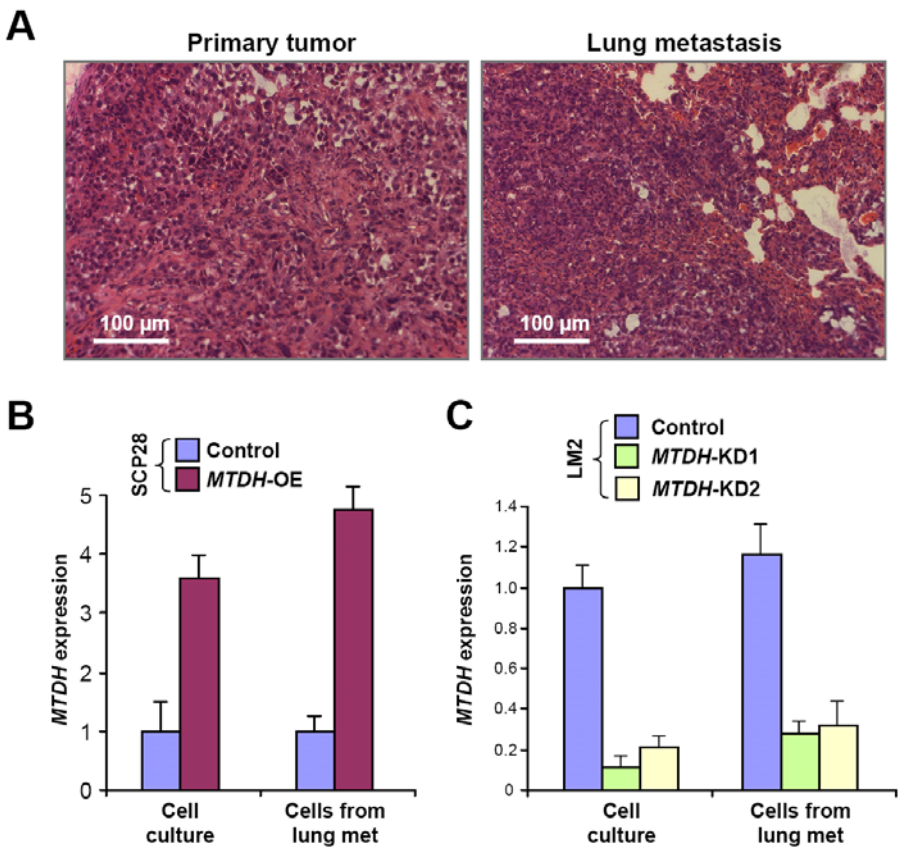

Figure S5. Analysis of Primary Tumors and Lung Metastases Generated by Variants of the MDA-MB-231 Cell Line

(A) H&E staining of primary tumors or lung metastases generated by LM2 cells.

(B and C)qRT-PCR analysis of *MTDH* expression levels in SCP28 (B) or LM2 cells (C) isolated from cell culture or xenograft lung metastases. Error bars represent ±SD.

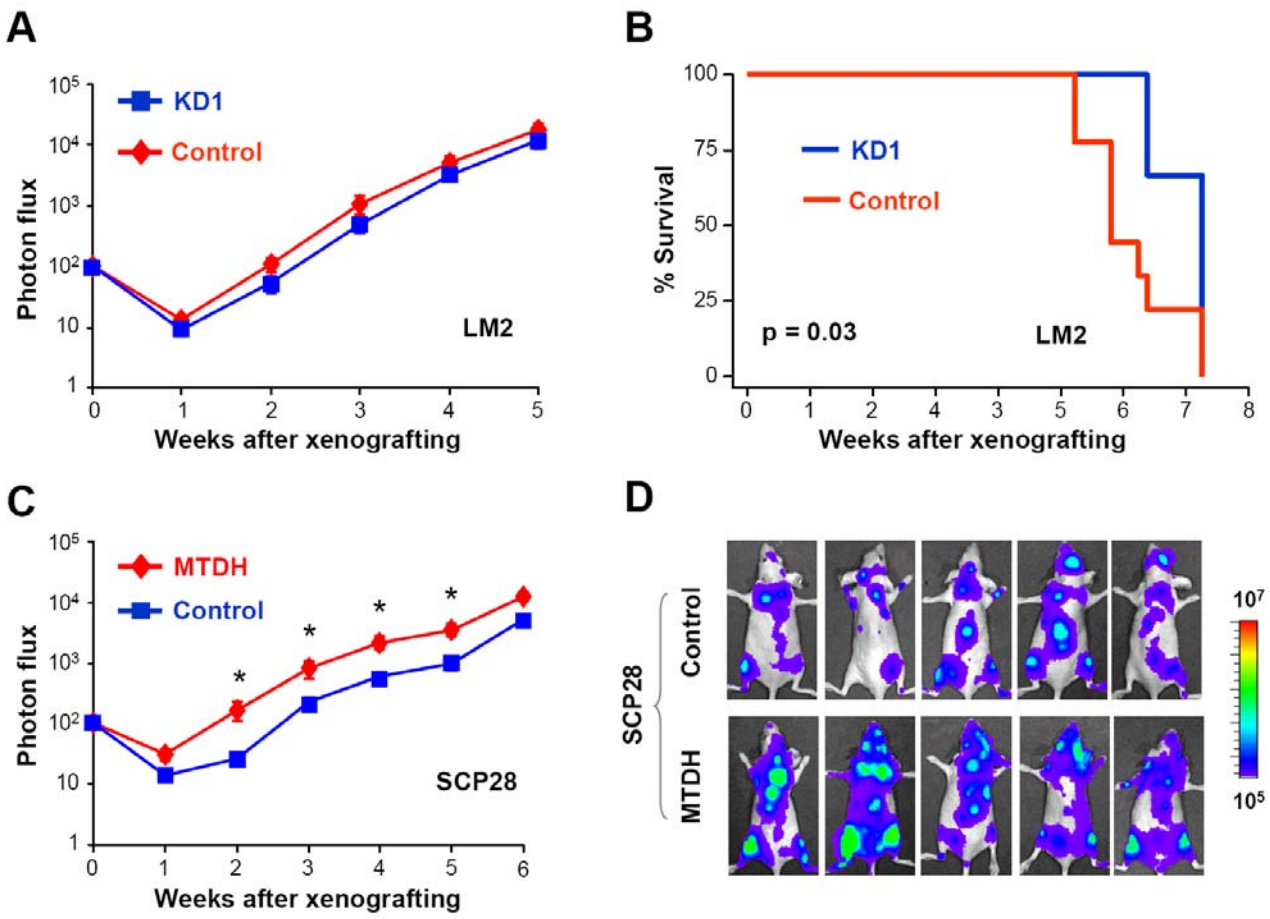

Figure S6. Organ-Specific Metastasis Mediated by MTDH

(A-C) LM2 cells with *MTDH* knockdown were inoculated via intracardiac injection into nude mice to generate bone and brain metastasis. A slight decrease in bone metastasis (A) and a modest but significant improvement of animal survival (B) were observed. n = 10. Reciprocally, *MTDH* overexpression in SCP28 led to a significant increase in bone metastasis propensity (C). n = 10. *p < 0.05 based on a two-sided Wilcoxon rank test.

(D) Representative BLI images of systemic metastasis burden in mice injected with SCP28 control and *MTDH*-overexpression cells.

In (A) and (C), error bars represent ±SEM.

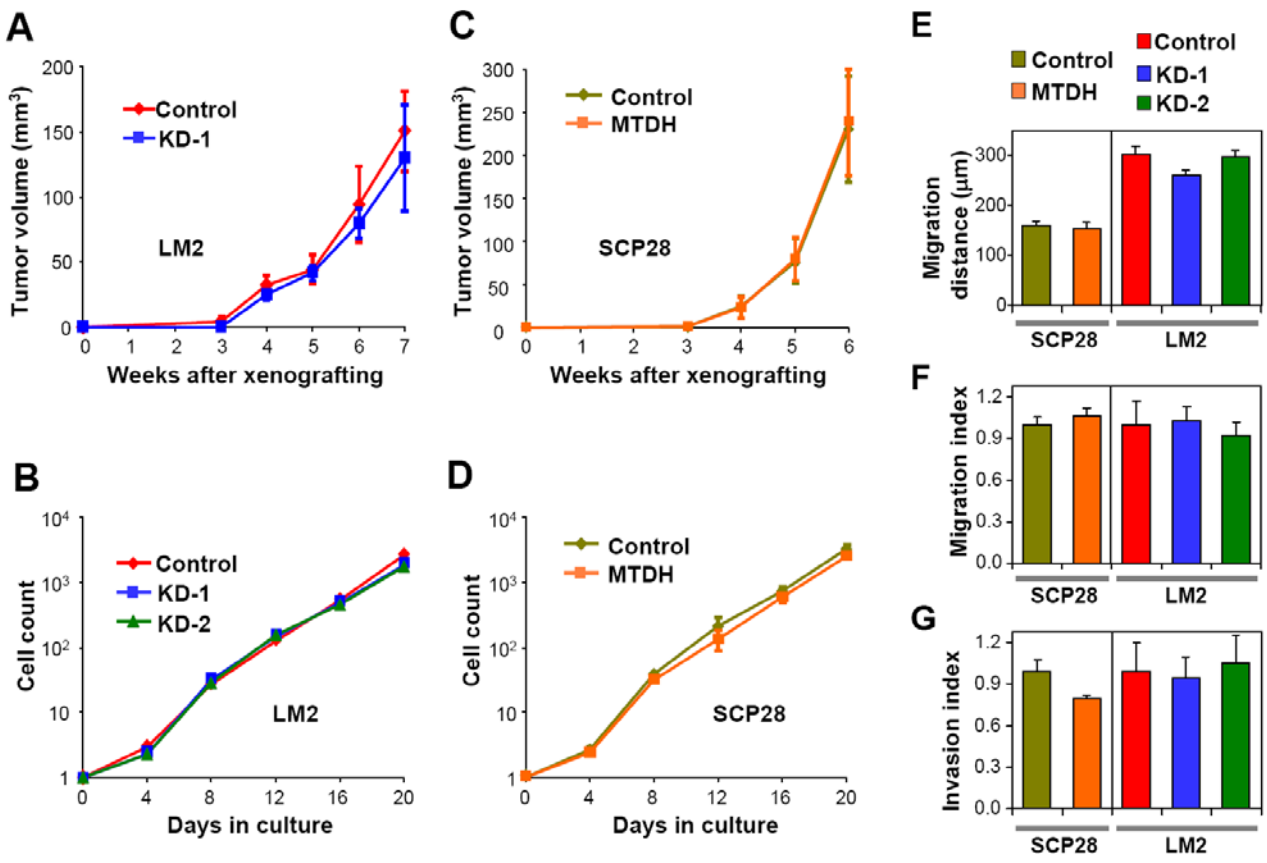

Figure S7. MTDH Does Not Influence the Growth, Migration, or Invasion of Tumor Cells
(A) LM2 cells with *MTDH* knockdown or control hairpin expression were inoculated into the #4 mammary fat pad of nude mice. Length and width of the primary tumors were measured, and the tumor volumes were calculated at the indicated time points.
(B) The *in vitro* proliferation rates of LM2 cells were not affected by *MTDH* knockdown.
(C) The growth curve of the SCP28 control or *MTDH* overexpression cells after inoculation into mammary fat pads.
(D) The *in vitro* proliferation rates of SCP28 cells.
In (A)-(D), error bars represent ±SEM.
(E-G) Alteration of *MTDH* expression in LM2 or SCP28 cells did not lead to change of migration and invasion properties of the cancer cells as measured by wound healing assays (E), Boyden two-chamber migration assay (F), and two-chamber Matrigel invasion assay (G).
In (E)-(G), Error bars represent ±SD.

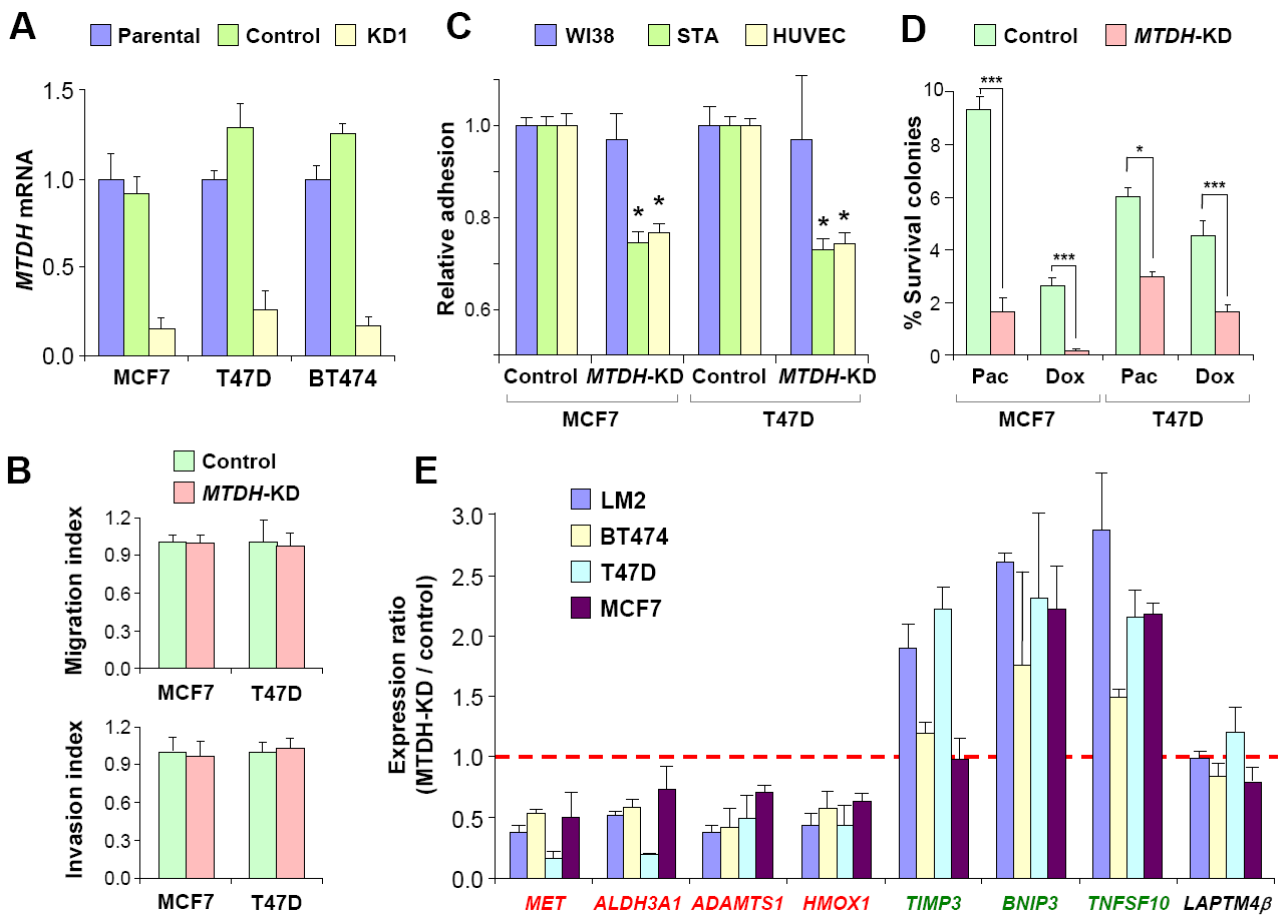

Figure S8. Functional Validation of MTDH in Additional Breast Cancer Cell Lines

(A) qPCR confirmation of *MTDH* knockdown in three breast cancer cell lines.

(B) Migration and invasion assays of MCF7 and T47D.

In (A) and (B), error bars represent ±SD.

(C) Endothelial adhesion assays with endothelial cells from lung (STA) (Krump-Konvalinkova et al., 2001) or umbilical vein (HUVEC), and control fibroblast cells (WI38).

(D) Relative colony numbers after chemo-therapeutic treatments normalized to non-treatment culture. Pac: paclitaxel; Dox: doxorubicin.

In (C) and (D), data represent average values ± SEM. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$ with a two-sided Student's t-test.

(E) Expression levels of genes regulated by MTDH were analyzed by qPCR in several breast cancer cell lines with or without *MTDH* knockdown. Genes in red are upregulated by MTDH, and genes in green downregulated by MTDH. *LAPTM4β*, a non-related gene, serves as control. All expression levels are normalized to *GAPDH*. Error bars represent ±SD.

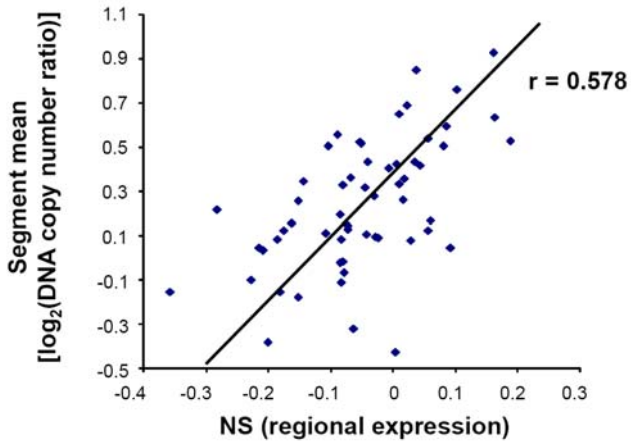

Figure S9. Correlation of 8q22 Copy Number and NS in NCI60 Cell Lines
8q22 DNA copy numbers are positively correlated with the gene expression levels of this region in the 58 human cancer cell lines of the NCI60 data. The 8q22 copy numbers were analyzed from SNP microarray data using the CBS algorithm and shown as the segment mean values. The overall 8q22 gene expression pattern is calculated as the neighborhood scores (NS) using the ACE algorithm.

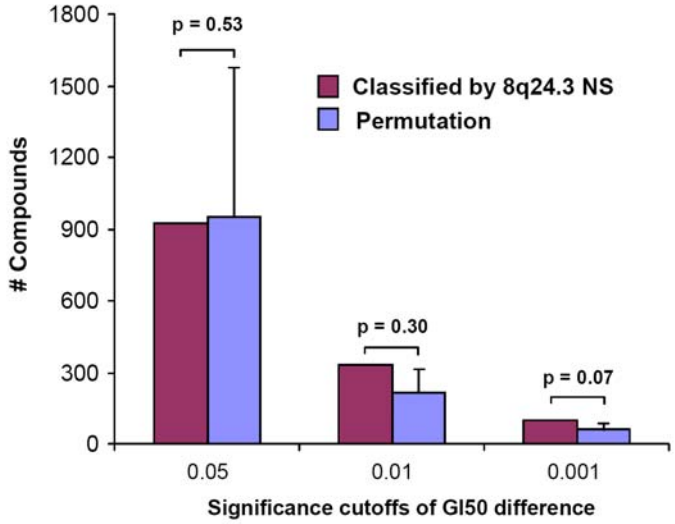

Figure S10. Genomic Gain at 8q24.3 Is Not Associated with Higher Resistance to Chemical Compounds in the NCI60 Human Cancer Cell Lines
ACE analysis identified 8q24.3 gain as another poor-prognosis-associated genomic event on the chromosome 8q arm. To investigate whether this region also contributes to chemoresistance, $logGI_{50}$ (drug concentration for 50% growth inhibition) of each of the 24,642 compounds in cell lines with 8q24.3 gain was compared to those in cells without genomic gain. The number of compounds with significantly ($p < 0.05$, $p < 0.01$ and $p < 0.001$) increased $logGI_{50}$ in cells of 8q24.3 gain was compared to a null distribution obtained by permuting the 8q24.3 copy numbers of the cell lines. Median values from permutations are shown with mean absolute deviation (MAD) as the error bar.

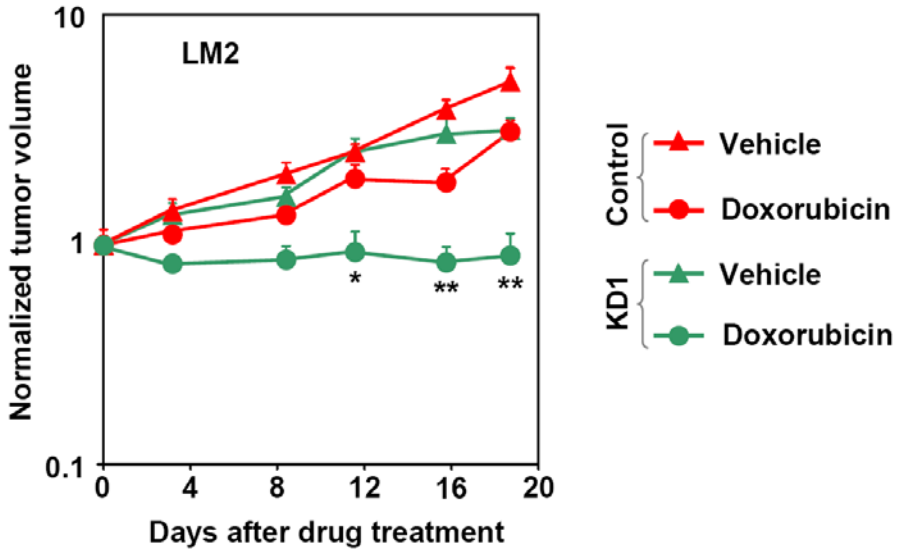

Figure S11. In Vivo Chemoresistance Assay with Doxorubicin Treatment
Shown are the xenograft tumor sizes from control LM2 or *MTDH*-KD cells when mice were treated with doxorubicin or drug vehicle. Data represent average ±SEM of 6 mice per group. *$p < 0.05$; **$p < 0.01$ with a two-sided Student's t-test to compare KD-1 cells with and without doxorubicin treatment. $p = 0.022$ with ANOVA analysis of repeated measurement to compare the whole growth curves of these two conditions.

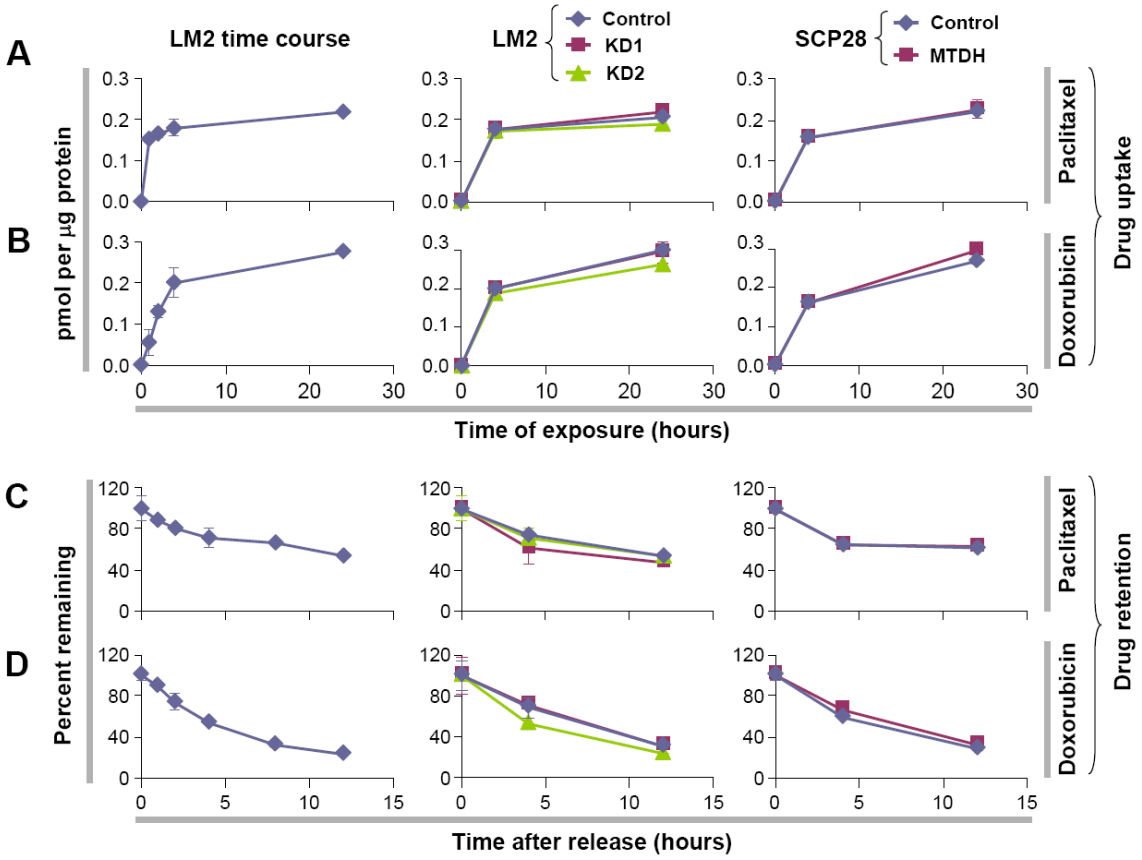

Figure S12. Drug Uptake and Retention in Cells with Modified *MTDH* Expression
(A and B) Drug update assay of paclitaxel (A) and doxorubicin (B) in LM2 parent cells (left panel), LM2 vector control and *MTDH* knockdown (middle panel), and SCP28 cells with *MTDH* overexpression and vector control (right panel). Cells were treated with radio-labeled paclitaxel or doxorubicin for up to 24 h and were harvested immediately after the indicated period of drug exposure. Drug uptake in the cells was measured by liquid scintillation counting. Results were normalized with cellular protein amount measured by Bradford assay and expressed as average ± SD of three replicates.
(C and D) Drug retention assay for paclitaxel (C) and doxorubicin (D) in various cell lines as in (A) and (B). For the retention study, cells were incubated with drug-containing medium for 4 h, followed by incubation in 2 ml drug-free medium for the indicated time and then harvested. Drug retention in the cells was measured by liquid scintillation counting and normalized with cellular protein amount measured by Bradford assay. Results were expressed as percentage of remaining drugs as compared to the amount at the end of exposure to drug-containing media and shown as average ± SD of three replicates.

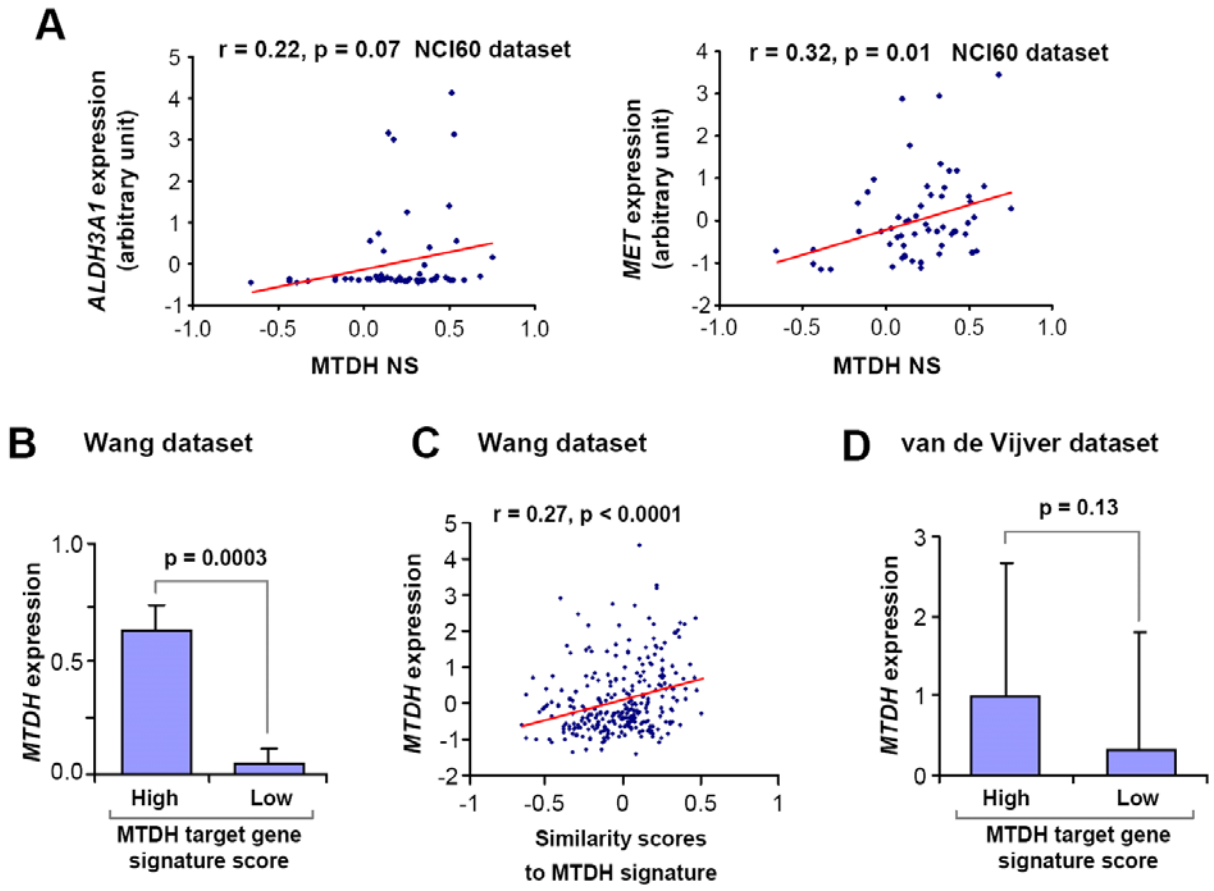

Figure S13. Correlation of MTDH Downstream Genes with *MTDH* Expression in NCI60 Cells and Primary Tumors from Human Patients

(A) Correlation of *ALDH3A1* and *MET* expression with *MTDH* status in NCI60 cell lines.

(B) Tumor samples in Wang et al. data set were classified as "high" and "low" MTDH target gene signature score groups according to the expression pattern of the MTDH-regulated gene set (24-genes mapped to the array platform). The observed *MTDH* expression was compared in the two groups.

(C) A similarity score was assigned to each tumor in the Wang et al. data set by comparing the expression pattern of MTDH-regulated genes to that of *MTDH*-expressing cell lines, and the scores were correlated to the observed *MTDH* levels in tumors.

(D) Sample classification was performed for the van de Vijver et al. data set as in (B). However, only 15 genes in the MTDH-regulated set could be mapped to the Rosetta microarray platform and were used for classification.

In (B) and (D), error bars represent ±SD.

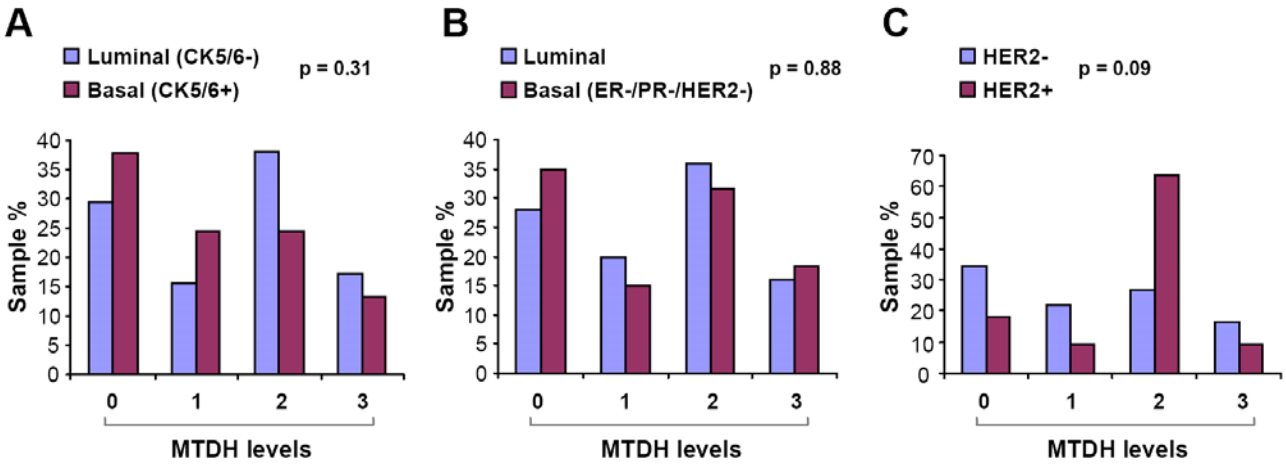

Figure S14. MTDH Overexpression Is Not Linked to Particular Cancer Subtypes
A breast cancer tissue microarray was immunostained with antibodies against MTDH and other markers. MTDH expression levels were compared in different cancer subtypes classified by status of CK5/6 (A), triple markers (ER/PR/HER2) (B), or HER2 (C). χ^2 test p values were shown.

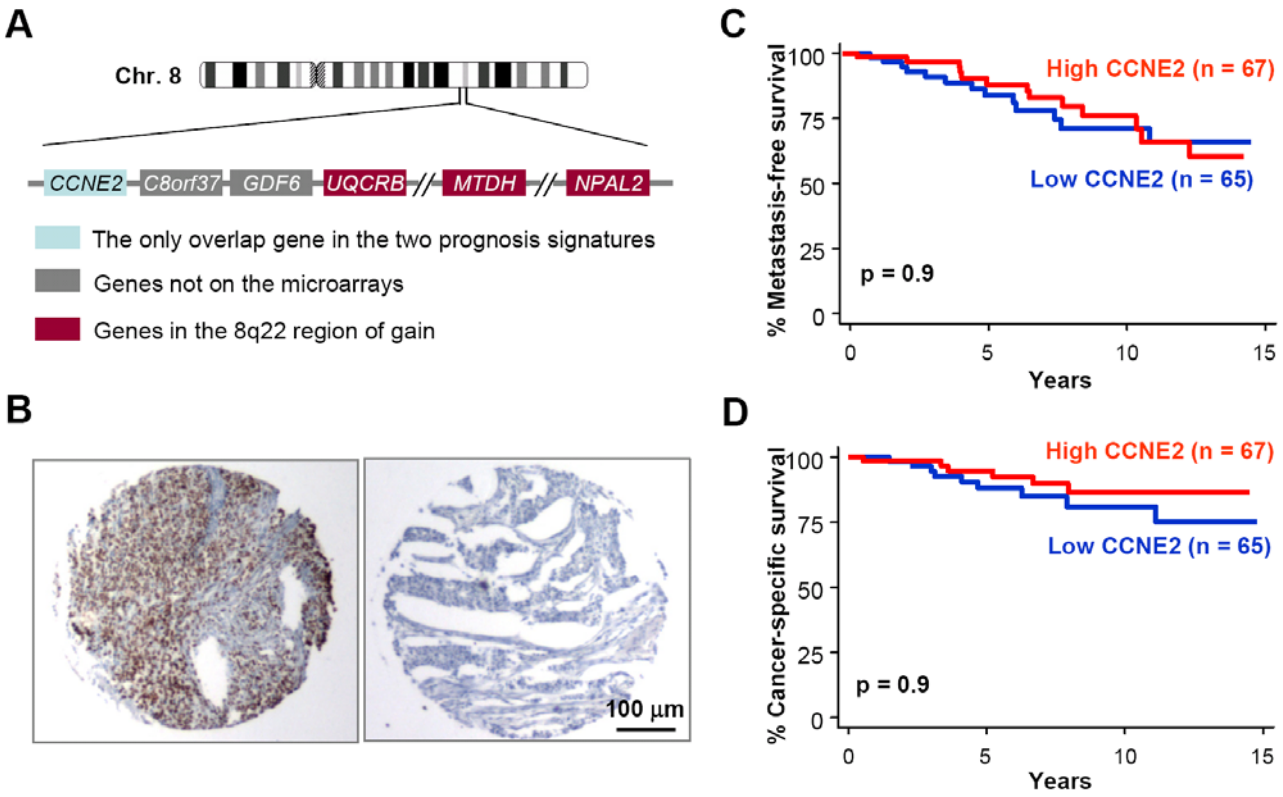

Figure S15. CCNE2 Is Not Associated with Clinical Outcomes in the Breast Cancer Tissue Array Analysis

(A) *CCNE2* is the only overlapping gene in the poor- prognosis gene signatures by van't Veer *et al*. and Wang *et al*., and is located immediately upstream of the 8q22 region of gain.

(B) A human beast cancer tissue array was stained with a CCNE2 antibody. A case of high CCNE2 expression (left) and a case of low CCNE2 expression (right) are shown.

(C and D) Kaplan-Meier analysis of patient metastasis and survival shows no significance of CCNE2 expression.

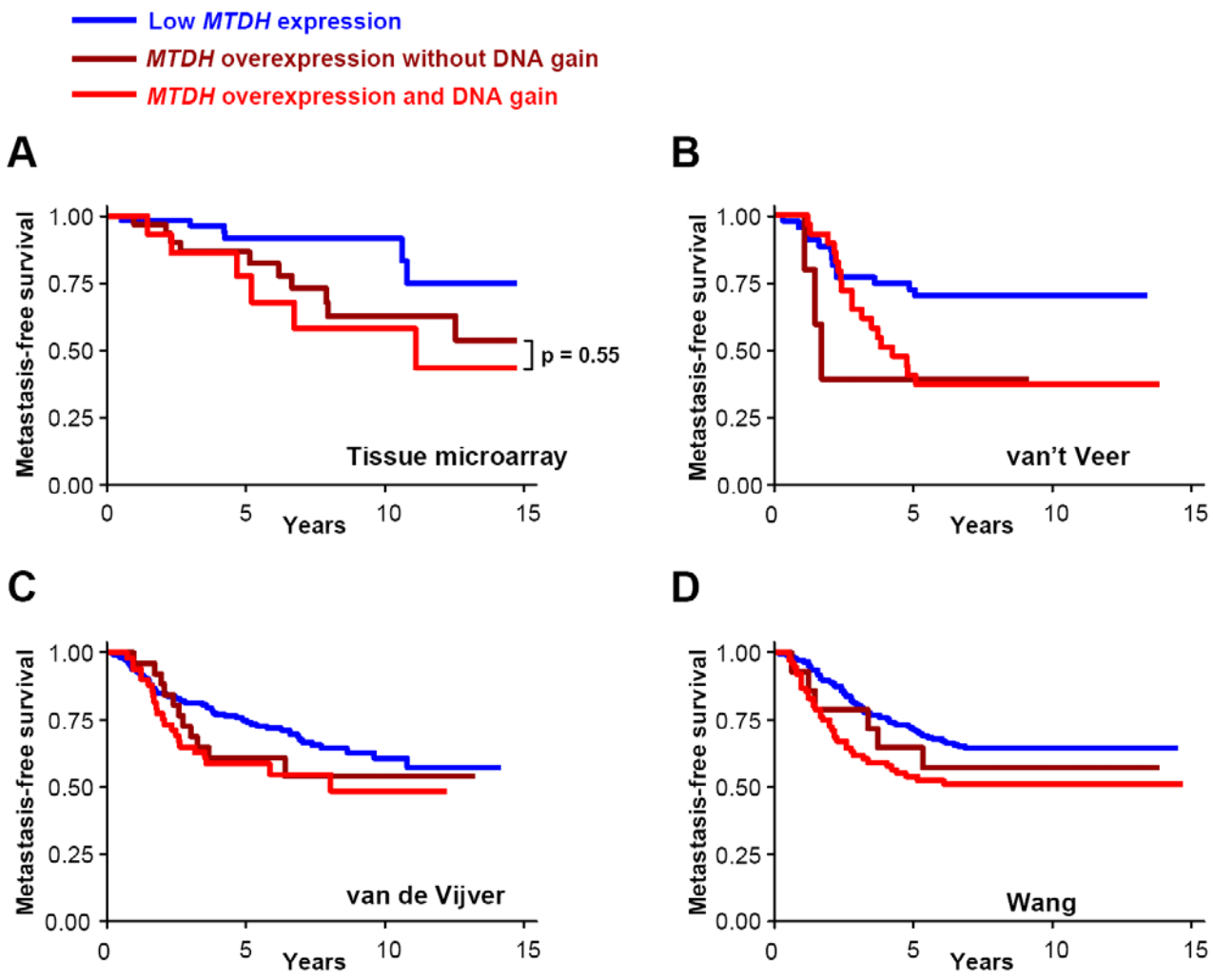

Figure S16. Similar Prognosis of Breast Tumors with *MTDH* Overexpression as the Result of 8q22 Gain or Other Mechanisms of Regulation

Metastasis-free survival analysis of tumor samples with *MTDH* overexpression and genomic gain, with *MTDH* overexpression only, or without *MTDH* overexpression. Analysis was performed on tumors in a tissue microarray (A), and in the three published data sets (C-D). No statistically significant difference was seen for samples with *MTDH* activated by genomic gain or by other mechanisms.

Supplemental Software

The ACE (Analysis of CNAs by Expression) software package and instruction manual is available for download from the *Cancer Cell* website.