Supplemental Methods

Animals and recurrence assays

All mice were housed and treated in accordance with protocols approved by the Institutional Animal Care and Use Committee at the University of Pennsylvania. MMTV-rtTA;TetO-HER2/neu, MMTV-rtTA;TetO-MYC and MMTV-rtTA;TetO-Wnt1;p53+/− mice were bred, housed, induced with 2 mg/ml doxycycline, monitored for tumor development and recurrence, and sacrificed as described (1-3).

Orthotopic recurrence assays were performed as described (4). Athymic nude mice (nu/nu) were purchased from Taconic (Germantown, NY). For fluorescent cell competition assays, EGFP and ECFP-labeled cells were admixed at a 1:1 ratio and injected orthotopically into the mammary glands of nu/nu mice. A fraction of the cell mixture was seeded in chamber slides to confirm equal input. Once primary tumors reached 3x3 in diameter, doxycycline was withdrawn from drinking water to down-regulate the expression of HER2/neu. Primary tumors and residual lesions were harvested at increasing time points following deinduction, as indicated in the text, and preserved in O.C.T. (Tissue-Tek). 4μm O.C.T. sections were fixed with 4% PFA and images were captured using a fluorescent microscope. GFP and CFP-positive cells were counted using Image-pro Express (Media Cybernetics).

For fluorescent cell competition assays evaluating the c-MET inhibitor, compound 22 (5), pK1 and Spsb1 expressing tumor cells were labeled with H2B-mCherry and H2B-EGFP, respectively. Equal numbers of pK1 (red) and SPSB1 (green) cells were admixed and injected into the mammary fat pads of nu/nu mice maintained on doxycycline. When primary tumors reached 3mm in diameter, mice were either sacrificed and the primary tumor grafts harvested, or doxycycline was withdrawn and mice were either treated with the c-MET inhibitor, compound
22 (Amgen), or vehicle control. Regressing tumor grafts were harvested four days after doxycycline withdrawal. Harvested tissues were preserved, analyzed and quantified as above.

**Cell lines**

BT474, MDA-MB-231 and Hs578T cells were purchased from ATCC and cultured at 37.0°C in 5% CO2. BT474 was cultured in RPMI-1640 medium with 10% fetal bovine serum (FBS, Gibco). MDA-MB-231 and Hs578T were cultured in DMEM medium with 10% FBS.

Primary tumor cells harvested from tumors arising in MMTV-rtTA; TetO-HER2/neu mice maintained on doxycycline were cultured as described (4) in the presence of doxycycline and transduced with retrovirus expressing either Spsb1 or shRNAs targeting a gene of interest. Following infection, puromycin was used to select stably-transduced cells.

**Cell growth, apoptosis and clonogenic assays**

For cell growth analysis, 2x10^5 primary tumor cells were seeded in 6-well plates. The following day, one set of cells was maintained in doxycycline-containing media while doxycycline was withdrawn from the other set of cells for 4 days, following which viable cells were counted using a Vi-Cell counter (Beckman-Coulter).

For cell apoptosis analysis, 10,000 cells were seeded in 96-well plates. The following day, cells were treated with varying concentrations of Lapatinib (B-Bridge, SY-Lapatinib) for 24 h. Apoptosis was measured by normalizing capspase-3/7 activity to cell number, which were quantified using Caspase-Glo 3/7 and Cell-titre-Glo assays (Promega), respectively.

For clonogenic assays using mouse tumor cells, 2x10^3 primary tumor cells were seeded in 60mm plates. The following day, plates were washed and re-fed with doxycycline-free media
without HGF (for pK1 and SPSB1 cells), or supplemented with 10ng/ml HGF (for shSPSB1-expressing and control cells) for 20 days, following which cells were fixed with 20% methanol containing 0.5% crystal violet. For experiments using the c-MET inhibitor, compound 22, 100nM inhibitor or DMSO-containing doxycycline-free media were fed to the cells three times per week for 14 days.

**Antibodies**

The following primary antibodies were used: rat anti-Cytokeratin-8 (TROMA-1, Developmental Studies Hybridoma Bank, University of Iowa), rat anti-E-cadherin (clone ECCD2, 13-1900, Zymed), rat anti-Ki67 (M7249, DAKO), rabbit anti-cleaved caspase-3 (9661, Cell signaling), rabbit anti-cleaved parp (9544, Cell signaling), rabbit anti-phospho-MET (3077, Cell signaling), mouse anti-c-MET (3127, Cell signaling), rabbit anti-phospho-AKT (9271, Cell signaling), rabbit anti-AKT (9272, Cell signaling), rabbit anti-phospho-STAT3 (9145, Cell signaling), mouse anti-phospho-STAT3 (9139, Cell signaling), rat anti-HA (11867423001, Roche), mouse anti-beta-tubulin (MU122-UC, BioGenex) and normal mouse IgG (12-371, Millipore). Rabbit anti-Spsb1 antibody was raised against a peptide (DMRDPTYRPLKQEL) from mouse Spsb1 and purified by affinity purification (Openbiosystems).

The following secondary antibodies were used: goat anti-rat Alexa 488 (Molecular Probe, A11006), goat anti-rabbit Alexa 594 (Molecular Probe, A11012), IRDye 800cw conjugated Goat anti-rabbit IgG (Licor, 926-32211) and IRDye 680LT Conjugated Goat (polyclonal) anti-mouse IgG (Licor, 926-68020).

**Western blot analysis**
Proteins for immunoprecipitation were extracted in IP buffer (50mM Tris 7.4, 1mM EDTA, 150mM NaCl and 1% Triton) supplemented with Halt Protease Inhibitor Cocktail (Thermo Scientific). Proteins for western blot analysis were extracted from tissues or cultured cells in RIPA buffer (50mM Tris [pH 8.0], 150mM NaCl, 0.5% sodium deoxycholate, 1% NP-40, 0.1% SDS) supplemented with Halt Protease Inhibitor Cocktail.

For immunoprecipitation analyses, cell lysates were precleared with protein G and mouse IgG prior to incubation with anti-MET antibody at 4°C for 4 h, followed by 1 h with protein G-conjugated agarose beads (Invitrogen, 15929-010). Protein-bound beads were washed five times with buffer, eluted by boiling in Laemli buffer and subjected to western blot analysis.

Immunoprecipitation reactions as well as proteins extracted from tumor tissues and cultured cells were separated by denaturing SDS polyacrylamide gel electrophoresis and transferred onto nitrocellulose membranes (Bio-rad), which were then incubated with the indicated primary and secondary antibodies. The Odyssey V3.0 system (Li-COR Biosciences) was used to visualize and quantify proteins of interest. For quantification of protein phosphorylation, the relative level of each phospho-protein was calculated by normalizing its expression to that of the corresponding total protein.

**Immunofluorescence studies**

For immunofluorescence studies on tumor grafts and residual lesions, samples were snap-frozen in O.C.T. and 8μm frozen sections fixed in 4% paraformaldehyde were stained with appropriate primary and secondary antibodies. For cell culture experiments, cells were fixed on 4-well chamber glass slides with 4% paraformaldehyde and processed in a manner analogous to
that of tumor samples. Microscopy was performed on a Leitz DMRXE microscope (Leica) and images were captured using a Spot RT Color Camera (Diagnostic Instruments).

**Quantitative RT-PCR**

For Quantitative RT-PCR, total RNA was isolated from tumor tissues using Trizol (Invitrogen) according to the manufacturer’s instructions. cDNA was synthesized using the cDNA archive kit (Applied Biosystems). Real-time PCR reactions were performed in triplicate for each sample using Taqman primers and probes (Applied Biosystems). Reactions were run on an ABI PRISM 7900HT Sequence Detection System (Applied Biosystems). The TATA-binding protein, \( \text{TBP} \), was used as a reference for all reactions. Relative expression levels were calculated using the comparative Ct method (6).

RNA was reverse transcribed with the cDNA archive kit (4368814, Applied Biosystems). TaqMan probes were purchased from Applied Biosystems: mouse \( \text{Tbp} \) (Mm00446973_m1), mouse \( \text{Spsb1} \) (Mm00499665_mH), human \( \text{TBP} \) (4333769T), and human \( \text{SPSB1} \) (Hs00228304_m). Reactions were run on an ABI PRISM 7900HT Sequence Detection System (Applied Biosystems).

**Retroviral cDNA and Hairpin RNA Constructs and viral packaging**

PCR primers were designed to fuse an HA tag at the N-terminus of \( \text{SPSB1} \) (forward primer, ATCTCGAGATGTACCCATACGATGTTCCAGATTACGCTGGCGGCGGCGGTCAG AAGG TCACAGG; reverse primer, ATATGAATTCCGGTCAGAAGGTCACAGGAGG), using a cDNA encoding mouse Spsb1 (Open Biosystems, MMM1013-98479055). The corresponding PCR product was cloned into the EcoRI and XhoI sites of pK1 (gift from Dr.
Warren Pear, University of Pennsylvania). Sequencing analysis was performed to confirm the fidelity of PCR amplification and cloning. Short hairpin RNAs (shRNA) targeting mouse Spsb1 were designed using software from RNAi central (http://katahdin.cshl.org/siRNA/RNAi.cgi?type=shRNA). shSpsb1a:

```
TGCTGGACAGTGACGGCTGGACCTGCTCAGACTAGTGAAGCC
```

shSpsb1b:

```
TGCTGGACAGTGACGGCTCAGATGTCACATACCATTAGTGAAGCC
```

Hairpin constructs were cloned in LMP vector (EAV4679, Openbiosystems). shMET specific to mouse c-Met was purchased from Open Biosystems (V2LMM_32199). Hairpins against human SPSB1 were purchased from Open Biosystems (shSPSB1a, V2LHS_645482; shSPSB1b, V2LHS_639403).

**Statistical analyses**

The two-sample Student t-test was used to compare unpaired experimental data between two groups that met the normality assumption. The one-sample t-test was used to test whether log-ratios of paired experimental data were significantly different from 0. Two-way ANOVA with Bonferroni pair-wise post-hoc tests was used to compare data with two experimental factors and that met the normality and the homoscedasticity assumptions. The normality assumption was tested using the Shapiro-Wilk test (7). The homoscedasticity assumption was tested using the Bartlett test (8). When these assumptions were not met, data were log-transformed and re-evaluated with respect to these assumptions. For data sets in which log transformation was inadequate to meet the analysis assumptions, the non-parametric Mann-Whitney test was used instead of the t-test. All data sets analyzed by ANOVA met the assumptions either before or
after log-transformation. Assumption tests were performed in R 2.15.0 using the Shapiro test and the Bartlett test functions. All other tests were performed in Prism version 5 for Windows (GraphPad Software, La Jolla CA USA, www.graphpad.com).

**Human breast cancer microarray data analysis**

Publicly available microarray data for 2,324 patients contained within 14 human primary breast cancer data sets (9-24) along with the corresponding clinical annotations were downloaded from NCBI GEO or original authors’ websites. Microarray data were converted to base 2 log scale where necessary. Affymetrix microarray data were re-normalized using Robust Multi-array Average (RMA (25)) when .CEL files were available.

Within each data set, the effect size of the association between *SPSB1* mRNA expression and 10-year relapse-free survival was estimated using two different methods: 1) hazard ratio from Cox proportional hazards regression in which *SPSB1* expression was modeled as a continuous variable; and 2) the concordance index (c-index (26)). Each type of effect size estimate was combined across data sets by meta-analysis using the inverse-variance weighting method (27). Between-study homogeneity of survival association was tested using chi-squared test on Cochran’s Q statistic (28), for which a p-value of less than 0.1 was interpreted as evidence of significant heterogeneity. In the presence of significant heterogeneity, the random-effect model (29) was used for meta-analysis. In the absence of significant heterogeneity, the fixed-effect model (30) was used.

Cox proportional hazards regression, concordance index analysis, and meta-analysis were performed using the “coxph” function in the “survival” package, the “survConcordance” function in the “survival” package and the “metagen” function in the “meta” packages in R
The precision of the meta-analysis p-values was verified using Monte Carlo permutation tests by randomly permuting sample labels and repeating the above meta-analyses 10,000 times. For data sets in which relapse-free survival information was not available, but distant metastasis-free survival information was available, metastasis-free survival was used for survival analysis.

Additional meta-analyses were performed in subsets of samples stratified by ER status, HER2 status or intrinsic molecular subtype, as well as in the subset of patients who, according to available treatment information, did not receive any adjuvant systemic treatment. Since HER2 immunohistochemical status was not available for several data sets, HER2 status was approximated by ERBB2 mRNA expression as measured by microarray in a similar fashion as the Cancer Outlier Profile Analysis (31). In each data set, HER2-positive and HER2-negative samples were defined as being above and below a cutoff of 1.5 median absolute deviations above the median, which resulted in average specificity of 98% and sensitivity of 78% in five validation data sets (11, 13, 32, 33). Due to the non-random association between ER and HER2 status, approximation of HER2 status was not attempted in data sets consisting entirely of hormone-positive or hormone-negative cancers. Assignments of intrinsic subtype were done using the PAM50 classifier (34) after expression data were median-centered for each gene. For each significant overall p-value derived by meta-analysis, influence analysis was performed to determine whether the significant result was independent of any single data set.

The association between SPSB1 mRNA expression and categorical prognostic variables in human breast cancers, including ER status, PR status, HER2 status, lymph node status, tumor size, tumor grade and intrinsic molecular subtype, was assessed by ANOVA in pooled microarray data sets. For each categorical prognostic variable, SPSB1 expression was normalized against the mean expression of the same baseline group in each data set, and pooled
across all data sets for which the prognostic variable was available. Baseline groups used for ER status, PR status, HER2 status, lymph node status, tumor grade, tumor size and intrinsic molecular subtype were ER-negative, PR-negative, HER2-negative, lymph node-negative, low tumor grade (Grade I and II), T1 tumor size, and Luminal A subtype, respectively. Baseline normalization was performed by subtracting mean \textit{SPSB1} expression (log2 scale) in the baseline group from \textit{SPSB1} expression in each sample.

To estimate relative c-MET activity in human breast cancer samples, we generated a MET signature containing 24 genes concordantly regulated in four gene expression microarray data sets: 1) a \textit{c-Met}-amplified mouse mammary tumor cell line treated with vehicle or 300nM compound 22 (Met inhibitor) for 24 h; 2) the \textit{c-MET}-amplified human gastric carcinoma cell line, GTL-16, treated with the MET inhibitor PHA or with \textit{c-MET} knockdown by inducible siRNA (GSE19043); 3) primary cultures of \textit{c-Met} wild type or knockout mouse hepatocytes treated with vehicle or 50ng/ml HGF (GSE4451); and 4) liver tissues of \textit{c-Met} wild type mice treated with vehicle or HGF (GSE13992). Signatures genes were selected from differentially expressed genes common to the four experiments. Following validation of the signature’s ability to estimate MET activity, a previously described scoring method (35) was used to estimate c-MET activity in human breast cancer samples.

Correlation between \textit{SPSB1} expression and estimated relative c-MET activity was assessed in human breast cancer data sets using Pearson correlation coefficient and summarized across data sets by meta-analysis. The correlation was assessed using all breast cancer samples as well as within each of the four sample partitions determined by the median expressions of \textit{c-MET} and \textit{HGF} within each data set. Only data sets for which microarray data were available for \textit{SPSB1}, \textit{c-MET}, and \textit{HGF} were used in the correlation analysis.