Cell

HSF1 Drives a Transcriptional Program Distinct from Heat Shock to Support Highly Malignant Human Cancers

Marc L. Mendillo,[1,8] Sandro Santagata,[1,2,8] Martina Koeva,[1,3] George W. Bell,[1] Rong Hu,[5,6] Rulla M. Tamimi,[5,6] Ernest Fraenkel,[3] Tan A. Ince,[7] Luke Whitesell,[1,*] and Susan Lindquist[1,4,*]

[1]The Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA
[2]Department of Pathology, Brigham and Women's Hospital, Boston, MA 02115, USA
[3]Department of Biological Engineering
[4]Howard Hughes Medical Institute, Department of Biology
Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[5]Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA
[6]Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA
[7]Department of Pathology, Braman Family Breast Cancer Institute and Interdisciplinary Stem Cell Institute, University of Miami Miller School of Medicine, Miami, FL 33136, USA
[8]These authors contributed equally to this work
*Correspondence: whitesell@wi.mit.edu (L.W.), lindquist_admin@wi.mit.edu (S.L.)
http://dx.doi.org/10.1016/j.cell.2012.06.031

SUMMARY

Heat-Shock Factor 1 (HSF1), master regulator of the heat-shock response, facilitates malignant transformation, cancer cell survival, and proliferation in model systems. The common assumption is that these effects are mediated through regulation of heat-shock protein (HSP) expression. However, the transcriptional network that HSF1 coordinates directly in malignancy and its relationship to the heat-shock response have never been defined. By comparing cells with high and low malignant potential alongside their nontransformed counterparts, we identify an HSF1-regulated transcriptional program specific to highly malignant cells and distinct from heat shock. Cancer-specific genes in this program support oncogenic processes: cell-cycle regulation, signaling, metabolism, adhesion and translation. HSP genes are integral to this program, however, many are uniquely regulated in malignancy. This HSF1 cancer program is active in breast, colon and lung tumors isolated directly from human patients and is strongly associated with metastasis and death. Thus, HSF1 rewires the transcriptome in tumorigenesis, with prognostic and therapeutic implications.

INTRODUCTION

A wide variety of environmental stressors can damage proteins. These include elevated temperatures, oxidative agents, heavy metals, and low pH. Organisms respond by inducing heat-shock proteins (HSPs), which act as molecular chaperones to restore protein homeostasis (Shamovsky and Nudler, 2008; Whitesell and Lindquist, 2005). This powerful adaptive mechanism, known as the heat-shock response, is unleashed by the heat-shock transcription factor HSF1. Upon heat shock, HSF1 is phosphorylated, trimerizes, and translocates to the nucleus. There, it induces chaperone gene expression by binding to DNA sequence motifs known as heat-shock elements (HSEs) (Pelham, 1982; Sakurai and Enoki, 2010). Major aspects of this classic response are conserved from yeast to humans and are vital in many stressful environments. HSF1 also functions as a critical regulator of longevity in some organisms (Chiang et al., 2012; Volovik et al., 2012). Consistent with this, recent work indicates that HSF1 helps cells accommodate the complex pathophysiological derangements in protein homeostasis that underlie many human diseases, especially those associated with aging (Morimoto, 2008).

We have previously shown in mice that HSF1 is co-opted by tumor cells to promote their survival, to the detriment of their hosts. The importance of HSF1 in supporting carcinogenesis, at least in model systems, is demonstrated by the dramatically reduced susceptibility of *Hsf1*-knockout mice to tumor formation. This has been established for cancers driven by oncogenic RAS, tumor suppressor *p53* mutations, and chemical carcinogens (Dai et al., 2007; Jin et al., 2011; Min et al., 2007). In addition to its role in tumor formation in mice, HSF1 fosters the growth of human tumor cells in culture. Depleting HSF1 from established human cancer lines markedly reduces their proliferation and survival (Dai et al., 2007; Meng et al., 2010; Min et al., 2007; Santagata et al., 2012; Zhao et al., 2011).

In mouse models, HSF1 enables adaptive changes in a diverse array of cellular processes, including signal transduction, glucose metabolism, and protein translation (Dai et al., 2007; Khaleque et al., 2008; Lee et al., 2008; Zhao et al., 2009, 2011). The commonly held view is that HSF1 exerts this broad influence in cancer simply by allowing cells to manage the

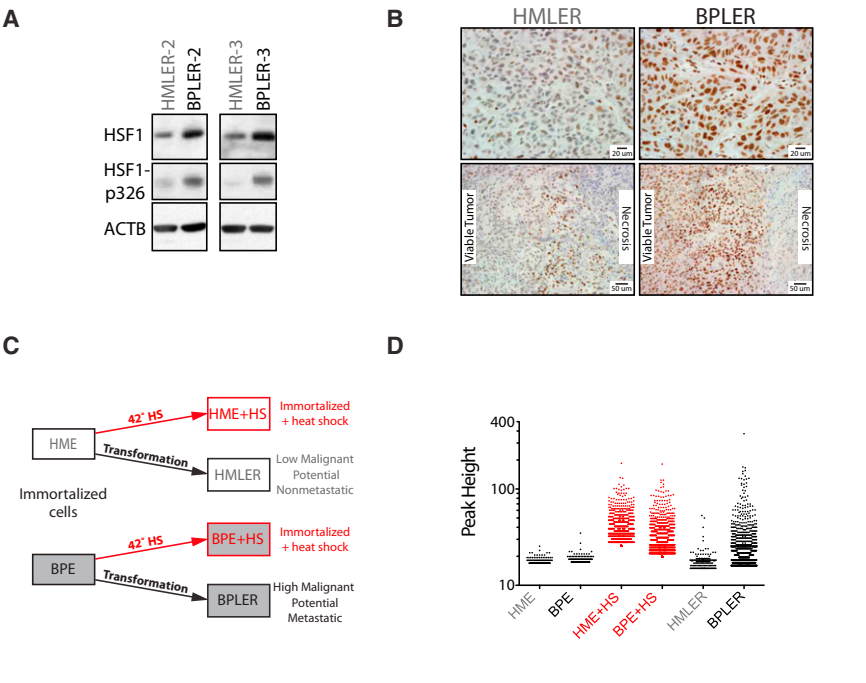

Figure 1. HSF1 Is Activated in Metastatic and Highly Tumorigenic Human Mammary Epithelial Cell Lines

(A) Equal amounts of total protein were immunoblotted with the indicated antibody.

(B) IHC staining with anti-HSF1 antibody of the indicated tumors xenografted in mice. Upper panels show regions of viable tumor (high magnification; scale bar, 20 μm) and lower panels show the viable tumor/necrotic interface (low magnification; scale bar, 50 μm).

(C) Schematic of experimental groups analyzed by HSF1 ChIP-Seq.

(D) Graph of ChIP-Seq peak heights for each region of HSF1 occupancy, normalized by the number of reads in the data set.

(E) Overlap of genes bound in malignant cells (BPLER, 37°C) and immortalized, nontumorigenic cells after heat shock (BPE or HME cells, 42°C).

(F) Representative genes bound in BPLER cells (*CKS2, LY6K, RBM23*) and bound in both BPLER cells and heat-shocked HME and BPE cells (*HSPA6, HSPA8*, *PROM2*). x axis: from −2kb from the transcription start site (TSS) to either +5, +6 or +10kb from the TSS for each gene; genes diagrams drawn to scale. See also Figure S1 and Tables S1 and S2.

imbalances in protein homeostasis that arise in malignancy. According to this view, the main impact of HSF1 on tumor biology occurs indirectly, through the actions of molecular chaperones like HSP90 and HSP70 on their client proteins (Jin et al., 2011; Solimini et al., 2007). An alternate, and to date unexplored, possibility is that HSF1 plays a more direct role, rewiring the transcriptome and, thereby, the physiology of cancer cells.

To investigate the HSF1-regulated transcriptional program in cancer and how it relates to the classical heat-shock response, we first took advantage of human breast cancer cell lines with very different abilities to form tumors and metastasize (Ince et al., 2007). Two types of primary mammary epithelial cells (HMEC and BPEC) have been isolated from normal breast tissue derived from the same donor during reductive mammoplasty (Ince et al., 2007). These pairs of isogenic cells were established by using different culture conditions that are believed to have supported the outgrowth of distinct cell types. The cells were immortalized (HME and BPE) and then transformed with an identical set of oncogenes (HMLER and BPLER). The resulting tumorigenic breast cell lines had very different malignant and metastatic potentials (low, HMLER, and high, BPLER) supporting the concept that the cell type from which a cancer arises ("cell-of-origin") can significantly influence its ultimate phenotype (Ince et al., 2007).

Here, using this well-controlled system, we identify changes in the HSF1 transcriptional program that occur during transformation and underlie the different malignant potentials of these cells. Chromatin immunoprecipitation coupled with massively parallel DNA sequencing (ChIP-Seq) revealed a surprisingly diverse transcriptional network coordinated by HSF1 in the highly malignant cells. We then extend analysis of this HSF1 cancer program to a wide range of well-established human cancer cell lines and to diverse types of tumors taken directly from patients. Finally, we establish the clinical relevance of our findings through in-depth analysis of HSF1 activation in cohorts of breast, colon, and lung cancer patients with known clinical outcomes. Thus, the breadth of HSF1 biology is far greater than previously appreciated.

RESULTS

HSF1 Is Activated in Highly Tumorigenic Cells

We first asked whether HSF1 expression differed in the highly malignant BPLER and the much less malignant HMLER breast cancer cells (Ince et al., 2007). We used two sets of such cells, each pair derived independently from a different donor. In both, HSF1 protein expression was higher in the more malignant member of the pair, the BPLER cells (Figure 1A). The BPLER cells

also had more phosphoserine-326-HSF1, a well established marker of HSF1 activation (Guettouche et al., 2005), than the HMLER cells (Figure 1A).

To determine whether these differences in HSF1 were simply an artifact of growth in cell culture, we implanted the cells into immunocompromised mice and allowed them to form tumors. HSF1 immunostaining was weak in the HMLER tumors. Moreover, it was largely restricted to nonmalignant, infiltrating stroma and to tumor areas bordering necrosis (Figure 1B), indicating that microenvironmental stress can influence the activation of HSF1. In BPLER tumors, however, HSF1 staining was strong, nuclear localized and very uniform thoughout (Figure 1B; Figure S1A available online). Thus, the dramatic difference in HSF1 expression we observe between BPLER and HMLER cells is due to stable, cell-autonomous factors intrinsic to these distinct cell types (Ince et al., 2007).

Given this evidence for the activation of HSF1 in the BPLER cell type, we asked whether such cells were more dependent on HSF1 than HMLER cells for growth and survival. Neither cell type was affected by negative control shRNA. With two independent shRNAs that knockdown HSF1 expression, however, cell growth and viability were far more strongly reduced in the BPLER than the HMLER cells (Figure S1B).

HSF1 Genome Occupancy in Cancer Is Distinct from Heat Shock

To determine whether the transcriptional program driven by HSF1 in highly malignant cells differs from that driven by a classical thermal stress, we used chromatin immunoprecipitation coupled with massively parallel DNA sequencing (ChIP-Seq), characterizing HSF1-binding sites genome wide. We first assessed the immortalized nontransformed progenitor cells, HME and BPE, grown at 37°C or following a 42°C heat shock (Figure 1C). We then related the genome-wide distribution of HSF1-binding sites to those of the oncogenically transformed HMLER and BPLER cells grown at 37°C.

In the HME and BPE parental cell lines, a limited number of genes were bound by HSF1 in the absence of heat shock, and these were bound weakly (Figure 1D; Table S1). Heat shock drove robust binding of HSF1 to ~800 genes in HME cells and to ~1,100 genes in BPE cells (Figure 1D; Table S1). These observations are consistent with a previous report that a large number of genes are bound by HSF1 in the mammalian heat-shock response (Page et al., 2006).

A small number of genes were bound by HSF1 under basal conditions in the transformed cells with low malignant potential, HMLER (37°C; Figure 1D). However, binding was more localized to promoter regions than in the parental cells (Figure S1C), suggesting some low level of HSF1 activation (MacIsaac et al., 2010). In sharp contrast, in the metastatic and highly tumorigenic BPLER cells, we identified ~900 genes bound by HSF1 at 37°C (Figure 1D; Table S1).

Surprisingly, a full 60% of the genes bound by HSF1 in BPLER cells were not bound in nontransformed parental lines, even after heat shock (Figure 1E). Examples included (Figure 1F) cyclin-dependent kinase interacting protein, *CKS2*, which enables proliferation under conditions of replicative stress common to malignant cells (Liberal et al., 2011); *LY6K*, which encodes a glycosylphosphatidyl-inositol (GPI)-anchored membrane protein implicated as a biomarker in lung and esophageal carcinomas (Ishikawa et al., 2007; Maruyama et al., 2010); and *RBM23*, which encodes an RNA-binding protein implicated in the regulation of estrogen-mediated transcription (Dowhan et al., 2005). Using the Molecular Signatures Database (MSigDB) (Subramanian et al., 2005), the genes bound by HSF1 in the BPLER cells, but not in either of the parental lines after heat shock, were most highly enriched in protein translation, RNA binding, metabolism, cell adhesion (Figure S1D; Table S2), and other processes vital in supporting the malignant state.

We analyzed the 100 bp genomic regions surrounding the peaks of HSF1 binding unique to BPLER cells using the ab initio motif discovery algorithm MEME (Machanick and Bailey, 2011). The canonical HSE was highly enriched in the HSF1-bound regions (p value = 1.4×10^{-97}; Figure S1E), strongly suggesting the genes that are constitutively bound by HSF1 in malignant cells are bona fide HSF1-binding targets.

The remaining 40% of genes bound by HSF1 in BPLER cells under basal conditions were also bound in at least one of the two parental lines following heat shock. As expected, these genes included many classical heat-shock genes, including *HSPA8*, the constitutively expressed HSC70 protein, and *HSPD1/E1*, which encodes HSP60 and HSP10 (Figures 1F and S1F). By the MSigDB, this large group of genes was enriched for protein folding categories (Figure S1D; Table S2).

Notably, however, for many of the genes bound in both cancer and heat shock, HSF1 binding differed. For example, the strongly heat-shock inducible *HSPA6* gene (encoding HSP70B') was highly bound in parental lines upon heat shock but only weakly bound in BPLER cells at 37°C (Figures 1F, S1G and S1H). Conversely, *PROM2,* which encodes a basal epithelial cell membrane glycoprotein, was weakly bound by HSF1 in parental lines following heat shock, but highly bound in BPLER cells (Figure 1F). Thus, HSF1 engages a regulatory program in the highly malignant state that is distinct from the classic heat-shock response.

To assess the functional significance of the HSF1 cancer program, we asked whether the genes comprising this program played a significant role in malignancy and used unbiased data from an independent investigation. The Elledge lab recently conducted a whole-genome siRNA screen to identify genes that are required to maintain growth when cells are transformed with a malignantly activated Ras gene (Luo et al., 2009). Among the ~1,600 genes identified in this screen, our HSF1-bound gene set was very strongly enriched (73 gene overlap; p value = 7.95×10^{-15}, Table S2). The HSF1-bound genes we identified as unique to the malignant state were more strongly enriched (49 gene overlap; p value = 1.1×10^{-12}) than those shared with heat-shocked cells (24 gene overlap; p value = .0004), but both sets of genes were important in supporting the malignant state.

HSF1 Regulates the Transcription of Genes It Binds in Malignant Cells

To investigate the consequences of HSF1 occupancy on gene expression, we compared RNA profiles in HMLER and BPLER cells transduced with control shRNA hairpins to those transduced with hairpins that knockdown HSF1. As we previously

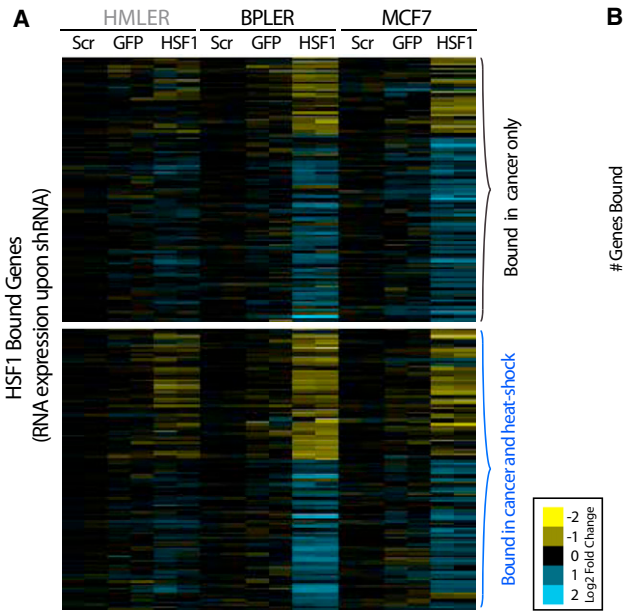

Figure 2. The Expression of HSF1-Bound Genes Is Altered by HSF1 Depletion

(A) Relative gene expression levels following shRNA-mediated knockdown of HSF1 in indicated cells. Scr and GFP were negative control shRNA. (B) Graph showing the number of genes positively regulated (reduced expression upon HSF1 depletion) or negatively regulated (increased expression upon HSF1 depletion) by HSF1 relative to site of gene occupancy by HSF1 (promoter versus distal). See also Figure S2 and Table S3.

reported, the growth and survival of malignant cells is compromised by prolonged depletion of HSF1 (Dai et al., 2007). Therefore, we only analyzed mRNA expression in the early stages of shRNA inhibition, where HSF1 knockdown was still incomplete (Figure S2), but cell viability was unimpaired. This provides a conservative assessment of the effects of HSF1 on gene expression in malignant cells.

Control hairpins that did not reduce HSF1 levels (Scr and GFP; Figure S2) had minimal effects on the expression of HSF1-bound genes (Figure 2A; Table S3). Targeted hairpins that did reduce HSF1 had a minor impact in HMLER cells but markedly changed expression in BPLER cells. The expression of many genes decreased indicating that they were positively regulated by the HSF1 transcription factor. Many genes increased, indicating that a larger number of genes than previously appreciated are negatively regulated by HSF1. Genes unique to the malignant state and genes shared by heat-shocked cells were affected equivalently. For example, expression of the malignancy-associated genes *CKS2* and *RBM23* and the heat-shock protein genes *HSPA8* (HSC70) and *HSP90AA1* (HSP90) were all reduced (by ~50%) following HSF1 knockdown (Table S3).

Relating the effects of HSF1 knockdown on gene expression to our earlier ChIP-Seq analysis, ~70% of genes positively regulated by HSF1 were bound at the promoter, whereas only ~30% of these genes were bound in distal regions (Figure 2B). Genes that were negatively regulated by HSF1 showed the opposite pattern (Figure 2B). This observation (p value = 0.00004) suggests that the direction of regulation (positive versus negative) in these cells is influenced by the location of the HSF1-binding site.

Next we examined the effects of HSF1 knockdown on gene expression in a cell line that had not been deliberately engineered. The MCF7 line was established from a human breast cancer metastasis. Moreover, as an estrogen receptor positive (ER+) line, its biology is fundamentally distinct from the hormone-receptor-negative HMLER and BPLER cell lines. Despite these differences, the changes in gene expression caused by HSF1 knockdown was very similar in BPLER cells and MCF7 cells (Figure 2A).

HSF1 Gene Occupancy Is Conserved across a Broad Range of Common Human Cancer Cell Lines

Next we used ChIP-qPCR to monitor HSF1 binding to a representative set of the HSF1-target genes in cell lines derived from patients with breast cancer. We used nine well-studied cancer lines (including MCF7 cells) representing all three major categories of breast cancer: ER^+, $HER2^+$and Triple-Negative (TN). Under basal conditions (at 37°C), we detected HSF1 binding in each of the major breast cancer subtypes (Figure S3A). A range of binding intensities was observed. Most notably, however, the distinct pattern of HSF1 gene occupancy we had identified in the highly malignant BPLER cells was also present in these naturally-arising malignant cells. This included genes unique to malignant cells, such as *CKS2* and *RBM23,* and genes shared by heat-shocked cells, such as *HSPD1/E1*. Again, the gene most strongly inducible by heat shock, *HSPA6*, was minimally bound across this entire panel of cancer lines under basal conditions (37°C; Figures S3A, S3B and S3C). We also analyzed HSF1 binding in the nontumorigenic breast cell line MCF10A. Comparable to the low malignancy HMLER cells, MCF10A cells had low levels of HSF1 occupancy across all genes examined (Figures S3A and S3C).

These ChIP-PCR data spurred us to employ ChIP-Seq to generate additional genome-wide high-resolution maps of HSF1 occupancy. We performed ChIP-Seq analysis on the non-tumorigenic MCF10A cell line grown either at 37°C or following a 42°C heat-shock. We compared these data with our prior data from the nontumorigenic cell lines HME and BPE and weakly tumorigenic HMLER cells. We then assessed HSF1 binding in a panel of human tumor lines that extended to other types of malignancy: duplicate samples of four breast, three lung and three colon cancer cell lines (Figures 3A and S3D), thus covering the human cancers with the highest total mortality in the developed world.

After heat shock, MCF10A cells exhibited an HSF1-binding profile that was comparable to that of heat-shocked HME and BPE cells. In the absence of heat shock, the overall magnitude

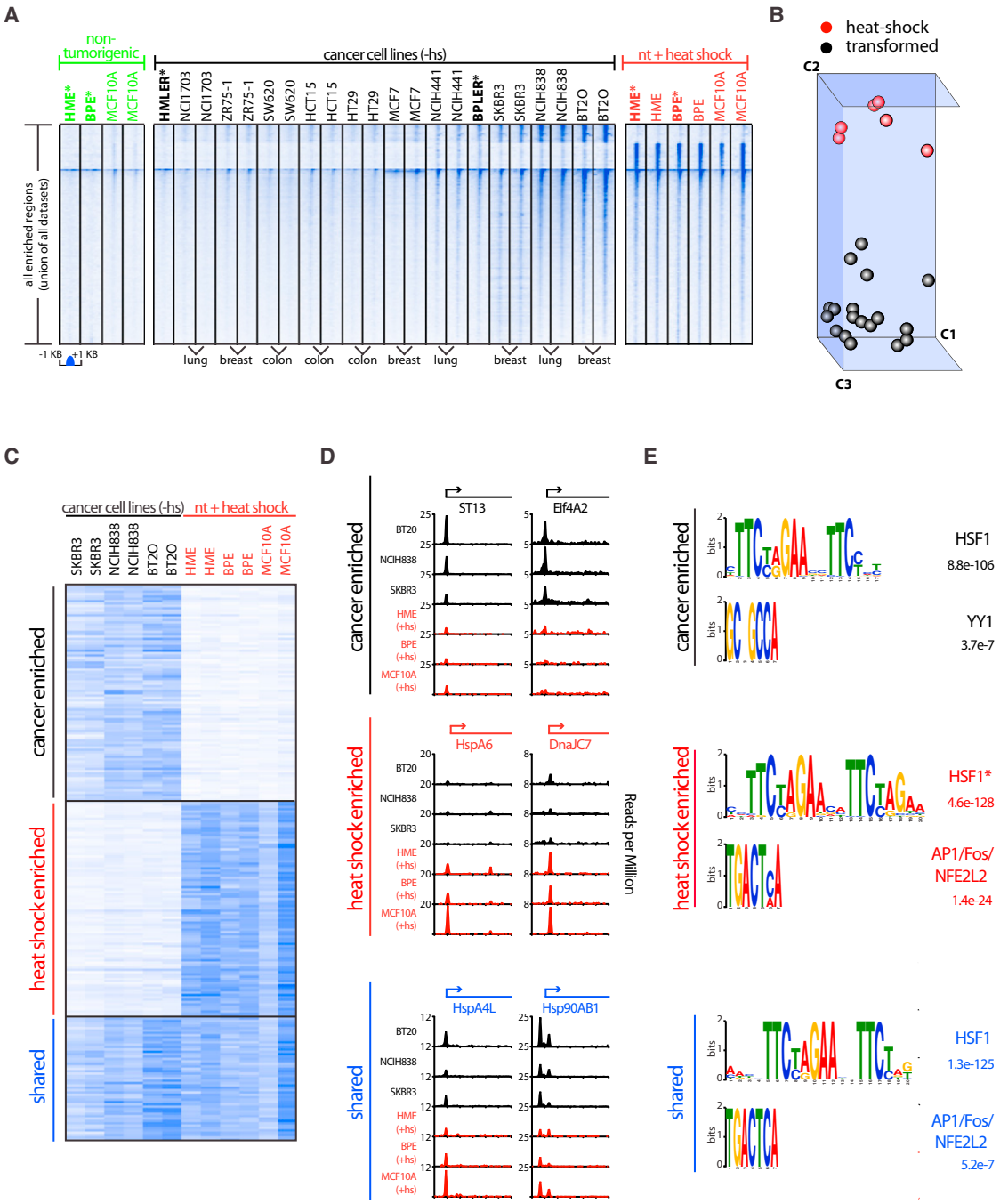

Figure 3. Genome-wide patterns of DNA occupancy by HSF1 across a broad range of common human cancer cell lines

(A) Heat map of ChIP-Seq read density for all HSF1 target regions (union of all HSF1-bound regions in all data sets). Genomic regions from −1kb to +1kb relative to the peak of HSF1 binding are shown. Regions are ordered the same in all data sets. Read density is depicted for nontumorigenic cells at 37°C (green), cancer lines at 37°C (black) and nontumorigenic (nt) lines following heat shock at 42°C (red). Asterisks indicate data sets also used for the analysis in Figure 1E.

(B) Principal component analysis (PCA) of HSF1 binding in heat-shocked parental lines (red) and cancer lines (black).

(C) ChIP-Seq density heat map of genomic regions differentially bound by HSF1 in cancer lines at 37°C, heat-shocked nontumorigenic lines, and regions shared under both conditions.

(D) HSF1 binding of representative genes in cancer lines at 37°C (black) and heat-shocked nontumorigenic lines (red). Examples of genes with distinct patterns of binding are presented: Enriched in cancer lines, heat-shocked nontumorigenic lines, or both.

(E) Motif analysis of 100 bp regions surrounding HSF1-binding peaks for genes enriched in cancer lines (BT20, NCIH838 and SKBR3), heat-shocked non-tumorigenic lines (HME, BPE, MCF10A) and both cancer and heat-shocked nontumorigenic lines. See also Figure S3 and Table S1.

of HSF1 binding in all of the nontumorigenic cell lines (nt) was uniformly very weak, and the total number of bound genes was small (Figure 3A; Table S1). In contrast, in the cancer lines a range of HSF1 binding was observed at 37°C (Figure 3A). For example, robust binding was observed in the lung adenocarcinoma line NCI-H838 and in the TN breast carcinoma line BT20. Less pronounced overall binding was seen in others lines such as the weakly malignant HMLER. Binding in BPLER cells was intermediate.

Irrespective of the level of binding, the distribution of HSF1 occupancy on a genome-wide scale was remarkably similar among the cancer cell lines and distinct from the pattern of binding in the heat-shocked cells (Figure 3A). The differences between the heat-shocked and malignant states were further probed using principal component analysis (PCA; Figure 3B). This unsupervised method of clustering sets of data clearly distinguished one cluster containing all cell lines exposed to heat-shock and a second cluster containing all cancer cell lines. This analysis confirmed the global nature of the differences in the HSF1-binding profiles.

Data from these multiple cell lines allowed us to confidently identify regions of HSF1 binding that were strong in cancer cells but not in heat-shocked cells, weak in cancer but strong in heat shock, or similarly strong in both (Figure 3C). Examples of genes that were strongly bound in cancer but not in heat shock included *CKS2*, *LY6K*, *RBM23*, *CCT6A*, *CKS1B*, *ST13*, *EIF4A2* (Figures S3E; Figure 3D). Genes that were weakly bound in cancer lines but strongly bound in heat shock included *HSPA6* and *DNAJC7* (Figure 3D). Genes that were strongly bound in both cell types included *HSPA4L* and *HSP90AB1* (Figure 3D).

We performed motif analysis to evaluate the 100 bp genomic regions surrounding the peaks of HSF1 binding in each of these groups. The HSE, comprised of adjacent inverted repeats of 5′-nGAAn-3′, was the most enriched motif in all three groups (Figure 3E). The regions strongly bound in cancer but not in heat shock were enriched in HSEs that had three such repeats (p value = 8.8×10^{-106}). They were also enriched in binding elements for YY1, the so called ''ying-yang'' transcription factor that is involved in activating and repressing a broad range of genes (p value = 3.7×10^{-7}). The regions strongly bound in heat-shocked cells but not cancer were enriched for expanded HSEs with a fourth 5′-nGAAn-3′ repeat (p value = 4.6×10^{-128}). They also were enriched in an AP1/Fos/NRF2 (NFE2L2) binding site (p value = 1.4×10^{-24}) as previously reported for mammalian heat-shock genes. This variation in binding motifs suggests the involvement of distinct coregulators in establishing differential patterns of HSF1 occupancy. The regions strongly bound by HSF1 in both cancer and in heat shock, had features of both groups. They were enriched for HSEs with three inverted repeats (p value = 1.3×10^{-125}). They were not enriched for the YY1 sites but were enriched for the AP1/Fos and NRF2 binding site (p value = 5.2×10^{-7}).

HSF1-Bound Genes form Distinct, Coordinately Regulated Modules

Integrating our diverse data sets revealed a direct and pervasive role for HSF1 in cancer biology (Figure 4A). Extending far beyond protein folding and stress, HSF1-bound genes were involved in many facets of tumorigenesis, including the cell cycle, apoptosis, energy metabolism, and other processes. To gain a more global view of the relationship between the genes most strongly bound by HSF1 in cancer cell lines, we generated an RNA expression correlation matrix through meta-analysis of pre-existing data sets (Figure 4B). We used the UCLA Gene Expression Tool (UGET) (Day et al., 2009) to query the extent to which the expression of each HSF1-bound gene correlated with every other HSF1-bound gene across all of the ~12,000 human expression profiles that have been generated with Affymetrix HG U133 Plus 2.0 arrays and made available through the Celsius database (Day et al., 2009). Hierarchical clustering of this gene-gene correlation matrix revealed five major transcription modules (Figure 4B).

The largest module was enriched for protein folding, translation, and mitosis. Genes within this dominant module showed the strongest positive correlation with the expression of *HSF1* mRNA itself. Many of these genes had indeed proven to be regulated by HSF1 in our HSF1 shRNA knockdown experiments (Figures 2, 4A and S4). A second, smaller module was positively correlated with the first and strongly enriched for RNA-binding genes. Many of these genes, too, were positively regulated by HSF1 in our knockdown experiments (Figures 2, 4A, and S4). The remaining three modules (center to lower right of the matrix) were enriched for processes involved in immune functions, insulin secretion, and apoptosis. All three of these modules were negatively correlated with the largest module, suggesting negative regulation by HSF1.

Activation of HSF1 in a Broad Range of Cancer Specimens Taken Directly from Patients

Recently, we evaluated HSF1 expression and localization in a cohort of breast cancer patients culled from the Nurses' Health Study (NHS) (Santagata et al., 2011). In that work, HSF1 was cytoplasmic and expressed at low levels in normal breast epithelial cells, but it accumulated in the nucleus of the majority of tumor specimens. Here, we confirm that finding (Figures 5A, 5B and S5), combining samples from two independent breast cancer collections representing all three major clinical subtypes (see Extended Experimental Procedures).

Next, because our ChIP-Seq analysis showed that the HSF1 cancer program is engaged not just in breast cancer cell lines but also in colon and lung cancer cell lines, we examined more than 300 formalin-fixed surgical specimens taken directly from patients. We included not only colon and lung cancer but also a wide variety of other tumor types.

Normal cells adjacent to the tumor demonstrated low HSF1 levels and cytoplasmic localization of the protein. In contrast, high-level expression of HSF1 and nuclear localization was common (Figure 5C) across every cancer type we examined, including carcinomas of the cervix, colon, lung, pancreas, and prostate as well as mesenchymal tumors such as meningioma. HSF staining was negative or weak in some tumors from each cancer type (Figure 5C). However, in those tumors, where expression was high, it was remarkably uniform across the sample, with nearly all tumor cells expressing similar levels of nuclear HSF1.

To determine whether the high-level nuclear localization of HSF1 detected by immunostaining was truly indicative of its

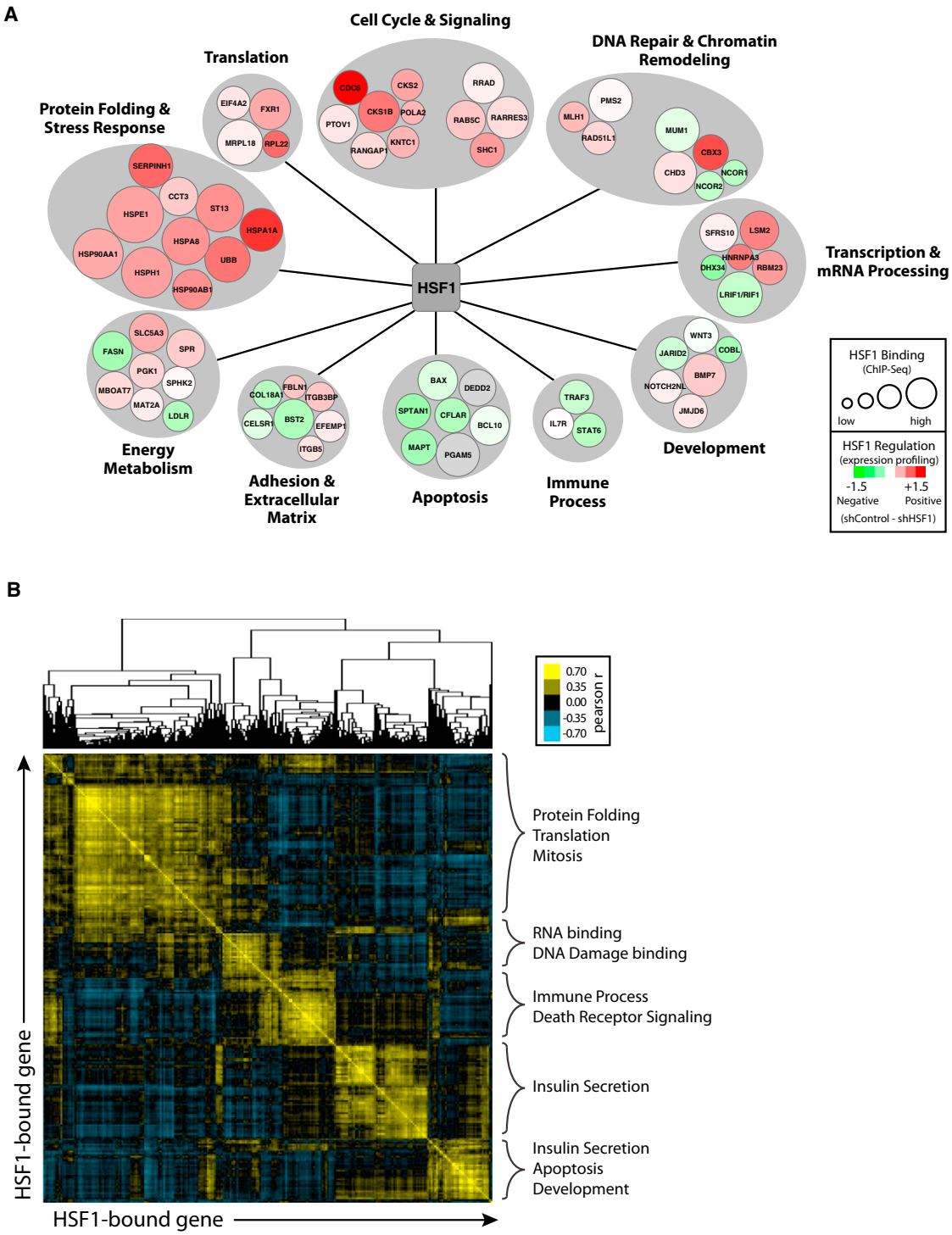

Figure 4. Distinct, Coordinately Regulated Modules of HSF1-Bound Genes

(A) Graphical representation of the HSF1 cancer program integrating information on gene binding, regulation and function. The peak height is reflected in the diameter of the circle (log2 peak height: range ~3 to 9) and color intensity reflects gene regulation (average of log2 fold change in BPLER and MCF7 cells upon HSF1 knockdown; red, positively regulated; green, negatively regulated; gray, no data available). Well-bound, differentially regulated genes as well as several genes of biological interest are displayed.

(B) Gene-gene expression correlation matrix of HSF1-bound genes. Pair-wise correlation map is presented of the genes that were bound by HSF1 in at least two of the three cancer cell lines (BT20, NCIH38, and SKBR3). The Pearson correlation coefficient relating normalized mRNA expression data for each gene pair was assessed in nearly 12,000 expression profiles. Enriched GO (gene-ontology) categories for each module are shown. See also Figure S4.

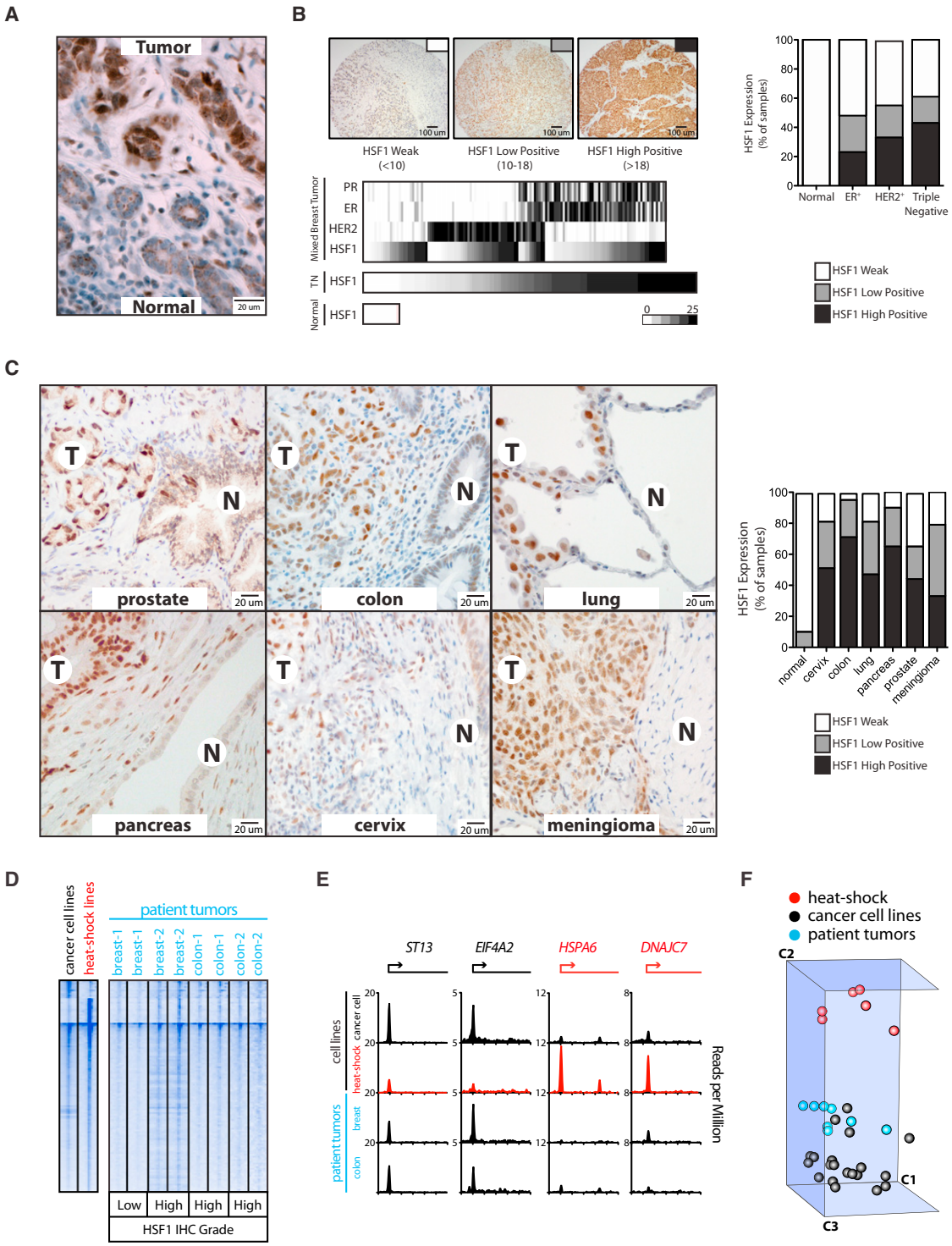

Figure 5. HSF1 Is Activated in a Broad Range of Human Tumors

(A) IHC shows strong nuclear HSF1 staining in human breast tumor cells (top) with adjacent normal breast epithelial cells (bottom) showing a lack of nuclear HSF1.

(B) Images of HSF1 IHC on breast cancer tissue microarray (TMA) cores. Heat map shows scoring of three TMAs. The top panel depicts data from two TMAs (BRC1501 and BRC1502), containing 138 breast tumors of all major breast cancer subtypes. Progesterone receptor (PR), ER, and HER2 were also evaluated. The middle panel shows data from 161 triple-negative (TN) breast cancer cases. The bottom panel shows the lack of HSF1 nuclear expression in 16 normal mammary tissue sections. A summary is provided in the bar graph (right).

(C) HSF1 IHC showing high-level nuclear staining in indicated tumors; T, Tumor; N, Normal adjacent tissue. A summary is provided in the bar graph.

activation, we obtained human tumor samples from breast and colon adenocarcinomas that had been cryopreserved and were of a quality suitable for ChIP-Seq analysis (Figures 5D and S5). Obtaining and processing such human tumor specimens for a technique as demanding as genome-wide ChIP-Seq is highly challenging. In addition, many potentially confounding factors are unavoidable (e.g., cell-type heterogeneity due to the presence of blood and stromal elements, areas of necrosis and micro-environmental stress). Despite these difficulties, the distinct HSF1-binding profile we had established with cultured cancer cell lines was clearly conserved in those tumors that expressed high levels of HSF1. Genes (such as *ST13* and *EIF4A2*) that were strongly bound by HSF1 in cancer lines but weakly bound after heat shock in nontransformed cells, were also strongly bound in these human tumor samples taken directly from patients (Figure 5E). Genes that were weakly bound by HSF1 in cancer lines but strongly bound after heat shock in nontransformed cells (such as *HSPA6* and *DNAJC7*) were also weakly bound in patient tumor samples (Figure 5E). These global similarities in HSF1-binding profiles between cancer cell lines and tumor samples, as well as their divergence from heat-shock profiles, were validated by principal component analysis (Figure 5F).

An HSF1-Cancer Signature Identifies Breast Cancer Patients with Poor Outcome

In our prior analysis of the Nurses' Health cohort, HSF1 overexpression and nuclear localization was associated with reduced survival (Santagata et al., 2011). That work, however, was based entirely on HSF1 immunohistochemistry, an approach that is inherently only semiquantitative. To acquire more precise and molecularly defined information about the effects of HSF1 activation in cancer, we asked whether malignant potential and long-term outcomes correlate with the HSF1 transcriptional program identified above. We distilled an "HSF1-cancer signature" of 456 genes that were bound by HSF1 near their transcription start sites (Figure 2). Expression of these genes (Table S4) was interrogated in ten publicly available mRNA data sets derived from breast cancer patients that had been followed for an average of 7.58 years and had known clinical outcomes (referenced in Table S5). In total, these cohorts encompassed nearly 1,600 individuals of diverse national and ethnic origin. We divided each data set into two groups, those with high (top 25%) and those with low (bottom 75%) expression of the HSF1-cancer signature. We performed Kaplan-Meier analysis independently on each data set to assess potential associations between the HSF1-cancer signature and patient outcome: metastasis-free, relapse-free, or overall survival, depending on the reported outcome parameter for that data set. One representative analysis is presented in Figure 6A, the remainder are shown in Figure S6. High expression of our HSF1-cancer signature had a remarkable correlation with poor prognosis (HSF1-CaSig; Figures 6B and S6). In nine of ten independent data sets reported over the past 10 years, the p values ranged from 0.05 to < 0.0001.

Next, we considered a recent finding that many published cancer signatures are not significantly better outcome predictors than random signatures of identical size (Venet et al., 2011). We performed Kaplan-Meier analysis on independent data sets to evaluate associations between 10,000 individual randomly generated gene signatures and patient outcome (compiled data Table S4, example Figure 6C). A meta-analysis of the breast data sets showed that the HSF1-CaSig outperformed all 10,000 random gene signatures (Monte Carlo p value across breast data sets < 0.0001, Table S4).

Our HSF1-cancer signature was more broadly associated with outcome than other well-established prognostic indicators (Figures 6B and S6) including the oncogene *MYC*, the proliferation marker Ki67 and even MammaPrint, an expression-based diagnostic tool used in routine clinical practice (Kim and Paik, 2010). Because various HSPs have been implicated as prognostic markers for a range of cancers, including breast cancer (Ciocca and Calderwood, 2005), we also tested many individual *HSP* transcripts for possible association with outcome. None of these genes, or even a panel of HSP genes, was as strongly associated with poor outcome as our broader HSF1-cancer signature (Figures 6B and S6).

HSF1 Activation Is an Indicator of Poor Outcome in Early Breast Cancer

At the time of diagnosis, the majority of breast cancer patients have ER+ tumors and early-stage disease (ER^+/lymph-node-negative tumors). A small fraction of these patients will experience a recurrence and might benefit from more aggressive treatment, but it is currently very difficult to identify them in advance. We found that our HSF1-cancer signature was significantly associated with metastatic recurrence in women initially diagnosed with ER^+/lymph-node-negative tumors (p value = 0.0149) (Figure 6D).

To further probe the potential prognostic value of HSF1 in this particularly challenging population, we returned to the Nurses' Health Study cohort because it provides one of the largest collections of patients with ER^+/lymph-node-negative tumors for evaluation (n = 947) and has the longest patient follow up. Because RNA samples are not available from this collection (initiated in 1976), we could assess only the levels and nuclear localization of HSF1. Survival decreased as HSF1 nuclear levels increased in a dose-dependent manner (p value = 0.0015; Figure 6E). This finding was validated by multivariate analysis that showed high-level nuclear HSF1 to be associated with a nearly 100% increase in mortality (Table S6).

(D) ChIP-Seq analysis of human breast and colon cancer surgical resection specimens (patient tumors). Heat map depicting ChIP-Seq read density for all HSF1 target regions defined in Figure 3A. For reference, the binding profiles for cancer cell lines in culture (black; average across BT20, NCIH838 and SKBR3) and parental heat-shocked cell lines (red) are included. HSF1 expression was evaluated by IHC in the same patient tumors used for ChIP-Seq (see Figure S5C) and scored as in (B).

(E) HSF1 binding in cell lines compared to patient tumors. Average binding across cancer cell lines in cell culture (black; average across BT20, NCIH838 and SKBR3), parental heat-shocked cell lines (red), and patient tumors (cyan) are depicted for the representative target genes indicated.

(F) PCA of HSF1 binding in heat-shocked parental cell lines (red), cancer cells lines (black) and patient tumors (cyan). See also Figure S5 and Table S1.

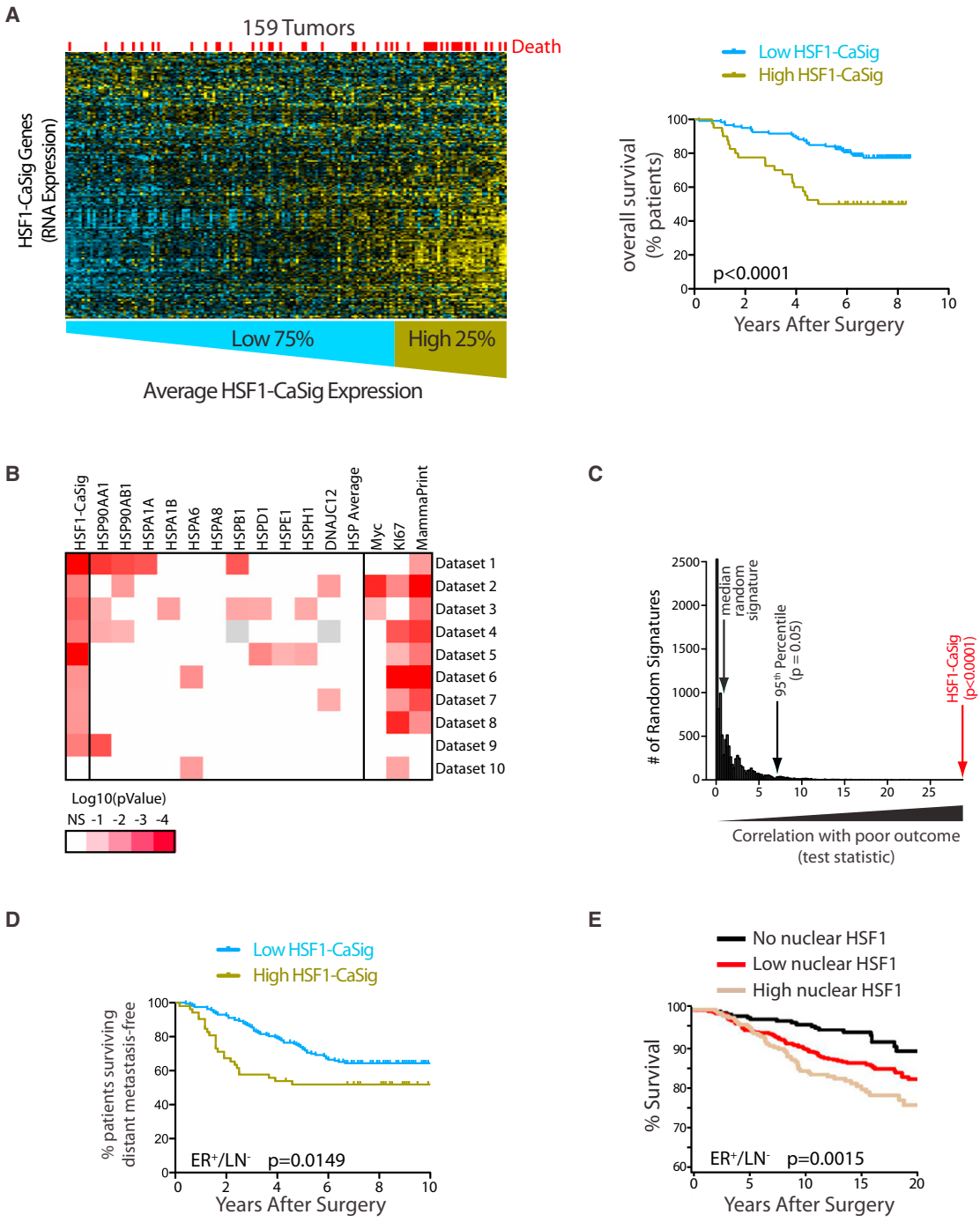

Figure 6. An HSF1-Cancer Signature Is Associated with Reduced Survival in Patients with Breast Cancer

(A) Representative data set (Pawitan et al., 2005) is shown from a meta-analysis of 10 publicly available mRNA expression data sets (Table S5) derived from human breast tumors with known clinical outcome and representing a total of 1594 patients. Each column corresponds to a tumor, and each row corresponds to a microarray probe for an HSF1-cancer signature (HSF1-CaSig) gene. Median levels of expression are depicted in black, increased expression in yellow, and decreased expression in blue. Tumors are ordered by average level of expression of the HSF1-cancer signature, from low (blue) to high (yellow). Red bars indicate deaths. Kaplan-Meier (KM) analysis of the tumors with high expression of the HSF1-cancer signature (top 25%, "High HSF1-CaSig," yellow) versus low expressors (bottom 75%, "Low HSF1-CaSig," blue) is shown.

(B) Log-rank p values for each of the indicated classifiers were calculated for each data set; results are displayed as a heat map. Corresponding KM curves are provided in Figure S6.

(C) Random gene signature analysis of a representative data set (Pawitan et al., 2005). KM analysis on the data set to evaluate associations between 10,000 individual randomly generated gene signatures and patient outcome. The random signatures are binned and ordered from least significant to most significant by

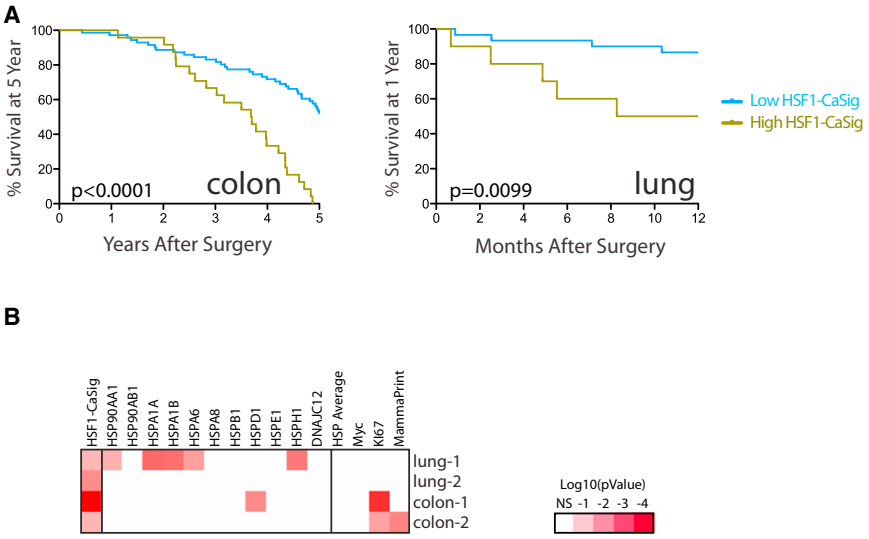

Figure 7. An HSF1-Cancer Signature Is Associated with Reduced Survival in Patients with Colon or Lung Cancers

(A) KM analysis of survival in patients with colon or lung cancer based on Low HSF1-CaSig (blue) or High HSF1-CaSig (yellow). Log-rank p values are shown.

(B) Heat map of log-rank p values for each of the indicated classifiers in four data sets is shown. Corresponding KM curves are in Figure S7. See also Tables S4 and S5.

HSF1-Cancer Signature Is Associated with Poor Outcome in Diverse Human Cancers

Next, we asked whether the HSF1-cancer signature might have prognostic value beyond breast cancer. Analyzing multiple independent gene expression data sets that include outcomes data, increased expression of the HSF1 cancer program in colon and lung cancers was strongly associated with reduced survival (Figures 7A and 7B). The HSF1-CaSig outperformed all 10,000 random gene signatures in these data sets (Monte Carlo p value across data sets < 0.0001, Table S4). Again, our HSF1-cancer signature was more significantly associated with outcome than any individual *HSP* transcript or even a panel of *HSP* genes (Figures 7B and S7). As expected, the MammaPrint expression signature, which was computationally derived by using breast cancers, was a poor indicator of outcome in lung and colon cancers (significant in one of four data sets). Additional HSF1 signatures also comprising positively regulated genes (from Module 1 and 2 of our gene-gene correlation analysis; HSF1-CaSig2) or containing both positively and negatively regulated genes (HSF1-CaSig3) were also strongly associated with patient outcome across tumor types (Table S4). We conclude that the HSF1 cancer program that we have identified supports the malignant state in a diverse spectrum of cancers because it regulates core processes rooted in fundamental tumor biology that ultimately affect outcome.

DISCUSSION

We have defined a distinct genome-wide transcriptional program that HSF1 coordinates in malignancy. This program includes some heat-shock proteins well known to be involved in oncogenic programs, such as HSP90 (Whitesell and Lindquist, 2005). However, it differs fundamentally from the HSF1 program induced by thermal stress, in that it includes many genes that are not induced by heat shock and does not include many that typically are. This cancer program is commonly activated in a wide variety of human malignancies. It is strongly associated with metastasis and death in at least the three cancers responsible for ~30% of all cancer-related deaths worldwide: those of the breast, colon and lung. The very broad range of tumors in which we see immunohistochemical evidence of HSF1 activation suggests it will play a pervasive role throughout tumor biology.

What types of cellular processes does HSF1 regulate in cancer? They constitute an astonishingly diverse group that extends far beyond protein folding and includes energy metabolism, cell cycle signaling, DNA repair, apoptosis, cell adhesion, extracellular matrix formation, and translation. Some of these processes were previously known to be affected by HSF1 (Dai et al., 2007; Jin et al., 2011; Zhao et al., 2009). However, the common assumption had been that HSF1's effects were mediated primarily by HSP chaperone activities (Jin et al., 2011; Meng et al., 2010; Solimini et al., 2007). The remarkable breadth of the HSF1 cancer program in humans explains why HSF1 is such a powerful modifier of tumorigenesis in multiple animal models (Dai et al., 2007; Jin et al., 2011; Zhao et al., 2009) and why HSF1 was identified as one of only six potent metastasis-promoting genes in a genome-wide screen for enhancers of invasion by malignant melanoma cells (Scott et al., 2011).

Not only is the repertoire of HSF1-regulated genes in cancer much larger than just heat-shock genes, but even the manner in which some of the classical heat-shock genes are regulated differs. For example, *HSPA6* (HSP70B'), a pillar of the heat-shock response, differs dramatically in these two states. Following heat stress, *HSPA6* is the most highly induced of all mRNAs, yet, surprisingly in cancer, *HSPA6* is only bound very weakly by HSF1. Its expression is not changed following HSF1 depletion and its transcript level does not correlate with that of HSF1 in our meta-analysis of 12,000 gene expression experiments.

the KM-generated test statistic. The Red arrow indicates the test statistic of the HSF1-CaSig. For reference, black arrows indicate the test statistic of the random signature with the median test statistic (5,000th) and the random signature with the 95th percentile test statistic.

(D) KM analysis of individuals with ER+/lymph-node-negative tumors (Wang et al., 2005) with Low HSF1-CaSig (blue) or High HSF1-CaSig (yellow).

(E) KM analysis of 947 individuals from the NHS with ER+, lymph-node-negative tumors expressing no, low, or high nuclear HSF1 as measured by IHC. Data are from the NHS (1976–1997). Log-rank p values are shown. See also Figure S6 and Tables S4, S5, and S6.

What could account for activation of a distinct HSF1-regulated program in cancer? After many years of investigation, we do not yet fully understand how HSF1 activity is regulated during the classic heat-shock response. Multiple mechanisms have been described. These include the release of HSF1 from its normal sequestration by chaperones when unfolded substrates compete for chaperone binding. But in addition, HSF1 is subject to an extensive array of posttranslational modifications (at least 30) including acetylation, sumoylation, and numerous phosphorylations (Anckar and Sistonen, 2011).

Some of these heat-shock regulatory mechanisms are likely shared by cancer cells. For instance, impaired protein homeostasis driven by the accumulation of mutant, misfolding-prone oncoproteins, aneuploidy, and the increased rate of translation in cancer could chronically stimulate HSF1 activation by releasing it from sequestration from chaperones (Anckar and Sistonen, 2011). Dysregulation of signaling pathways in cancer could also drive posttranslational modifications to HSF1. Some of these (such as those responsible for phosphorylation at serine 326) will likely be shared with heat-shocked cells. But others will likely be unique to cancer. Indeed, it seems extremely likely that different mechanisms of activation will operate in different cancers. Several pathways activated in cancer such as EGFR/HER2 axis (Zhao et al., 2009), the RAS/MAPK (Stanhill et al., 2006) or the insulin/IGFI-like growth factor system (Chiang et al., 2012) have all been reported to alter HSF1 activity. Additional modes of cancer-specific regulation might include epigenetic states common to cancer and proliferating cells and transcriptional coregulators.

How might the distinct transcriptional program regulated by HSF1 in malignancy have arisen? The association of this program with metastasis and death points to an evolutionary origin distinct from cancer itself. The broad range of cancer types in which we find HSF1 activated suggests that this program originated to support basic biological processes. Indeed, the sole heat-shock factor in yeast (yHSF), even at basal temperatures, binds many genes that are involved in a wide-range of core cellular functions (Hahn et al., 2004). These transcriptional targets allow yeast not only to adapt to environmental contingencies but also to modulate metabolism and maintain proliferation under normal growth conditions (Hahn et al., 2004; Hahn and Thiele, 2004). As a result, yHSF is essential for viability, paralleling the importance of HSF1 for the survival and proliferation of cancer cells (Dai et al., 2007). Activation of HSF1 may also be required in animals in states of high proliferation and altered metabolism such as immune activation and wound healing (Rokavec et al., 2012; Xiao et al., 1999; Zhou et al., 2008). Moreover, in diverse eukaryotes, HSF1 is a well-validated longevity factor; nonclassical activation of this transcription factor could be highly relevant in this context (Chiang et al., 2012; Volovik et al., 2012).

Ironically, the evolutionarily ancient role played by HSF1 in helping cells to adapt, survive, and proliferate is co-opted frequently to support highly malignant cancers. By enabling oncogenesis, the activation of this ancient prosurvival mechanism thereby actually impairs survival of the host. HSF1 activation in a particular tumor may reflect the degree to which accumulated oncogenic mutations have disrupted normal physiology even before overt invasion or metastasis occurs. This interpretation would explain the impressively broad prognostic value of our HSF1-cancer signature across disparate cancers and even at early stages of disease. Clinical implementation will require further refinement of the signature and validation in tissue and RNA samples from multiple clinical cohorts. Such studies are certainly warranted. As just one potential application, it might aid in the identification of indolent tumors that do not require intervention, reducing the burdens of unnecessary treatment (Kalager et al., 2012). In addition to its prognostic value, HSF1 and diverse regulators that activate it might prove useful targets for cancer therapeutics.

Our understanding of the extensive role played by HSF1 in supporting cancers continues to mature. The protein has been defined for decades by its ability to coordinate chaperone protein expression and enhance survival in the face of heat stress (Christians et al., 2002; Ritossa, 1962). Although appreciating the importance of these classical mechanisms, the role of HSF1 is clearly much broader and deeper.

EXPERIMENTAL PROCEDURES

Cell Culture Methods

Cell lines were cultured as detailed in Extended Experimental Procedures.

ChIP-Seq and ChIP-qPCR

ChIP-qPCR and ChIP-Seq experiments were performed as described previously (Lee et al., 2006), with modifications detailed in Extended Experimental Procedures.

Gene Expression Analysis

Lentiviral shRNA methods were described previously (Dai et al., 2007). Gene expression analysis was performed as described in Extended Experimental Procedures. Microarray data were deposited in NCBI Gene Expression Omnibus. RT-PCR analysis, gene-gene correlation analysis of HSF1-bound genes, and correlation of HSF1-bound gene expression with outcome is detailed in Extended Experimental Procedures.

Immunohistochemistry and The NHS Analysis

Paraffin sections were stained with HSF1 antibody (Thermo Scientific, RT-629-PABX) as detailed in Extended Experimental Procedures. The NHS is a prospective cohort study initiated in 1976 (Hu et al., 2011; Tamimi et al., 2008). For design and study population, and analysis, see Extended Experimental Procedures.

Statistical Analysis

Correlation of gene expression with location of HSF1 occupancy was performed by using a two-tailed Fisher's Exact Test. P values for significance of overlap between pairs of gene sets were generated by using the hypergeometric distribution. Statistical methods for ChIP-Seq analysis and the Nurses' Health Study outcome data analysis are detailed in Extended Experimental Procedures. Kaplan-Meier analysis was used to compare outcome events. P values were generated by using the logrank test. For all other data, mean ± standard deviation is reported and statistical significance between means was determined by using a two-tailed t test.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2012.06.031.

ACKNOWLEDGMENTS

We thank G. Frampton, I. Barrasa and S.Gupta for bioinformatic assistance. We thank T. Mazor, T. Volkert and the WIBR-GTC for sequencing support. We thank the Lindquist lab and T. Lee for discussion and K. Allendoerfer, B. Bevis, G. Karras, R. Shouval and K. Matlack for comments. The work was supported by the J&J COSAT focused funding program (L.W.) and the Marble Fund (S.L.). S.L. is an Investigator of the Howard Hughes Medical Institute. M.L.M. is supported by American Cancer Society New England Division-SpinOdyssey (PF-09-253-01-DMC). S.S. is supported by NIH (K08NS064168), the Brain Science Foundation and the V Foundation. Additional support was provided by GSK (WE234 EPI40307); Public Health Service Grants CA087969, and SPORE in Breast Cancer CA089393, from the NCI, NIH, Dept. of Health and Human Services. T.A.I is supported by the Breast Cancer Research Foundation, NCI (R01-CA146445-01) and DoD-CDMRP Breast Cancer Research Program (W81XWH-08-1-0282 BC-07456). T.A. Ince receive royalty payments for WIT medium and was a consultant to Stemgent Inc. during 2008–2010. We thank the participants and staff of the NHS, and state cancer registries: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY.

Received: October 13, 2011
Revised: April 10, 2012
Accepted: June 4, 2012
Published: August 2, 2012

REFERENCES

Anckar, J., and Sistonen, L. (2011). Regulation of HSF1 function in the heat stress response: implications in aging and disease. Annu. Rev. Biochem. *80*, 1089–1115.

Chiang, W.C., Ching, T.T., Lee, H.C., Mousigian, C., and Hsu, A.L. (2012). HSF-1 regulators DDL-1/2 link insulin-like signaling to heat-shock responses and modulation of longevity. Cell *148*, 322–334.

Christians, E.S., Yan, L.J., and Benjamin, I.J. (2002). Heat shock factor 1 and heat shock proteins: Critical partners in protection against acute cell injury. Crit. Care Med. *30*, S43–S50.

Ciocca, D.R., and Calderwood, S.K. (2005). Heat shock proteins in cancer: diagnostic, prognostic, predictive, and treatment implications. Cell Stress Chaperones *10*, 86–103.

Dai, C., Whitesell, L., Rogers, A.B., and Lindquist, S. (2007). Heat shock factor 1 is a powerful multifaceted modifier of carcinogenesis. Cell *130*, 1005–1018.

Day, A., Dong, J., Funari, V.A., Harry, B., Strom, S.P., Cohn, D.H., and Nelson, S.F. (2009). Disease gene characterization through large-scale co-expression analysis. PLoS ONE *4*, e8491.

Dowhan, D.H., Hong, E.P., Auboeuf, D., Dennis, A.P., Wilson, M.M., Berget, S.M., and O'Malley, B.W. (2005). Steroid hormone receptor coactivation and alternative RNA splicing by U2AF65-related proteins CAPERalpha and CAPERbeta. Mol. Cell *17*, 429–439.

Guettouche, T., Boellmann, F., Lane, W.S., and Voellmy, R. (2005). Analysis of phosphorylation of human heat shock factor 1 in cells experiencing a stress. BMC Biochem. *6*, 4.

Hahn, J.S., and Thiele, D.J. (2004). Activation of the Saccharomyces cerevisiae heat shock transcription factor under glucose starvation conditions by Snf1 protein kinase. J. Biol. Chem. *279*, 5169–5176.

Hahn, J.S., Hu, Z., Thiele, D.J., and Iyer, V.R. (2004). Genome-wide analysis of the biology of stress responses through heat shock transcription factor. Mol. Cell. Biol. *24*, 5249–5256.

Hu, R., Dawood, S., Holmes, M.D., Collins, L.C., Schnitt, S.J., Cole, K., Marotti, J.D., Hankinson, S.E., Colditz, G.A., and Tamimi, R.M. (2011). Androgen receptor expression and breast cancer survival in postmenopausal women. Clin. Cancer Res. *17*, 1867–1874.

Ince, T.A., Richardson, A.L., Bell, G.W., Saitoh, M., Godar, S., Karnoub, A.E., Iglehart, J.D., and Weinberg, R.A. (2007). Transformation of different human breast epithelial cell types leads to distinct tumor phenotypes. Cancer Cell *12*, 160–170.

Ishikawa, N., Takano, A., Yasui, W., Inai, K., Nishimura, H., Ito, H., Miyagi, Y., Nakayama, H., Fujita, M., Hosokawa, M., et al. (2007). Cancer-testis antigen lymphocyte antigen 6 complex locus K is a serologic biomarker and a therapeutic target for lung and esophageal carcinomas. Cancer Res. *67*, 11601–11611.

Jin, X., Moskophidis, D., and Mivechi, N.F. (2011). Heat shock transcription factor 1 is a key determinant of HCC development by regulating hepatic steatosis and metabolic syndrome. Cell Metab. *14*, 91–103.

Kalager, M., Adami, H.O., Bretthauer, M., and Tamimi, R.M. (2012). Overdiagnosis of invasive breast cancer due to mammography screening: results from the Norwegian screening program. Ann. Intern. Med. *156*, 491–499.

Khaleque, M.A., Bharti, A., Gong, J., Gray, P.J., Sachdev, V., Ciocca, D.R., Stati, A., Fanelli, M., and Calderwood, S.K. (2008). Heat shock factor 1 represses estrogen-dependent transcription through association with MTA1. Oncogene *27*, 1886–1893.

Kim, C., and Paik, S. (2010). Gene-expression-based prognostic assays for breast cancer. Nat Rev Clin Oncol *7*, 340–347.

Lee, T.I., Johnstone, S.E., and Young, R.A. (2006). Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat. Protoc. *1*, 729–748.

Lee, Y.J., Lee, H.J., Lee, J.S., Jeoung, D., Kang, C.M., Bae, S., Lee, S.J., Kwon, S.H., Kang, D., and Lee, Y.S. (2008). A novel function for HSF1-induced mitotic exit failure and genomic instability through direct interaction between HSF1 and Cdc20. Oncogene *27*, 2999–3009.

Liberal, V., Martinsson-Ahlzen, H.S., Liberal, J., Spruck, C.H., Widschwendter, M., McGowan, C.H., and Reed, S.I. (2011). Breast Cancer Special Feature: Cyclin-dependent kinase subunit (Cks) 1 or Cks2 overexpression overrides the DNA damage response barrier triggered by activated oncoproteins. Proc. Natl. Acad. Sci. USA *109*, 2754–2759.

Luo, J., Emanuele, M.J., Li, D., Creighton, C.J., Schlabach, M.R., Westbrook, T.F., Wong, K.K., and Elledge, S.J. (2009). A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. Cell *137*, 835–848.

Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics *27*, 1696–1697.

MacIsaac, K.D., Lo, K.A., Gordon, W., Motola, S., Mazor, T., and Fraenkel, E. (2010). A quantitative model of transcriptional regulation reveals the influence of binding location on expression. PLoS Comput. Biol. *6*, e1000773.

Maruyama, M., Yoshitake, H., Tsukamoto, H., Takamori, K., and Araki, Y. (2010). Molecular expression of Ly6k, a putative glycosylphosphatidylinositol-anchored membrane protein on the mouse testicular germ cells. Biochem. Biophys. Res. Commun. *402*, 75–81.

Meng, L., Gabai, V.L., and Sherman, M.Y. (2010). Heat-shock transcription factor HSF1 has a critical role in human epidermal growth factor receptor-2-induced cellular transformation and tumorigenesis. Oncogene *29*, 5204–5213.

Min, J.N., Huang, L., Zimonjic, D.B., Moskophidis, D., and Mivechi, N.F. (2007). Selective suppression of lymphomas by functional loss of Hsf1 in a p53-deficient mouse model for spontaneous tumors. Oncogene *26*, 5086–5097.

Morimoto, R.I. (2008). Proteotoxic stress and inducible chaperone networks in neurodegenerative disease and aging. Genes Dev. *22*, 1427–1438.

Page, T.J., Sikder, D., Yang, L., Pluta, L., Wolfinger, R.D., Kodadek, T., and Thomas, R.S. (2006). Genome-wide analysis of human HSF1 signaling reveals a transcriptional program linked to cellular adaptation and survival. Mol. Biosyst. *2*, 627–639.

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

Pelham, H.R. (1982). A regulatory upstream promoter element in the Drosophila hsp 70 heat-shock gene. Cell *30*, 517–528.

Ritossa, F. (1962). A new puffing pattern induced by temperature shock and DNP in Drosophila. Cellular and Molecular Life Sciences *18*, 571–573.

Rokavec, M., Wu, W., and Luo, J.L. (2012). IL6-mediated suppression of miR-200c directs constitutive activation of inflammatory signaling circuit driving transformation and tumorigenesis. Mol. Cell *45*, 777–789.

Sakurai, H., and Enoki, Y. (2010). Novel aspects of heat shock factors: DNA recognition, chromatin modulation and gene expression. FEBS J. *277*, 4140–4149.

Santagata, S., Hu, R., Lin, N.U., Mendillo, M.L., Collins, L.C., Hankinson, S.E., Schnitt, S.J., Whitesell, L., Tamimi, R.M., Lindquist, S., and Ince, T.A. (2011). High levels of nuclear heat-shock factor 1 (HSF1) are associated with poor prognosis in breast cancer. Proc. Natl. Acad. Sci. USA *108*, 18378–18383.

Santagata, S., Xu, Y.M., Wijeratne, E.M., Kontnik, R., Rooney, C., Perley, C.C., Kwon, H., Clardy, J., Kesari, S., Whitesell, L., et al. (2012). Using the heat-shock response to discover anticancer compounds that target protein homeostasis. ACS Chem. Biol. *7*, 340–349.

Scott, K.L., Nogueira, C., Heffernan, T.P., van Doorn, R., Dhakal, S., Hanna, J.A., Min, C., Jaskelioff, M., Xiao, Y., Wu, C.J., et al. (2011). Proinvasion metastasis drivers in early-stage melanoma are oncogenes. Cancer Cell *20*, 92–103.

Shamovsky, I., and Nudler, E. (2008). New insights into the mechanism of heat shock response activation. Cell. Mol. Life Sci. *65*, 855–861.

Solimini, N.L., Luo, J., and Elledge, S.J. (2007). Non-oncogene addiction and the stress phenotype of cancer cells. Cell *130*, 986–988.

Stanhill, A., Levin, V., Hendel, A., Shachar, I., Kazanov, D., Arber, N., Kaminski, N., and Engelberg, D. (2006). Ha-ras(val12) induces HSP70b transcription via the HSE/HSF1 system, but HSP70b expression is suppressed in Ha-ras(val12)-transformed cells. Oncogene *25*, 1485–1495.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

Tamimi, R.M., Baer, H.J., Marotti, J., Galan, M., Galaburda, L., Fu, Y., Deitz, A.C., Connolly, J.L., Schnitt, S.J., Colditz, G.A., and Collins, L.C. (2008). Comparison of molecular phenotypes of ductal carcinoma in situ and invasive breast cancer. Breast Cancer Res. *10*, R67.

Venet, D., Dumont, J.E., and Detours, V. (2011). Most random gene expression signatures are significantly associated with breast cancer outcome. PLoS Comput. Biol. *7*, e1002240.

Volovik, Y., Maman, M., Dubnikov, T., Bejerano-Sagie, M., Joyce, D., Kapernick, E.A., Cohen, E., and Dillin, A. (2012). Temporal requirements of heat shock factor-1 for longevity assurance. Aging Cell *11*, 491–499.

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.

Whitesell, L., and Lindquist, S.L. (2005). HSP90 and the chaperoning of cancer. Nat. Rev. Cancer *5*, 761–772.

Xiao, X., Zuo, X., Davis, A.A., McMillan, D.R., Curry, B.B., Richardson, J.A., and Benjamin, I.J. (1999). HSF1 is required for extra-embryonic development, postnatal growth and protection during inflammatory responses in mice. EMBO J. *18*, 5943–5952.

Zhao, Y.H., Zhou, M., Liu, H., Ding, Y., Khong, H.T., Yu, D., Fodstad, O., and Tan, M. (2009). Upregulation of lactate dehydrogenase A by ErbB2 through heat shock factor 1 promotes breast cancer cell glycolysis and growth. Oncogene *28*, 3689–3701.

Zhao, Y., Liu, H., Liu, Z., Ding, Y., Ledoux, S.P., Wilson, G.L., Voellmy, R., Lin, Y., Lin, W., Nahta, R., et al. (2011). Overcoming trastuzumab resistance in breast cancer by targeting dysregulated glucose metabolism. Cancer Res. *71*, 4585–4597.

Zhou, J.D., Luo, C.Q., Xie, H.Q., Nie, X.M., Zhao, Y.Z., Wang, S.H., Xu, Y., Pokharel, P.B., and Xu, D. (2008). Increased expression of heat shock protein 70 and heat shock factor 1 in chronic dermal ulcer tissues treated with laser-aided therapy. Chin. Med. J. (Engl.) *121*, 1269–1273.

EXTENDED EXPERIMENTAL PROCEDURES

Cell Culture Methods

HME, HMLER and MCF10A cells were cultured in MEGM medium supplemented as specified by the manufacturer (Lonza). BPE and BPLER cells were cultured in WIT-I and WIT-T medium, respectively, in accordance with recommendations by the manufacturer (Stemgent). The HME, BPE, HMLER and BPLER cells are available from the Ince laboratory upon request. BT474, H441, H838, H1703, HCC38, HCC1954, HCT15, HT29, SKBR3, SW620 and ZR75-1 cells were cultured in RPMI-1640 medium supplemented with 10% fetal bovine serum. BT20, MDA-MB-231, MCF7 and T47D cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum.

ChIP Antibodies

For ChIP-Seq, HSF1 antibody (Santa Cruz, sc-9144) and normal rabbit IgG (Santa Cruz, sc-2027) were used. For ChIP-qPCR, HSF1 antibody (Santa Cruz, sc-9144) and, as a control, a second HSF1 antibody (Thermo Scientific, RT-629-PABX), were used. Similar results were obtained and RT-629-PABX antibody data are reported. Additionally, (RNA polymerase II CTD repeat YSPTSPS antibody [4H8] (Abcam, ab5408) and normal rabbit IgG (Santa Cruz, sc-2027) were used, as indicated.

ChIP-Seq and ChIP-qPCR

For ChIP-Seq, $5x10^7$ cells were used for each immunoprecipitation. For heat-shock, cells were transferred to a 42° (5% CO_2) incubator for 1 hr. ChIP and ChIP-Seq experiments were performed as described previously (Lee et al., 2006) with several modifications (Novershtern et al., 2011). In place of RIPA buffer, immunoprecipitations were washed sequentially with buffer B (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 2 mM EDTA, pH 8.0, 0.1% SDS and 1.0% Triton X-100), buffer C (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 2 mM EDTA, pH 8.0, 0.1% SDS and 1.0% Triton X-100), buffer D (10 mM Tris-HCl, pH 8.0, 250 mM LiCl, 1 mM EDTA, pH 8.0, 1.0% Na-Deoxycholate and 1.0% IGEPAL CA-630), and buffer TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0, 50 mM NaCl). Preparation of the ChIP-Seq DNA library and deep sequencing using an Illumina Solexa genome analyzer were performed as described previously (Yu et al., 2009).

Images acquired from the Illumina sequencer were processed through the bundled Illumina image extraction pipeline. ChIP-Seq reads were aligned to HG18 with ELAND software (Illumina). Identification of enriched genomic regions was performed as described previously (Guenther et al., 2008). Briefly, each ChIP-Seq read (a maximum of two repeat reads were allowed) was extended 100 bp to approximate the middle of the sequenced fragment. The extended fragments were subsequently allocated to 25 bp bins across the genome. Read density for each bin was calculated and enriched bins were identified by comparison to a Poisson background model using a p value threshold of 10^{-12}. The minimum ChIP-seq read density required to meet this threshold for each data set is indicated in Table S1. Enriched bins within 200 bp were combined to form enriched regions. Enriched regions less than 100 bp were removed. Because of the nonrandom nature of background reads, enriched bins and regions were also required to have an eight-fold greater ChIP-seq density versus a nonspecific control IgG immunoprecipitation performed under identical conditions. All RefSeq genes that were within 8 kb of enriched regions were considered to be enriched genes. Enriched genes and regions for all experiments are listed in Table S1. The raw data have been deposited in a public database (NCBI Gene Expression Omnibus).

The unions of all HSF1 enriched regions identified by ChIP-Seq (Table S1) in each sample were merged to identify a global set of regions. Short reads overlapping these regions were quantified by using HTSeq-count (http://www-huber.embl.de/users/anders/HTSeq/doc/count.html). The counts matrix was median-normalized by using the total number of mapped reads. After adding 1 pseudocount, counts were log2-normalized and analyzed by principal components as implemented by the MADE4 program in Bioconductor (Culhane et al., 2005).

Preparation of human breast and colon tumors for ChIP-seq was performed by using 300 mg of cryopreserved material. Frozen tumor tissue was retrieved from the Brigham and Women's Hospital (BWH) Tissue Bank in accordance with the regulations for excess tissue use stipulated by the BWH institutional review board. Frozen sections for immunohistochemistry were prepared using a cryostat from adjacent tissue. Frozen samples were processed for ChIP-Seq using a tissue pulverizer, and this material was subsequently suspended in PBS and passed serially through needles of increasing gauge. This suspension was then fixed for 10 min and the pellet was processed as described above.

For ChIP-qPCR, 5×10^6 cells were used for each immunoprecipitation. The protocol was modified as described above. RT^2 SYBR Green qPCR Mastermix (SABiosciences) was used with the indicated oligo pairs (Table S7) on a 7700 ABI Detection System.

Gene Expression Analysis

Lentiviral shRNA sequences, viral production and transduction of cells have been described previously (Dai et al., 2007). RNA was purified following extraction with TRIzol reagent (Invitrogen, #15596-026), 60 hr after viral infection. Protein lysates of concurrent infections were prepared in TNES buffer consisting of 50 mM Tris, pH 7.4; NP-40 1%; EDTA 2 mM; NaCl 200 mM plus protease inhibitor cocktail (Roche Diagnostics, Cat# 11836153001). Protein concentration was measured by BCA assay (Thermo Fisher Scientific 23227) and 15 μg total protein/lane was analyzed by SDS-PAGE and immunoblotting using rat monoclonal anti-HSF1 antibody cocktail (Ab4, Thermo Scientific, 1:1,000 dilution) and Actin Monoclonal Antibody (mAbGEa; clone DM1A, Thermo Scientific, 1:1,000). Because prolonged depletion of HSF1 is toxic to malignant cells (Dai et al., 2007), we analyzed mRNA expression early, before

HSF1 knockdown was complete and cell viability was grossly impaired. Thus, results likely underestimate the effects of HSF1 on gene expression in malignant cells. For gene expression after heat-shock, cells were transferred to a 42°C (5% CO_2) incubator for 1 hr and allowed to recover for 30 min in a 37°C (5% CO_2) incubator before RNA extraction. Gene expression analysis was performed by using an Affymetrix GeneChip HT Human Genome U133 96-Array Plate and data were analyzed by using previously described methods (Ince et al., 2007). All microarray raw data were deposited in a public database (NCBI Gene Expression Omnibus GSE38232).

For evaluating the effects of HSF1 knockdown on the expression of target genes, HSF1 was depleted by using siRNA (Dharmacon, Lafayette, CO): M012109-01 siGenome SMART pool, Human HSF1 (target sequences: UAGCCUGCCUGGACAAGAA;CCACUUG GAUGCUAUGGAC; GAGUGAAGACAUAAAGAUC; AGAGAGACGACACGGAGUU). siGLO RISC-Free siRNA (D-001600-01) and siGENOME Nontargeting siRNA #5 (D-001210-05) were used as controls. Cells were transfected by using Lipofectamine RNAiMAX Transfection Reagent (Invitrogen, #13778) and were harvested in Trizol (Invitrogen, #15596-026). RNA was purified by using Direct-zol RNA MiniPrep (Zymo Research, Irving, CA). Quantitative PCR to evaluate mRNA levels was performed as described above by using RT^2 SYBR Green qPCR Mastermix (SABiosciences) and primer assay pairs (SABiosciences; Valencia, CA) on a 7700 ABI Detection System.

Gene-Gene Correlation Analysis

Correlation values of HSF1-bound genes were determined by using the UCLA Gene Expression Tool (genome.ucla.edu/projects/UGET) to query gene expression profile data collected in Celsius, a data warehousing system that aggregates Affymetrix CEL files and associated metadata (Day et al., 2007; Day et al., 2009). Nearly 12,000 Affymetrix HG-U133 Plus 2.0 human gene expression profiles, predominantly representing neoplasms of highly diverse human origin, were interrogated. A pair-wise correlation matrix was built by assessing genes bound in at least two of the three cell lines with most robust HSF1 activation (BT20, NCIH838, SKBR3). This generated 1042 genes. The final map as displayed contains 709 unique genes, with genes required to have an absolute value of the correlation coefficient > 0.3 ($|a| > 0.3$) with at least 100 other genes. Data were ordered by using hierarchical clustering (correlation centered, average linkage).

Correlation of Gene Expression with Outcome

The "HSF1-CaSig" was generated from the 456 genes that were bound in BPLER cells by HSF1 near their transcription start sites (bound from −8kb to +2kb of the TSS) (Table S4 for list of genes). The HSF1-CaSig2 was generated from the genes found in Modules 1 and 2 of our gene-gene correlation analysis (Figure 4B). Genes within Module 1 showed strong positive correlation with the expression of *HSF1* mRNA itself, and Module 2 was positively correlated with Module 1 (Table S4 for list of genes). The HSF1-CaSig3 was derived by using three training data sets (Hou et al., 2010; Jorissen et al., 2009; Pawitan et al., 2005) Briefly, the 300 genes most positively correlated and the 150 most negatively correlated (by t test statistic) with poor outcome were identified in each data set. Genes present in at least two of three data sets in each group were assembled in the final HSF1-CaSig3 gene signature (Table S4 for list of genes).

We used all breast cancer data sets with reported clinical outcome available in the Oncomine database (Rhodes et al., 2007) containing at least 70 tumors, excluding several data sets based on older microarray platforms that were missing many currently annotated genes. This left 10 high-quality data sets, the majority of which contained more than 150 tumors (Table S5). We stratified each data set into two groups of tumors based on high (highest 25%) and low (lowest 75%) average expression of the gene or gene signature being queried. For analysis of the MammaPrint and the HSF1-CaSig3 gene signature, the subset of genes that positively correlate with poor outcome was positively weighted and the subset of genes that negatively correlate with poor outcome was negatively weighted, as described previously (van 't Veer et al., 2002; van de Vijver et al., 2002). Data for the three versions of the HSF1-CaSig for KM analysis were retrieved from Oncomine (Rhodes et al., 2007).

All data for comparisons with random signatures were obtained from NCBI GEO and KM analysis was repeated (reported in Table S4). If CEL files were available, Affymetrix microarrays were processed with RMA by using Bioconductor; otherwise, preprocessed expression matrices were obtained from NCBI GEO or author web sites. Monte Carlo cross validation was applied to contrast HSF1-CaSig signatures with random signatures of genes of the same number. Random sets of signatures containing the same number of probesets as each HSF1 signature were generated for each data set with a particular emphasis on U133A probesets (present on both U133A and U133 Plus 2.0 arrays). The 10,000 random signatures were processed in the same manner as the original signature, sorting samples by increasing mean expression of each mean-centered probeset. Cancer samples, partitioned into the high and low HSF1-CaSig as before, were then analyzed for survival with the log-rank test, producing 10,000 test statistics. Median p values were calculated across a tumor subtype and Monte Carlo cross validation was applied.

Xenografts

$5x10^6$ HMLER and BPLER cells in a 50/50 mix of PBS/Matrigel were inoculated subcutaneously in the right inguinal region of each mouse by using a 27 g needle. Tumors were removed, and fixed in 10% formalin. Following standard tissue processing, 5μM sections were cut and immunostained as described below.

Immunohistochemistry of Tissues and Scoring

Paraffin blocks of human tumor and normal tissue were obtained from the archives of BWH in accordance with the regulations for excess tissue use stipulated by the BWH institutional review board. Tissue microarrays were purchased from Pantomics (Richmond, CA) for carcinoma of the breast (BRC1501, BRC1502), cervix (CXC1501), colon (COC1503), lung (LUC1501), pancreas (PAC481) and prostate (PRC1961). Whole sections of 40 meningioma specimens were retrieved from the archives of BWH. A TMA of triple-negative breast cancer cases was kindly provided by Dr. Andrea Richardson (BWH). Normal tissue cores on the TMAs and adjacent normal tissues in the whole sections were used to evaluate expression of HSF1 in nonneoplastic tissues.

Formalin-fixed, paraffin-embedded (FFPE) sections were first deparaffinized. Frozen sections were first postfixed in 10% formalin. FFPE or fixed-frozen sections were blocked with 3% H2O2 and antigen retrieval was performed by using a pressure cooker with Dako citrate buffer (pH 6.0) at 120°C ± 2°C, 15 ± 5 PSI. Slides were blocked by using 3% normal rabbit serum, primary HSF1 antibody (1:2000) was applied at room temperature for 40 min, followed by a 30 min incubation with Dako Labeled Polymer-HRP anti-rat IgG as a secondary antibody. Visualization was achieved with 3, 3′ - diaminobenzidine (DAB) as a chromogen (Dako Envision+ System). Counterstaining was performed with Mayer-hematoxylin.

Immunostained sections were scored independently by two pathologists (SS and TAI) by using light microscopy. HSF1 immunostains of FFPE tumor sections were scored by using a 0 to 25 scale in Figure 5. The percent of tumor cells staining with HSF1 was quantified as (0) = 0%; (1+) = 1%–20%; (2+) = 21%–40%; (3+) = 41%–60%; (4+) = 61%–80%; (5+) = 81%–100%. The intensity of nuclear staining was quantified 0 to 5+ relative to negative normal cells. The total HSF1 score was derived by multiplying the percent score with the intensity score. Three tiers of HSF1 staining were defined based on total combined scores of less than 10 (Weak HSF1); 10-18 (Low-Positive HSF1), > 18 (High-Positive HSF1).

For the Nurses' Health Study analysis, an individual value was assigned to each tissue core based on a semiquantitative review of staining intensity with 0 for no nuclear staining, 1 for low-level nuclear staining and 2 for strong nuclear staining. Pathologists were blinded to the survival outcomes of the participants and to the scores given by the other pathologist. Scoring averages were determined per case from values assigned to all evaluable cores from the two independent readings. The case status was recorded as missing if diagnostic tissue was absent or if the staining was uninterpretable for all three cores from a case. The ER, PR and HER2 status of each case was determined as previously described (Dawood et al., 2011).

Immunofluorescence was performed by using 1:250 dilution of rat monoclonal anti-HSF1-antibody cocktail (Ab4, Thermo Scientific, 1:1,000 dilution), 1:100 dilution of rabbit polyclonal anti-p53 (Santa Cruz, #sc-6243) and with fluorescence labeled secondary antibodies. The slides were then reviewed by standard fluorescence microscope.

The Nurses' Health Study Design and Population, Exclusion Criteria and Statistical Analysis

121,700 female US-registered nurses between the ages of 30-55 completed a questionnaire on factors relevant to women's health with follow-up biennial questionnaires used to update exposure information and ascertain nonfatal incident diseases. The NHS breast cancer tissue block collection and tissue microarray (TMA) assembly have been described previously (Hu et al., 2011; Tamimi et al., 2008). This study included women with invasive breast carcinoma that were diagnosed between 1976 and 1996. Participants were excluded from outcome analysis if they had in situ carcinoma only, stage IV breast cancer at the time of diagnosis or if HSF1-status could not be evaluated due to missing cores. Outcome analysis was performed on 947 women. Kaplan-Meier analysis and multivariate analysis are shown for this subset of early-stage ER-positive participants with lymph-node-negative disease. Covariates considered in the multivariate model include age and date of diagnosis, estrogen receptor status, disease stage, tumor grade, radiation treatment, chemotherapy and hormonal treatment. For statistical analysis the survival endpoint was death from breast cancer. Deaths from other causes were censored. Survival curves were estimated by the Kaplan-Meier method and the log-rank test was used to assess statistical significance. Cox proportional hazards regression models were used to evaluate the relationship between HSF1-status and breast cancer-specific mortality after adjusting for covariates. All analyses were run with SAS version 9.1 statistical software.

SUPPLEMENTAL REFERENCES

Bild, A.H., Yao, G., Chang, J.T., Wang, Q., Potti, A., Chasse, D., Joshi, M.B., Harpole, D., Lancaster, J.M., Berchuck, A., et al. (2006). Oncogenic pathway signatures in human cancers as a guide to targeted therapies. Nature *439*, 353–357.

Culhane, A.C., Thioulouse, J., Perrière, G., and Higgins, D.G. (2005). MADE4: an R package for multivariate analysis of gene expression data. Bioinformatics *21*, 2789–2790.

Dai, C., Whitesell, L., Rogers, A.B., and Lindquist, S. (2007). Heat shock factor 1 is a powerful multifaceted modifier of carcinogenesis. Cell *130*, 1005–1018.

Dawood, S., Hu, R., Homes, M.D., Collins, L.C., Schnitt, S.J., Connolly, J., Colditz, G.A., and Tamimi, R.M. (2011). Defining breast cancer prognosis based on molecular phenotypes: results from a large cohort study. Breast Cancer Res. Treat. *126*, 185–192.

Day, A., Carlson, M.R., Dong, J., O'Connor, B.D., and Nelson, S.F. (2007). Celsius: a community resource for Affymetrix microarray data. Genome Biol. *8*, R112.

Day, A., Dong, J., Funari, V.A., Harry, B., Strom, S.P., Cohn, D.H., and Nelson, S.F. (2009). Disease gene characterization through large-scale co-expression analysis. PLoS ONE *4*, e8491.

Desmedt, C., Piette, F., Loi, S., Wang, Y., Lallemand, F., Haibe-Kains, B., Viale, G., Delorenzi, M., Zhang, Y., d'Assignies, M.S., et al.; TRANSBIG Consortium (2007). Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. Clin. Cancer Res. *13*, 3207–3214.

Guenther, M.G., Lawton, L.N., Rozovskaia, T., Frampton, G.M., Levine, S.S., Volkert, T.L., Croce, C.M., Nakamura, T., Canaani, E., and Young, R.A. (2008). Aberrant chromatin at genes encoding stem cell regulators in human mixed-lineage leukemia. Genes Dev. *22*, 3403–3408.

Hou, J., Aerts, J., den Hamer, B., van Ijcken, W., den Bakker, M., Riegman, P., van der Leest, C., van der Spek, P., Foekens, J.A., Hoogsteden, H.C., et al. (2010). Gene expression-based classification of non-small cell lung carcinomas and survival prediction. PLoS ONE *5*, e10312.

Hu, R., Dawood, S., Holmes, M.D., Collins, L.C., Schnitt, S.J., Cole, K., Marotti, J.D., Hankinson, S.E., Colditz, G.A., and Tamimi, R.M. (2011). Androgen receptor expression and breast cancer survival in postmenopausal women. Clin. Cancer Res. *17*, 1867–1874.

Ince, T.A., Richardson, A.L., Bell, G.W., Saitoh, M., Godar, S., Karnoub, A.E., Iglehart, J.D., and Weinberg, R.A. (2007). Transformation of different human breast epithelial cell types leads to distinct tumor phenotypes. Cancer Cell *12*, 160–170.

Jorissen, R.N., Gibbs, P., Christie, M., Prakash, S., Lipton, L., Desai, J., Kerr, D., Aaltonen, L.A., Arango, D., Kruhøffer, M., et al. (2009). Metastasis-Associated Gene Expression Changes Predict Poor Outcomes in Patients with Dukes Stage B and C Colorectal Cancer. Clin. Cancer Res. *15*, 7642–7651.

Lee, T.I., Johnstone, S.E., and Young, R.A. (2006). Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat. Protoc. *1*, 729–748.

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

Loi, S., Haibe-Kains, B., Desmedt, C., Wirapati, P., Lallemand, F., Tutt, A.M., Gillet, C., Ellis, P., Ryder, K., Reid, J.F., et al. (2008). Predicting prognosis using molecular profiling in estrogen receptor-positive breast cancer treated with tamoxifen. BMC Genomics *9*, 239.

Minn, A.J., Gupta, G.P., Siegel, P.M., Bos, P.D., Shu, W., Giri, D.D., Viale, A., Olshen, A.B., Gerald, W.L., and Massagué, J. (2005). Genes that mediate breast cancer metastasis to lung. Nature *436*, 518–524.

Novershtern, N., Subramanian, A., Lawton, L.N., Mak, R.H., Haining, W.N., McConkey, M.E., Habib, N., Yosef, N., Chang, C.Y., Shay, T., et al. (2011). Densely interconnected transcriptional circuits control cell states in human hematopoiesis. Cell *144*, 296–309.

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

Rhodes, D.R., Kalyana-Sundaram, S., Mahavisno, V., Varambally, R., Yu, J., Briggs, B.B., Barrette, T.R., Anstet, M.J., Kincead-Beal, C., Kulkarni, P., et al. (2007). Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. Neoplasia *9*, 166–180.

Schmidt, M., Böhm, D., von Törne, C., Steiner, E., Puhl, A., Pilch, H., Lehr, H.A., Hengstler, J.G., Kölbl, H., and Gehrmann, M. (2008). The humoral immune system has a key prognostic impact in node-negative breast cancer. Cancer Res. *68*, 5405–5413.

Smith, J.J., Deane, N.G., Wu, F., Merchant, N.B., Zhang, B., Jiang, A., Lu, P., Johnson, J.C., Schmidt, C., Bailey, C.E., et al. (2010). Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. Gastroenterology *138*, 958–968.

Tamimi, R.M., Baer, H.J., Marotti, J., Galan, M., Galaburda, L., Fu, Y., Deitz, A.C., Connolly, J.L., Schnitt, S.J., Colditz, G.A., and Collins, L.C. (2008). Comparison of molecular phenotypes of ductal carcinoma in situ and invasive breast cancer. Breast Cancer Res. *10*, R67.

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.

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.

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.

Yu, M., Riva, L., Xie, H., Schindler, Y., Moran, T.B., Cheng, Y., Yu, D., Hardison, R., Weiss, M.J., Orkin, S.H., et al. (2009). Insights into GATA-1-mediated gene activation versus repression via genome-wide chromatin occupancy analysis. Mol. Cell *36*, 682–695.

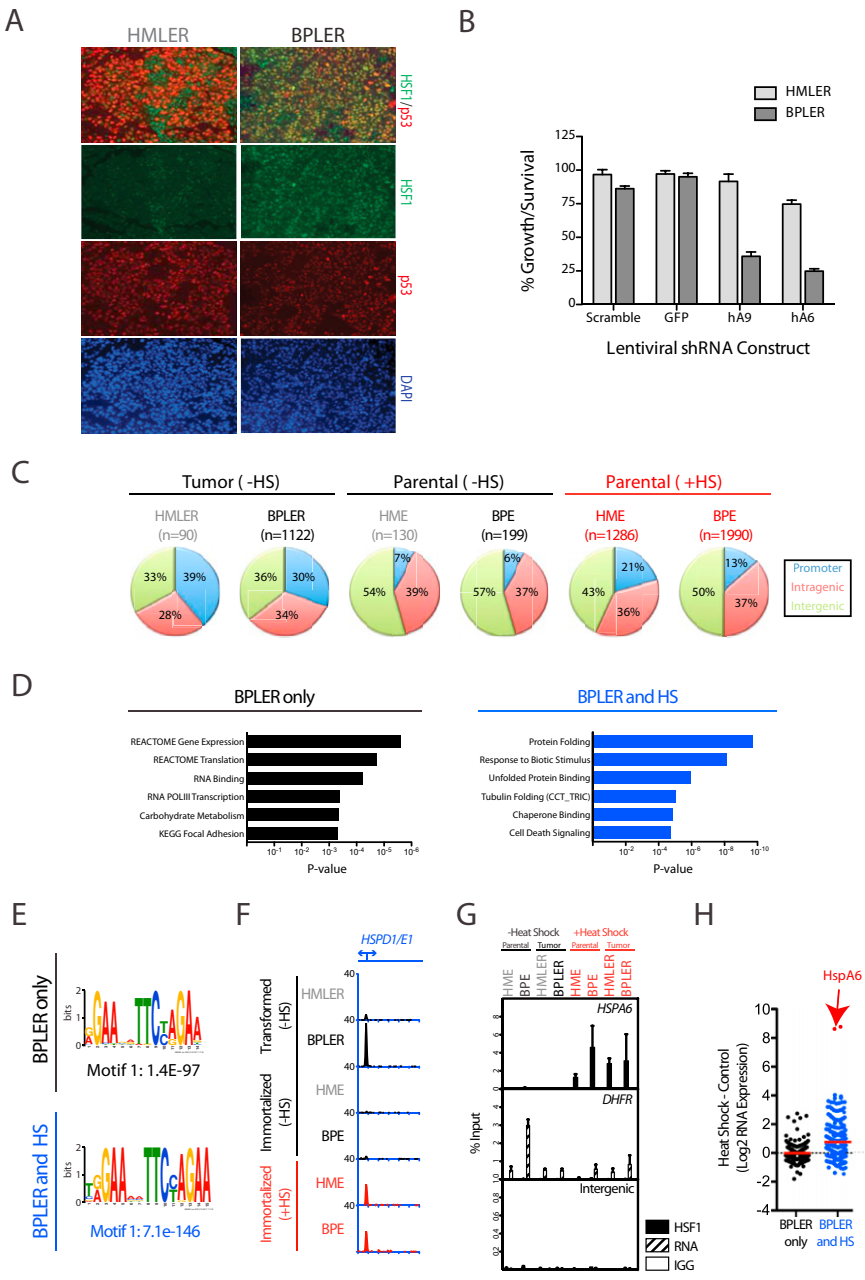

Figure S1. BPLER Cells Are Highly Dependent on HSF1 for Survival and HSF1 Activation during Malignancy Is Distinct from its Activation by Heat Shock

(A) HSF1 (green) and p53 (red) detected by immunofluorescence in HMLER or BPLER xenograft tumors established in mice. Staining for p53 identifies HMLER and BPLER tumor cells. In HMLER cells, HSF1 signal is predominantly seen in p53-low stromal cells.

(B) Cells were plated and transduced with either control lentiviral shRNAi constructs (Scramble or GFP) or lentiviral shRNAi constructs that target HSF1 (hA9, ha6). Four days after transduction, the relative viable cell number was measured by a standard dye reduction assay (Alamar blue).

(C) Genomic distribution of the regions of HSF1 occupancy (promoter, intragenic or intergenic).

(D) Gene set enrichment analysis (GSEA) was performed by using the MSigDB web service (http://www.broadinstitute.org/gsea/index.jsp) on genes bound strongly by HSF1 in cancer only (BPLER, only) or bound strongly by HSF1 in both cancer and heat-shocked cells (BPLER and HS). Complete GSEA results are provided in Table S2.

(E) The sequence motif corresponding to the HSE is strongly enriched within regions bound strongly by HSF1 in BPLER at 37°C (BPLER only, top panel) and genes that were well bound in both BPLER cells at 37°C and in the parental lines (HME and BPE) following heat shock at 42°C (lower panel). The ab initio motif discovery algorithm MEME was used to analyze the 100 bp genomic regions surrounding t HSF1-binding peaks.

(F) HSF1 binding of the HSPD1/E1 locus in HMLER, BPLER, HME and BPE cells at 37°C and HME and BPE cells following heat shock at 42°C. Arrows indicate the transcription start site of each gene. Reads per million total reads are shown.

(G) ChIP was performed from HME, BPE, HMLER or BPLER cells with or without a 1 hr heat shock at 42°C by using the indicated antibodies (RNA: RNA polymerase II, IGG: pre-immune control). Quantitative PCR was performed on enriched DNA with primers for either the promoter of *HSPA6* (top panel), the promoter of *DHFR* (middle panel) or an intergenic region (bottom panel) and normalized to input DNA. For clarity, *HSPA6* enrichment in the RNA Polymerase IP (top panel) is not shown.

(H) mRNA expression analysis showing the effect of heat shock on genes identified as strongly HSF1-bound in BPLER at 37°C (left) and genes identified as bound strongly in both BPLER cells at 37°C and parental HME and BPE cells following heat shock (right). The latter group is more heat-shock responsive than the former group. The two probes corresponding to *HspA6* (HSP70B') are indicated by an arrow. Related to Figure 1.

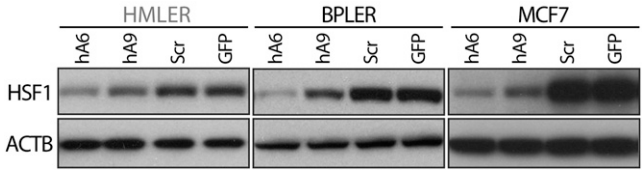

Figure S2. HSF1 Depletion by shRNA in HMLER, BPLER and MCF7 Cells

Equal amounts of total protein isolated from cells following infection with the indicated lentiviral shRNA constructs were subjected to immunoblotting by using an HSF1 antibody (Ab4). ACTB (beta-Actin) was used as a loading control. Related to Figure 2.

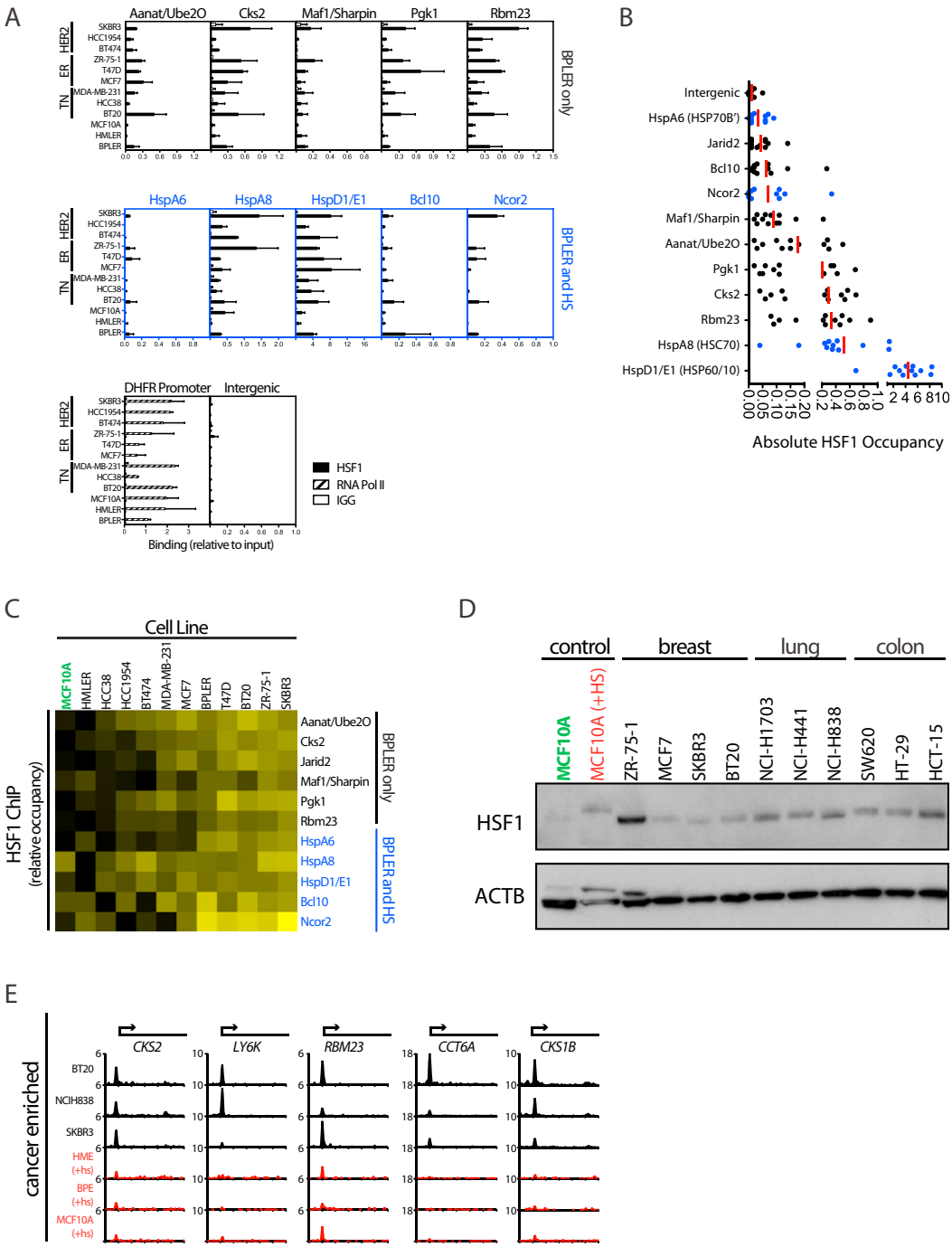

Figure S3. Spectrum of HSF1 Binding across Select Genes in Established Breast Cell Lines

(A) ChIP, with indicated antibody, was performed by using chromatin from the indicated cell lines. Quantitative PCR was performed on enriched DNA with primers corresponding to the indicated genomic regions and normalized to input DNA. Two biological replicates, each of which contained three technical replicates were performed. Data are shown as mean ± standard deviation.

(B) Scatter plot of HSF1 occupancy at the indicated genes in 12 breast cell lines. Genes are ordered by average level of HSF1 binding, from low (intergenic, top) to high (*HspD/E1*, bottom).

(C) Heat map of the HSF1-binding data depicted in (A). Low-level HSF1 binding is indicated in black and higher levels of HSF1 binding are depicted in yellow. Cell lines are ordered by average level of HSF1 occupancy across all genes, from low (MCF10A) to high (SKBR3).

(D) Immunoblot showing HSF1 levels in the cell lines used for the ChIP-Seq experiment presented in Figure 3. Beta-actin (ACTB) was used as a loading control.

(E) HSF1 binding for representative genes (*CKS2*, *LY6K*, *RBM23*, *CCT6A*, and *CKS1B)* is shown. Arrows indicate transcription start site of each gene. Reads per million total reads are shown. Related to Figure 3.

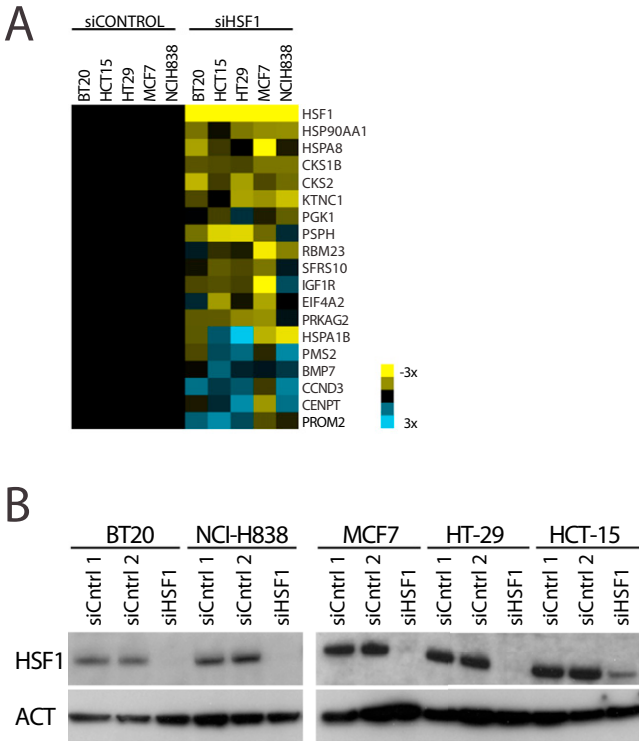

Figure S4. Regulation of HSF1-Target Genes

(A) Quantitative PCR was performed to evaluate expression of selected genes after knockdown of HSF1 by using siRNA oligos (48 hrs posttransfection) in five cells lines (Breast: BT20, MCF7; Colon: HCT15, HT29; Lung NCIH838). Heat map depicts the average fold-change following transfection with two control siRNA (siGLO RISC-Free siRNA and siGENOME Nontargeting siRNA #5) and the fold-change induced by HSF1 knockdown with siGenome SMART pool siRNA-Human HSF1. Yellow: positively regulated; blue: negatively regulated.

(B) Western blot of HSF1 (Ab4 antibody) from cell lysates harvested in parallel with samples used to generate mRNA for the quantitative PCR shown in (A). siCntrl 1: siGLO RISC-Free siRNA; siCntrl 2: siGENOME Nontargeting siRNA #5. siHSF1: siGenome SMART pool siRNA-Human HSF1. ACTB is the loading control. Related to Figure 4.

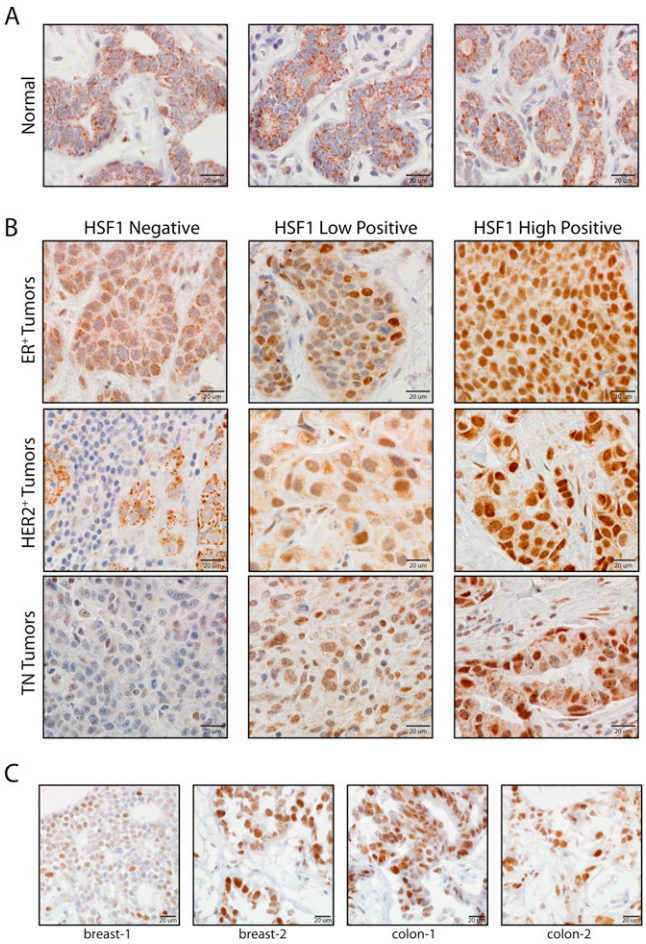

Figure S5. HSF1 Expression by IHC in Additional ER+, HER2+, Triple-Negative Tumors as Well as Patient Breast and Colon Tumors

Immunohistochemical (IHC) staining (brown) with an anti-HSF1 antibody (Ab4) of formalin-fixed paraffin-embedded human surgical specimens of (A) normal mammary tissue or (B) the indicated breast cancer subtypes. Blue staining nuclei with Mayer-hematoxylin counterstain are negative for HSF1. ER+ (estrogen receptor positive); TN (triple negative).

(C) IHC staining of frozen sections of breast and colon tumors used for tumor ChIP-seq analysis in Figure 5D. The level of nuclear HSF1 signal is reported in Figure 5D as HSF1 IHC Grade. Related to Figure 5.

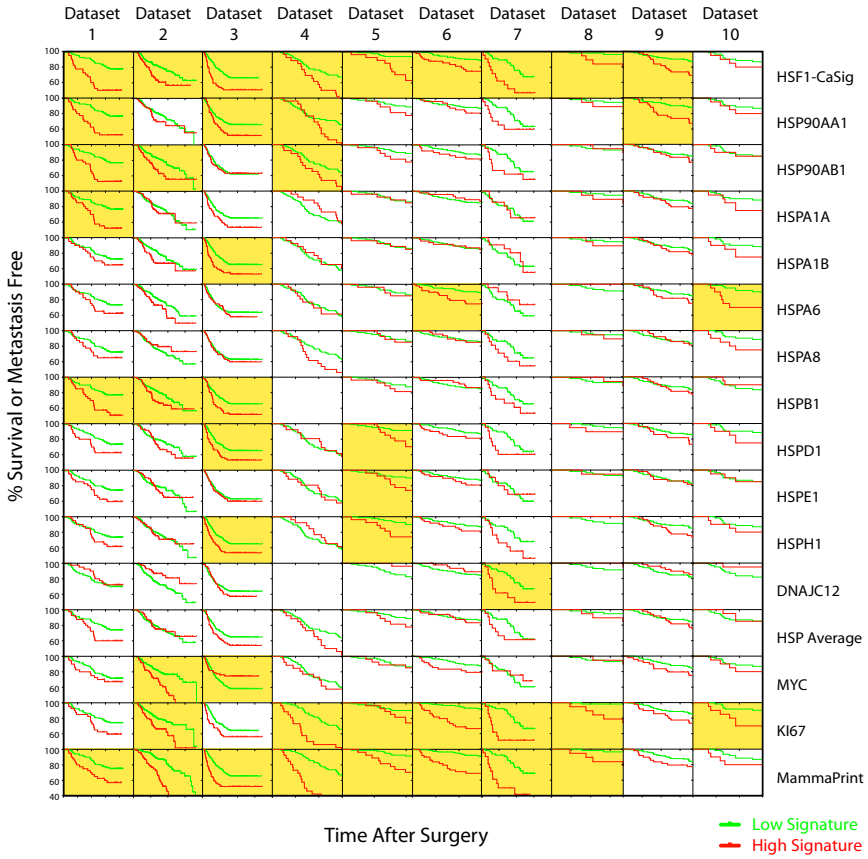

Figure S6. Kaplan-Meier Outcome Curves for Each of the Breast Cancer Data Sets Evaluated in Figure 6B

Meta-analysis of 10 publicly available mRNA expression data sets of breast cancer patients. Kaplan-Meier (KM) analysis of patient outcome using the indicated classifiers is shown. For HSF1 activation, tumors with an average expression value of the HSF1-cancer signature in the top 25th percentile were called "High HSF1-CaSig" (red) and the remaining tumors were called "Low HSF1-CaSig" (green). KM curves highlighted in yellow had log-rank p values < 0.05. Related to Figure 6.

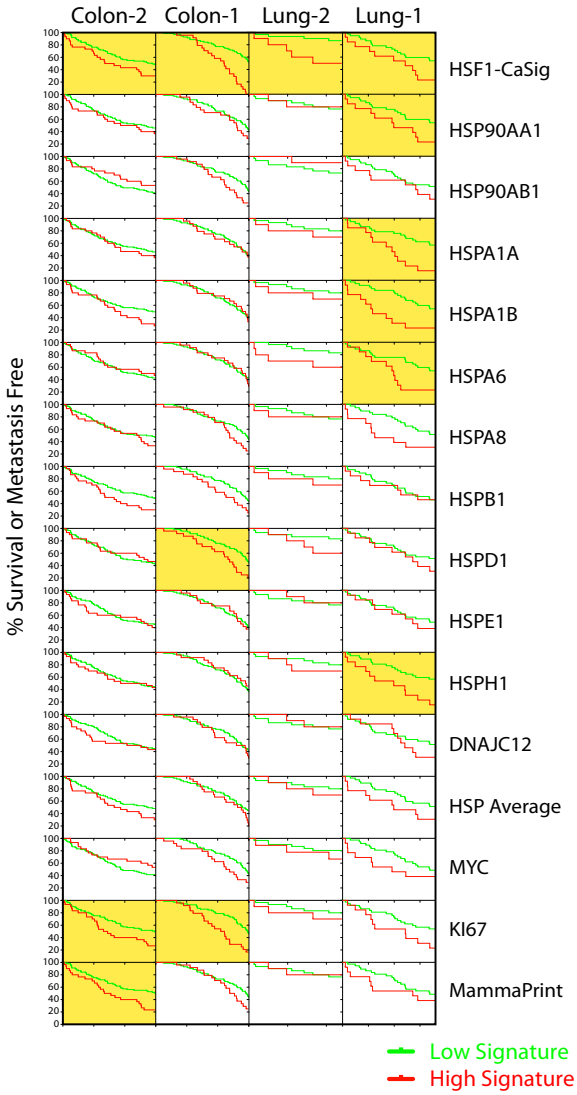

Figure S7. Kaplan-Meier Outcome Curves for Each of the Colon and Lung Cancer Data Sets Evaluated in Figure 7B

Meta-analysis of four publicly available mRNA expression data sets of colon and lung cancer patients. Kaplan-Meier (KM) analysis of patient outcome using the indicated classifiers is shown. For HSF1 activation, tumors with an average expression value of the HSF1-cancer signature in the top 25th percentile were called "High HSF1-CaSig" (red) and the remaining tumors were called "Low HSF1-CaSig" (green). KM curves highlighted in yellow had log-rank p values < 0.05. Related to Figure 7.