Title: Identification of a gene signature to assess HSF1 transcriptional activity

Authors: *Richard L. Carpenter[1,2,3], Haimanti Ray[1], John Wang[1], Sha Cao[3,4].

[1]Medical Sciences, Indiana University School of Medicine, Bloomington, IN, 47405, USA, [2]Department of Biochemistry and Molecular Biology, [3]Melvin and Bren Simon Comprehensive Cancer Center, [4]Department of Biostatistics and Health Data Science, Indiana University School of Medicine, Indianapolis, IN, 46202, USA

*Corresponding Author Contact Information:

Address: 308 Biology Building, 1001 E. 3rd St., Bloomington, IN 47405
Email: richcarp@iu.edu
Phone: (812) 855-8214
Fax: (812) 855-4436

Running Title: HSF1 gene signature

Abstract

Cells have evolved mechanisms for response to external stressors in order to maintain cell viability in the face of changes in the environment. The response to heat stress is driven by the transcription factor heat shock factor 1 (HSF1), which is considered the master regulator of the heat shock response. Here we report a new HSF1 gene signatured termed "HSF1 Activity Signature," or HAS, which is a 23-gene signature that more specifically assesses HSF1 transcriptional activity. We first identified genes that are direct transcriptional targets of HSF1 utilizing more than 40 ChIP-Seq samples in the public domain. Genes identified from ChIP-Seq were then reduced by removing genes whose expression was not altered or showed increased expression with the knockdown of HSF1. These genes were then assessed for their correlation with every other gene across 11 cancer expression datasets to develop an adjacency matrix to identify subgroups of genes that have high inter-gene correlation. The resulting 23-gene set (HAS) was then tested for its performance of assessing HSF1 activity wherein it was found to decreased when HSF1 was knocked down and increase in response to heat shock. The HAS was also able to predict outcomes of breast cancer patients that are previously observed such as high HAS was associated with worse overall survival and metastasis-free survival. This association with cancer patient outcomes extended to many other tumor types. In total, the 23-gene set we termed HAS, is an accurate tool to assess HSF1 activity from transcript expression data that is not limited to cancer-related functions of HSF1.

Introduction

Cells have evolved mechanisms for response to external stressors in order to maintain cell viability in the face of changes in the environment. Cellular exposure to heat stress induces protein denaturation throughout the cell that necessitates a response to promote the proper refolding of proteins to avoid cell death. The response to heat stress is driven by the transcription factor heat shock factor 1 (HSF1), which is considered the master regulator of the heat shock response (1-4). HSF1 and the heat shock response is conserved across a wide range of species (5). HSF1 is expressed across all cell types but the HSF1 protein largely remains inactive within a protein complex located in the cytoplasm in the absence of heat or other stressors. Exposure to heat stress induces a rapid activation process for the HSF1 protein that includes dissociation from the inactivating protein complex, nuclear localization, and trimerization prior to DNA binding and induction of target gene expression. HSF1 has a wide range of target genes, the bulk of which are made up of chaperone proteins that are expressed to assist the refolding of cellular proteins to maintain the integrity of the proteome in the face of heat stress.

Gene signatures, or sets of genes that represent higher-level biological functions or features, have gained popularity since the first introduction of gene set enrichment analysis (GSEA) and an early version of the molecular signatures database (mSigDB) (6). Since its inception, the mSigDB has expanded to include more than 30,000 gene sets that represent biological functions from cancer phenotypes to immune function to metabolic function and many more (6-9). In general, the mSigDB contains signatures with a wide variance in methodologies for producing a gene set to represent a specific biological function. The subcollection TFT (transcription factor targets) previously was a collection of gene sets for many transcription factors that were either gene sets of conserved instances in promoters of mammalian species or genes with regulatory motifs extracted from the TRANSFAC database (now termed TFT legacy) (10). A recent and significant update to the TFT now includes the gene transcription regulation database (GRTD) wherein a more integrated analysis that includes ChIP-Seq and RNA-Seq were utilized to develop gene sets that are more accurate for transcription factor function (7).

While the GRTD does not contain a gene set for HSF1 as of this report, the legacy TFT contains an HSF1 signature derived from genes that contain the HSF1 binding motif within 4 kb around their transcription start sites. However, this gene set does not contain some of the primary gene targets of HSF1, including HSPA1A, perhaps the most often used gene to report HSF1 activity. Consequently, the TFT gene target list is not a reliable set of gene targets for HSF1. There was a report of an HSF1 signature in 2012 within a set of elegant studies that identified cancer-specific gene targets for HSF1 that were distinct from HSF1 targets under heat shock (11). The HSF1-cancer signature (HSF1 CaSig) from this study contained 456 genes bound by HSF1 near their transcription start sites in BPLER (high malignancy) cells (11). The HSF1 CaSig has shown enormous utility across many publications. Here we report a new HSF1 gene signatured termed “HSF1 Activity Signature,” or HAS, which is a 23-gene signature that more specifically assesses HSF1 transcriptional activity. Additionally, we report a novel method of determining a gene set for transcription factors to predict their transcriptional activity incorporating direct gene targets from ChIP-Seq, gene expression, and intra-gene correlation to narrow down the genes to a defined list to accurately represent transcriptional activity.

Results

Identification of an HSF1 Activity Signature (HAS).

HSF1 was originally discovered for its role as the master regulator of the heat shock response (2, 3). There has been a steady rise in publications focusing on HSF1 across many more contexts than the heat shock response in the years since its discovery that indicate HSF1 participates in many other cellular processes. The discovery of new functions of HSF1 has particularly been fruitful in the context of cancer where it has been found to regulate a target gene set that is distinct from the genes targeted by HSF1 in response to heat shock and play a role in many processes that promote malignancy (11, 12). To better determine what cancer cell processes are associated with HSF1 transcriptional activity, we sought to identify a computational approach by which we could associate HSF1 transcriptional activity with cancer-related processes and with outcomes of cancer patients. HSF1 has a complex post-translational activation process whereby it must localize to the nucleus where it undergoes trimerization and DNA binding in addition to phosphorylation prior to recruiting general transcription factors to initiate transcription at target genes. Because of this complex activation process at the protein level, RNA levels of the HSF1 gene are not highly predictive for HSF1 transcriptional activity and do not significantly increase expression in response to heat shock (Suppl Fig. 1A-C). We reasoned that the most predictive gene set for HSF1 transcriptional activity will be RNA expression levels of genes that 1) are direct HSF1 target genes, 2) have high correlation within the gene set across multiple datasets, 3) decrease expression when HSF1 is knocked down or inhibited, 4) increase expression in response to heat shock, and 5) show a trend of increased expression in cancer samples compared to normal. Therefore, we attempted to identify a set of genes that meet these criteria utilizing the gene selection procedures outlined in Figure 1.

We first identified genes that are direct transcriptional targets of HSF1 utilizing more than 40 ChIP-Seq samples in the public domain (11, 13, 14) that include cancer cell lines from multiple cancer types, human tumor samples from multiple cancer types, and non-cancer cell lines under heat stress. All unique genes identified from these ChIP-Seq samples were included in the initial gene list as potential HSF1 target genes (Gene Set 1). To identify target genes from this initial list that are dependent on HSF1 for their expression, we removed genes from the initial list whose expression was not altered or showed increased expression with the knockdown of HSF1 in human breast cancer cells (Gene Set 2). The genes in this reduced gene set were then assessed for their correlation with every other gene in this gene set across 11 different cancer expression datasets to develop an adjacency matrix to identify subgroups of genes that have high inter-gene correlation. The adjacency matrix identified nine different gene sets for consideration with a variable number of genes comprising each gene set ranging from 6 to 34 genes. One gene set, hereafter referred to as the HSF1 Activity Signature (HAS), consisted of 23 genes and was found to have the highest intra-gene correlation (Fig. 2A). Of the 23 genes in the HAS, 17 were identified as HSF1 direct targets in both cancer cells and under heat shock conditions whereas the remaining 6 genes were identified as direct HSF1 targets in cancer cells (Fig. 2B).

HSF1 Activity Signature (HAS) detects changes in HSF1 transcriptional activity.

Following the identification of 9 gene sets from the adjacency matrix, including the 23-gene HAS, we sought to determine the ability of these gene sets to detect changes in HSF1 transcriptional activity in multiple contexts. First, we assessed the expression of the genes within these gene sets in response to the knockdown of HSF1, which results in decreased expression and activity of HSF1. The HAS gene set was found to outperform all other gene sets when HSF1 was knocked down in Hela cells in the presence or absence of heat shock (Fig. 3A, Suppl. Fig. 2A-H). Specifically the HAS gene set had a more uniform change in expression among the genes within the gene set compared to the other gene sets. The HAS gene set also showed consistently decreased expression among the genes within the gene set when HSF1 was knocked down in various cancer cell lines (Fig. 3B-C; Suppl. Fig. 3A). Similarly, the HAS gene set was also similarly decreased in response to the HSF1 inhibitor DTHIB (15) (Fig. 3D). While HSF1 activity should be expected to decrease upon gene knockdown or small molecule inhibition of HSF1, it should also be expected to increase in response to heat shock. Therefore, we next assessed whether the HAS increased in response to heat stress in various cell lines. The HAS showed consistent increases

after heat stress in various cell lines and the genes within the HAS maintained a high correlation (Fig. 3E-F; Suppl. Fig. 3B). The previously published HSF1-CaSig, a 456-gene signature which was identified as a biomarker for cancer patient outcomes across several cancer types, showed an overall decrease in response to HSF1 knockdown (Suppl. Fig. 4A-D) or an increase in response to heat shock (Suppl. Fig. 4D-F) but which less uniformity and has a lower intra-gene correlation compared to HAS (Suppl. Fig. 5). To try and quantifiably compare the HAS across samples within these studies, principal component analysis (PCA) was performed. In these experiments with HSF1 knockdown or heat shock, PC1 of the HAS accounted for 60-85% of the variance across these experiments (Suppl. Fig. 6A-G). To statistically compare HAS across samples, PC1 for each sample was calculated and was used to perform a Welch's t-test wherein the HAS was significantly decreased in samples with HSF1 knockdown and significantly increased in samples with heat shock compared to controls (Suppl. Fig. 6A-G). These data indicate the 23 gene set HAS, can reliably detect changes in HSF1 transcriptional activity as validated against known conditions that will decrease HSF1 activity (gene knockdown) or increase HSF1 activity (heat shock).

HSF1 activity is associated with breast cancer patient outcomes and molecular characteristics.

HSF1 has previously been associated with several cancer phenotypes with clear evidence that HSF1 frequently has increased expression and transcriptional activity in cancer cells (12). The increased expression of HSF1 is at least partially caused by a low percentage of cancer patients having amplification of the HSF1 gene (16, 17). However, detection of the active form of HSF1, either through nuclear HSF1 levels or HSF1 phospho-S326 levels, has been shown to be higher in cancer tissues and associated with worse outcome of cancer patients (11, 13, 18, 19), indicating that not only is the expression of HSF1 relevant to cancer patient outcomes but also HSF1 activity. We assessed whether the HAS is increased in a breast cancer population compared to matched normal adjacent tissue using the TCGA breast cancer cohort. The HAS genes showed a clear increase in expression in tumor samples compared to normal adjacent tissue (Fig. 4A), indicating the HAS was able to detect an increase in HSF1 activity in tumors. Expression of the HSF1 gene itself is also increased in tumor compared to adjacent normal (Suppl. Fig. 7A). We next assessed whether the HAS could serve as a biomarker for outcomes of breast cancer patients as previous studies show active HSF1 is associated with worse outcomes (11, 13, 17-20). First looking at overall survival, higher HAS was associated with worse overall survival in the TCGA breast cancer cohort (Fig. 4B) and the METABRIC breast cancer cohort (Fig. 4C) whereas expression of the HSF1 gene was not associated with overall survival in the TCGA cohort but was in the METABRIC cohort (Suppl. Fig. 7B-C). In addition to Kaplan Meier and Log Rank tests, age-adjusted hazard ratio for HAS was also significantly associated with worse overall survival in the TCGA (HR=2.42, 95% CI: 1.62-3.62; $p<0.001$) and the METABRIC (HR=1.58, 95% CI: 1.33-1.88; $p<0.001$) cohorts. In addition to overall survival, the HAS was also associated with worse metastasis-free survival in two cohorts (Fig. 4D-E) whereas expression of the HSF1 gene was not significantly associated with metastasis-free survival (Suppl. Fig. 7D). Additionally, HAS was seen to be higher in tumors that eventually metastasized compared to tumors that never metastasized (Fig. 4F-G), further supporting a potential role for HSF1 in metastasis as previously suggested (11, 21, 22). The previously identified HSF1-CaSig also performs well as a biomarker as it was designed to do so but has a less uniform increase in tumor compared to adjacent normal (Suppl. Fig. 8A-D). These data indicate the HAS can predict breast cancer patient outcomes consistent with previous findings for HSF1 activity (11, 18, 21).

We next assessed the relationship of the HAS with molecular subtypes of breast cancer. HAS was observed to have concerted upregulation within the Luminal B (LumB), HER2-enriched (HER2-E), and Basal subtypes in both the METABRIC and TCGA cohorts compared to the Normal-Like (NL) and Luminal A (LumA) subtypes, which showed to be statistically significant when comparing PC1 of the HAS across these subtypes (Fig. 5D-E, Suppl. Fig. 9A-B). These results are consistent with previous studies suggesting HSF1 is upregulated in triple-negative breast cancer (TNBC) cells and HER2-expressing breast cancer cells (21-26). However, the HAS was a significant predictor for overall survival in LumB and HER2-enriched subtypes (Suppl. Fig. 9C). The HAS was also assessed in the METABRIC Integrated Clusters (IntClust) where it was identified the HAS is upregulated in several integrated clusters that appear to mirror HAS activation from molecular subtypes. HAS was increased in IntClust 10 that is mostly associated with the basal molecular subtype, IntClust 5 that is closely associated with the HER2-enriched molecular subtype (Fig. 5F-G) (27). Interestingly, HAS was also elevated in clusters 1, 6, and 9 that are

primarily ER-positive cancers that includes overlap with the LumB molecular subtype and likely points towards an interaction between HSF1 and ERα, which a recent report identified HSF1 cooperates with ERα in breast cancer, and an interaction between HSF1 and HER2, which has been established. Further investigating a link between ER status and the HAS it was observed that high HAS significantly reduced overall survival in ER+ patients but not in ER- patients in both the METABRIC and TCGA BRCA cohorts (Fig. 5H-I). The HAS was still observed to significantly predict for overall survival in these cohorts when adjusted for both age and ER status (TCGA BRCA HR: 2.03, 1.30-3.18, p=0.002; METABRIC HR: 1.39, 1.17-1.66, p<0.001). These results suggest HSF1 may be most relevant to the breast cancer subtypes wherein there are molecules expression that functionally affect HSF1, such as Erα and HER2, but HSF1 does also appear to be hyperactivated in the basal subtype so it may still show to play a role in these cancers as well.

HSF1 activity is associated with outcomes of a broad number of cancer types.

After identifying a role for HSF1 activity in breast cancer, we next assessed the relevance of the HAS in other cancer types. The HAS was assessed for its association with both disease-free survival and overall survival across the spectrum of TCGA cancer types wherein the age-adjusted hazard ratio was computed. A forest plot of the HAS hazard ratios indicates an association of HSF1 activity with outcomes of many cancer types. HSF1 has previously been associated with outcomes for many cancer types and the HAS reflected many of these known associations including liver, lung, pancreatic, melanoma, esophageal, cervical, and head/neck cancers (Fig. 6A-B). The known roles of HSF1 in these various cancers illustrates its pleiotropic functions in cancer. HSF1 appears to be critical for supporting MYC and inflammation in hepatocellular carcinoma or it acts downstream of dysregulated EGFR signaling and interferes with AMPK and metabolism in pancreatic cancer. Interestingly, there is a strong association of the HAS with outcomes of kidney papillary cell carcinoma (KIRP) patients (Fig. 6A-B), which has been observed in previous pan-cancer analyses but a clearly defined role for HSF1 in KIRP is yet undefined. One unifying characteristic of the cancers the HAS predicts poor outcomes is they are all epithelial-derived cancers, which is consistent with known functions of HSF1 supporting epithelial-to-mesenchymal transitions (EMT) through regulation of EMT-promoting transcriptional repressors such as Slug or Snail. There also appears to be a group of these cancers that are typically initiated in response to environmental stressors, including SKCM, LUAD, ESCA, and BLCA, that could also suggest the canonical stress response and cell survival function of HSF1 could be playing a role.

Discussion

HSF1 has been known as the master regulator of the heat shock response since the mid-1980s (2, 3). It was first observed to be altered in cancer in metastatic prostate cancer and has since been found to play a pleiotropic role in cancer where HSF1 can regulate many functions in cancer cells from metabolism to proliferation. Because of its many roles in cancer cells and across many cancer types, there is significant interest in studying HSF1, its activity, and its functions in cancer cells and identifying whether HSF1 activity has an impact on patient outcomes. Due to the complex activation of HSF1 protein activity, the transcript levels of HSF1 has poor utility in predicting or assessing HSF1 transcriptional activity. To this end, we developed here a gene signature, named HAS, that is comprised of 23 genes that are direct HSF1 gene targets and depend on HSF1 for their expression. Additionally, they have a high intra-gene correlation amongst these 23 genes such that increases in HSF1 activity largely result in the majority of the 23 genes to increase in expression. This gives an accurate and sensitive readout of HSF1 activity using transcript data, which was confirmed with studies wherein HSF1 gene was knocked down and wherein cells were exposed to heat shock, both of which have predictable effects on HSF1 activity. There is one previously reported HSF1 gene signature, the CaSig, which was developed from elegant studies identifying many roles for HSF1 in cancer cells (11). While both the CaSig and HAS predict similar outcomes for cancer patients, the HAS does outperform the CaSig in measuring HSF1 transcriptional activity as measured by HSF1 knockdown or heat shock. The improved performance of the HAS in this instance is likely due to the intra-gene correlation as the CaSig (456 genes) had low intra-gene correlation. The application of the HAS will allow for future analysis of HSF1 transcriptional activity with cancer cell functions, patient outcomes, therapeutic response, and many others.

Methods and Materials

RNA-Sequencing: MDA-MB-231 cells expressing control or HSF1-directed shRNA were subjected to lysis and total RNA collection using the PureLink RNA extraction kit with DNAse treatment (ThermoFisher). Total RNA was then subjected to mRNA-sequencing using an Illumina HiSeq 4000. Raw reads were processed to RPKM as we have done previously (28).

Datasets: MCF7 and BPLER cells with HSF1 knockdown or heat shock were accessed from GSE38232. Hela cells with HSF1 knockdown and heat stress were accessed from GSE3697. A549 cells with heat shock were accessed from GSE83844. U2OS cells with heat shock were accessed from GSE115973. The breast cancer TCGA dataset was accessed through the TCGA data portal hosted by the Broad Institute. The METABRIC dataset was accessed through the European Genome-Phenome Archive (EGA) under study ID EGAS00000000083. The first breast cancer dataset was generated from four breast cancer cohorts (GSE2034, GSE2603, GSE5327, GSE12276, and GSE14020) as we did previously (29). The second breast cancer dataset was accessed from GSE47561.

Heat Map Generation and Principal Component Analysis: Expression from relevant samples were converted to z-scores for each gene within the respective signature gene lists. Z-scores were then used to generate heat maps using Morpheus and genes were subjected to hierarchical clustering using one minus Pearson correlation. Gene expression values for the respective genes within each signature were subjected to principal component analysis. PC1 scores were utilized to compared gene signatures within datasets using t-tests or one-way ANOVA where appropriate.

Survival Assessment: Patient survival was assessed using Kaplan-Meier plots within Prism 9. Log-Rank test were used for determination of statistical differences between groups. Cox proportional hazard ratios were computed using PC1 scores from gene signatures with outcomes of overall survival, disease-free survival, or metastasis-free survival using SPSS 28.0 and calculated 95% confidence interval and computed p-values. Forest plots of hazard ratios were plotted with Prism 9.

Acknowledgments

We would like to acknowledge the funding source of National Cancer Institute (NCI-K22CA207575; RLC).

References

1. Ritossa, F. (1962) A new puffing pattern induced by temperature shock and DNP in Drosophila. *Experientia* **18**, 571-573
2. Parker, C. S., and Topol, J. (1984) A Drosophila RNA polymerase II transcription factor contains a promoter-region-specific DNA-binding activity. *Cell* **36**, 357-369
3. Topol, J., Ruden, D. M., and Parker, C. S. (1985) Sequences required for in vitro transcriptional activation of a Drosophila hsp 70 gene. *Cell* **42**, 527-537
4. Sorger, P. K., Lewis, M. J., and Pelham, H. R. (1987) Heat shock factor is regulated differently in yeast and HeLa cells. *Nature* **329**, 81-84
5. Gomez-Pastor, R., Burchfiel, E. T., and Thiele, D. J. (2018) Regulation of heat shock transcription factors and their roles in physiology and disease. *Nature reviews. Molecular cell biology* **19**, 4-19
6. Mootha, V. K., Lindgren, C. M., Eriksson, K. F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., Houstis, N., Daly, M. J., Patterson, N., Mesirov, J. P., Golub, T. R., Tamayo, P., Spiegelman, B., Lander, E. S., Hirschhorn, J. N., Altshuler, D., and Groop, L. C. (2003) PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**, 267-273
7. Kolmykov, S., Yevshin, I., Kulyashov, M., Sharipov, R., Kondrakhin, Y., Makeev, V. J., Kulakovskiy, I. V., Kel, A., and Kolpakov, F. (2021) GTRD: an integrated view of transcription regulation. *Nucleic acids research* **49**, D104-d111
8. Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015) The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell systems* **1**, 417-425
9. Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdottir, H., Tamayo, P., and Mesirov, J. P. (2011) Molecular signatures database (MSigDB) 3.0. *Bioinformatics (Oxford, England)* **27**, 1739-1740
10. Xie, X., Lu, J., Kulbokas, E. J., Golub, T. R., Mootha, V., Lindblad-Toh, K., Lander, E. S., and Kellis, M. (2005) Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* **434**, 338-345
11. Mendillo, M. L., Santagata, S., Koeva, M., Bell, G. W., Hu, R., Tamimi, R. M., Fraenkel, E., Ince, T. A., Whitesell, L., and Lindquist, S. (2012) HSF1 drives a transcriptional program distinct from heat shock to support highly malignant human cancers. *Cell* **150**, 549-562
12. Carpenter, R. L., and Gokmen-Polar, Y. (2019) HSF1 as a Cancer Biomarker and Therapeutic Target. *Current cancer drug targets* **19**, 515-524
13. Kourtis, N., Moubarak, R. S., Aranda-Orgilles, B., Lui, K., Aydin, I. T., Trimarchi, T., Darvishian, F., Salvaggio, C., Zhong, J., Bhatt, K., Chen, E. I., Celebi, J. T., Lazaris, C., Tsirigos, A., Osman, I., Hernando, E., and Aifantis, I. (2015) FBXW7 modulates cellular stress response and metastatic potential through HSF1 post-translational modification. *Nature cell biology* **17**, 322-332
14. Santagata, S., Mendillo, M. L., Tang, Y. C., Subramanian, A., Perley, C. C., Roche, S. P., Wong, B., Narayan, R., Kwon, H., Koeva, M., Amon, A., Golub, T. R., Porco, J. A., Jr., Whitesell, L., and Lindquist, S. (2013) Tight coordination of protein translation and HSF1 activation supports the anabolic malignant state. *Science (New York, N.Y.)* **341**, 1238303
15. Dong, B., Jaeger, A. M., Hughes, P. F., Loiselle, D. R., Hauck, J. S., Fu, Y., Haystead, T. A., Huang, J., and Thiele, D. J. (2020) Targeting therapy-resistant prostate cancer via a direct inhibitor of the human heat shock transcription factor 1. *Sci Transl Med* **12**
16. Chen, F., Fan, Y., Cao, P., Liu, B., Hou, J., Zhang, B., and Tan, K. (2021) Pan-Cancer Analysis of the Prognostic and Immunological Role of HSF1: A Potential Target for Survival and Immunotherapy. *Oxid Med Cell Longev* **2021**, 5551036

17. Gökmen-Polar, Y., and Badve, S. (2016) Upregulation of HSF1 in estrogen receptor positive breast cancer. *Oncotarget* **7**, 84239-84245
18. 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. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 18378-18383
19. Yasuda, K., Hirohashi, Y., Mariya, T., Murai, A., Tabuchi, Y., Kuroda, T., Kusumoto, H., Takaya, A., Yamamoto, E., Kubo, T., Nakatsugawa, M., Kanaseki, T., Tsukahara, T., Tamura, Y., Hirano, H., Hasegawa, T., Saito, T., Sato, N., and Torigoe, T. (2017) Phosphorylation of HSF1 at serine 326 residue is related to the maintenance of gynecologic cancer stem cells through expression of HSP27. *Oncotarget* **8**, 31540-31553
20. Tsukao, Y., Yamasaki, M., Miyazaki, Y., Makino, T., Takahashi, T., Kurokawa, Y., Miyata, H., Nakajima, K., Takiguchi, S., Mimori, K., Mori, M., and Doki, Y. (2017) Overexpression of heat-shock factor 1 is associated with a poor prognosis in esophageal squamous cell carcinoma. *Oncology letters* **13**, 1819-1825
21. Carpenter, R. L., Sirkisoon, S., Zhu, D., Rimkus, T., Harrison, A., Anderson, A., Paw, I., Qasem, S., Xing, F., Liu, Y., Chan, M., Metheny-Barlow, L., Pasche, B. C., Debinski, W., Watabe, K., and Lo, H. W. (2017) Combined inhibition of AKT and HSF1 suppresses breast cancer stem cells and tumor growth. *Oncotarget* **8**, 73947-73963
22. Xi, C., Hu, Y., Buckhaults, P., Moskophidis, D., and Mivechi, N. F. (2012) Heat shock factor Hsf1 cooperates with ErbB2 (Her2/Neu) protein to promote mammary tumorigenesis and metastasis. *The Journal of biological chemistry* **287**, 35646-35657
23. Carpenter, R. L., Paw, I., Dewhirst, M. W., and Lo, H. W. (2015) Akt phosphorylates and activates HSF-1 independent of heat shock, leading to Slug overexpression and epithelial-mesenchymal transition (EMT) of HER2-overexpressing breast cancer cells. *Oncogene* **34**, 546-557
24. 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
25. Moreno, R., Banerjee, S., Jackson, A. W., Quinn, J., Baillie, G., Dixon, J. E., Dinkova-Kostova, A. T., Edwards, J., and de la Vega, L. (2021) The stress-responsive kinase DYRK2 activates heat shock factor 1 promoting resistance to proteotoxic stress. *Cell Death Differ* **28**, 1563-1578
26. Schulz, R., Streller, F., Scheel, A. H., Rüschoff, J., Reinert, M. C., Dobbelstein, M., Marchenko, N. D., and Moll, U. M. (2014) HER2/ErbB2 activates HSF1 and thereby controls HSP90 clients including MIF in HER2-overexpressing breast cancer. *Cell Death Dis* **5**, e980
27. Curtis, C., Shah, S. P., Chin, S. F., Turashvili, G., Rueda, O. M., Dunning, M. J., Speed, D., Lynch, A. G., Samarajiwa, S., Yuan, Y., Gräf, S., Ha, G., Haffari, G., Bashashati, A., Russell, R., McKinney, S., Langerød, A., Green, A., Provenzano, E., Wishart, G., Pinder, S., Watson, P., Markowetz, F., Murphy, L., Ellis, I., Purushotham, A., Børresen-Dale, A. L., Brenton, J. D., Tavaré, S., Caldas, C., and Aparicio, S. (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346-352
28. Huang, Z., Johnson, T. S., Han, Z., Helm, B., Cao, S., Zhang, C., Salama, P., Rizkalla, M., Yu, C. Y., Cheng, J., Xiang, S., Zhan, X., Zhang, J., and Huang, K. (2020) Deep learning-based cancer survival prognosis from RNA-seq data: approaches and evaluations. *BMC Med Genomics* **13**, 41
29. Xing, F., Liu, Y., Sharma, S., Wu, K., Chan, M. D., Lo, H. W., Carpenter, R. L., Metheny-Barlow, L. J., Zhou, X., Qasem, S. A., Pasche, B., and Watabe, K. (2016) Activation of the c-Met Pathway Mobilizes an Inflammatory Network in the Brain Microenvironment to Promote Brain Metastasis of Breast Cancer. *Cancer research* **76**, 4970-4980

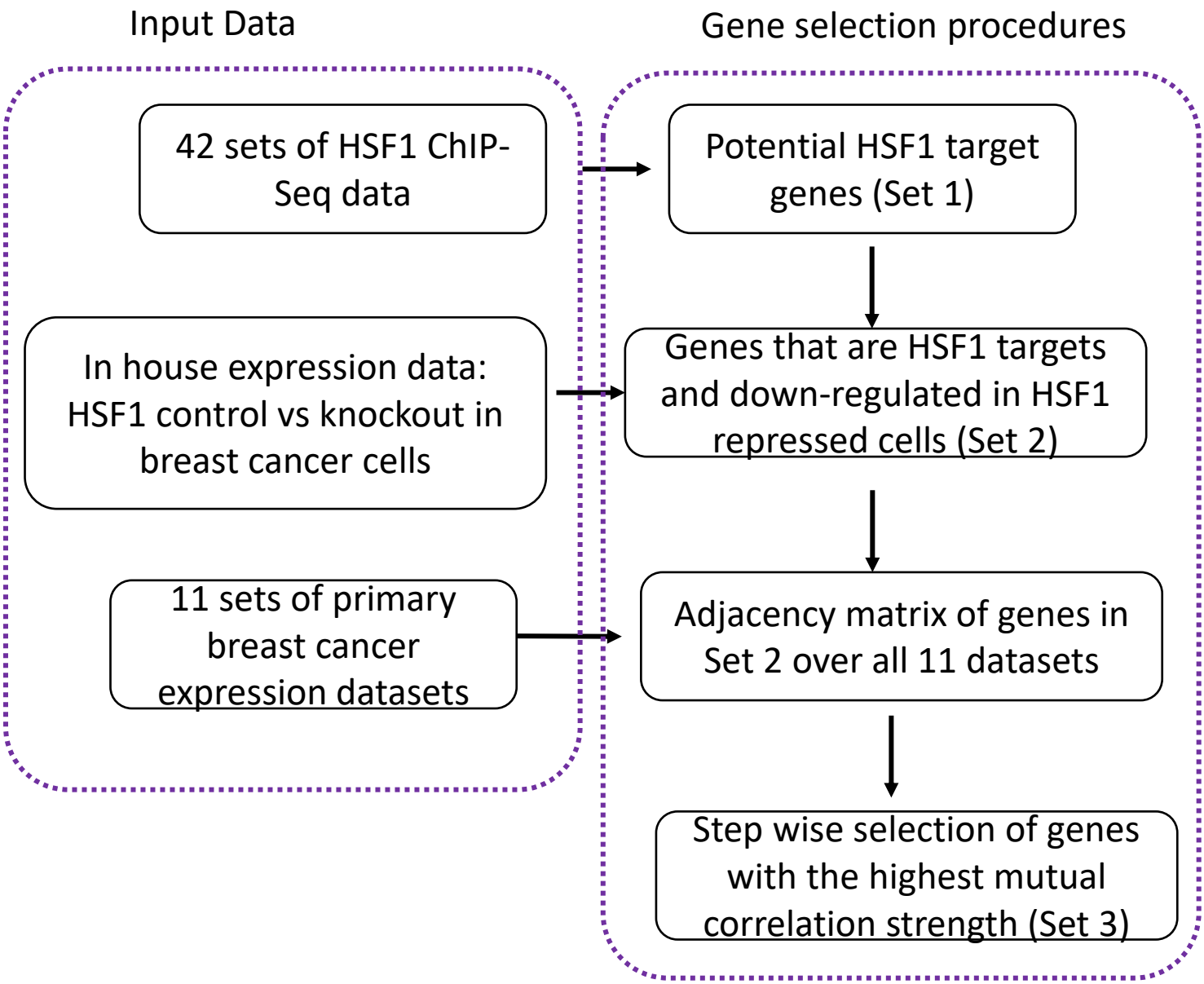


Figure 1: Schematic plot of the selection of HSF1 genes.

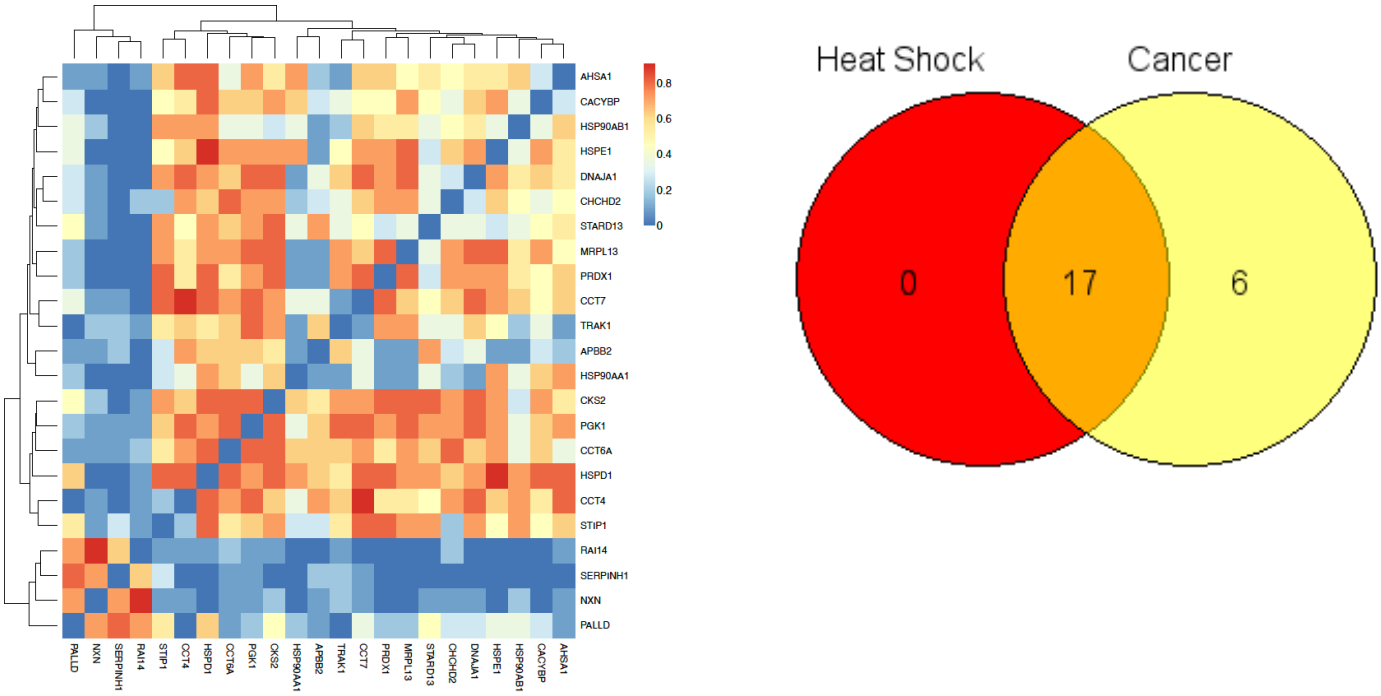


Figure 2: the average correlation strength of our signature genes (Left) and CaSig genes (Right) over 11 breast cancer datasets.

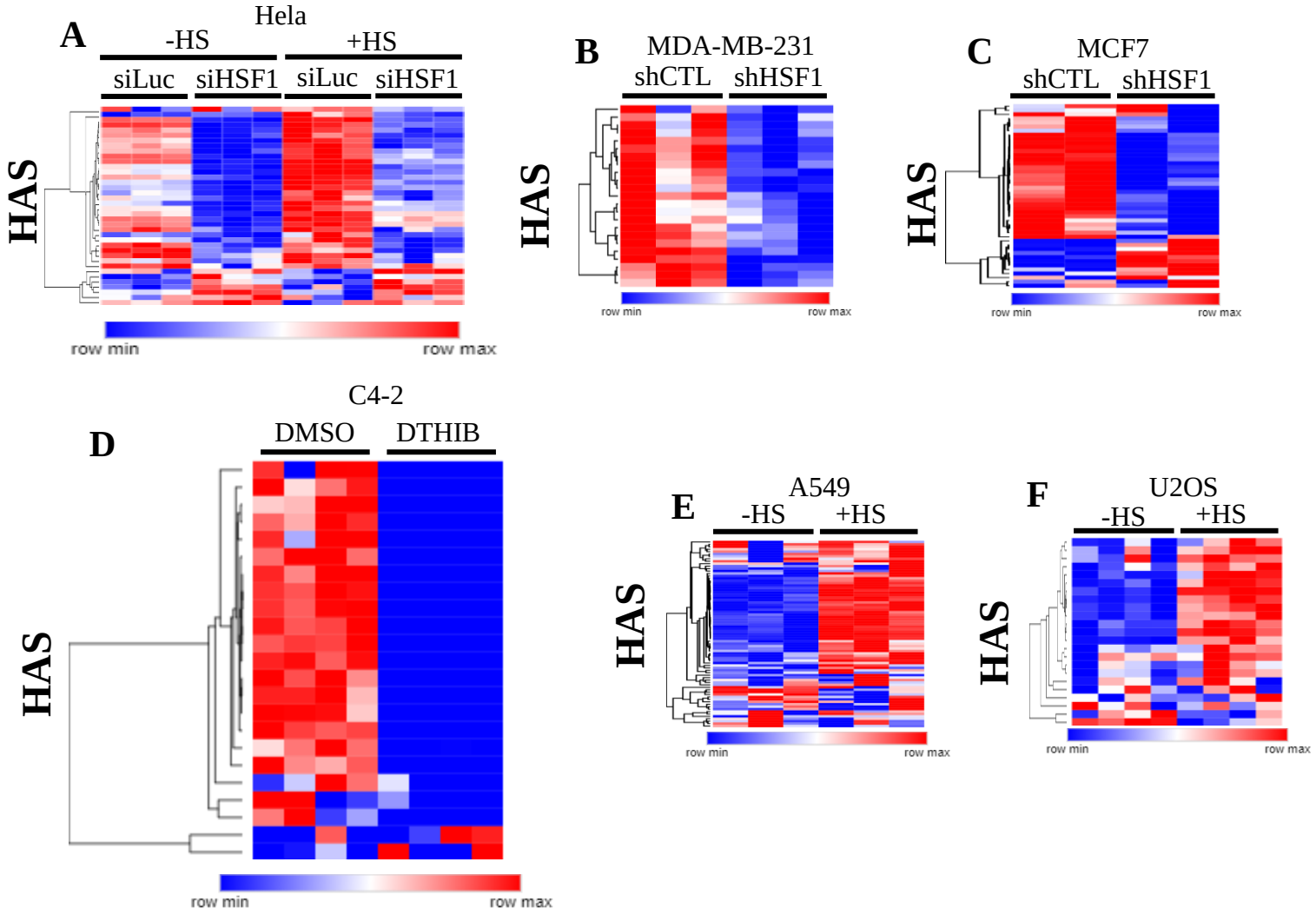


Figure 3: HSF1 Activity Signature accurately assess HSF1 activity in both cancer and non-cancer contexts.

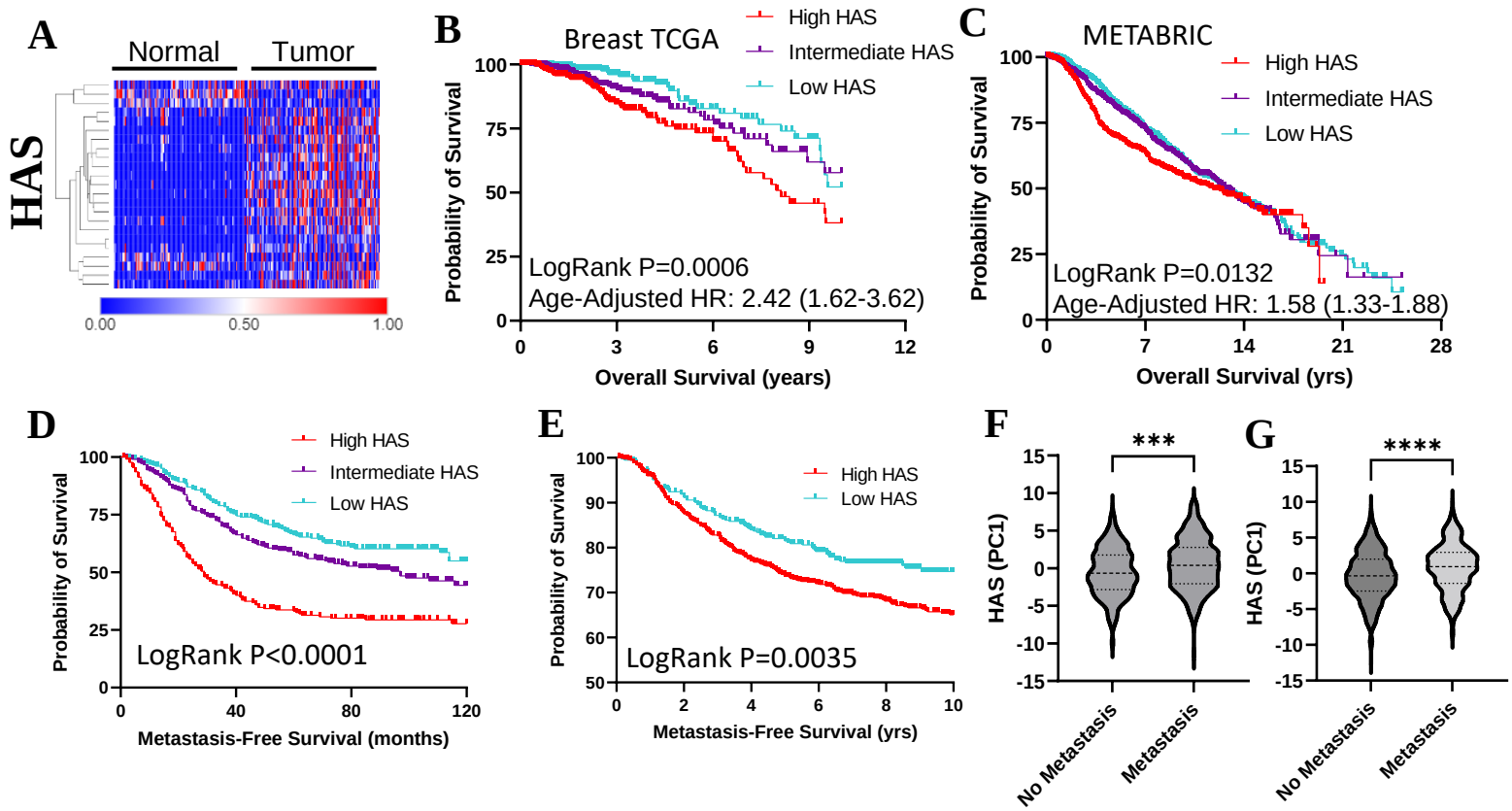


Figure 4: HSF1 Activity Signature is associated with breast cancer patient outcomes.

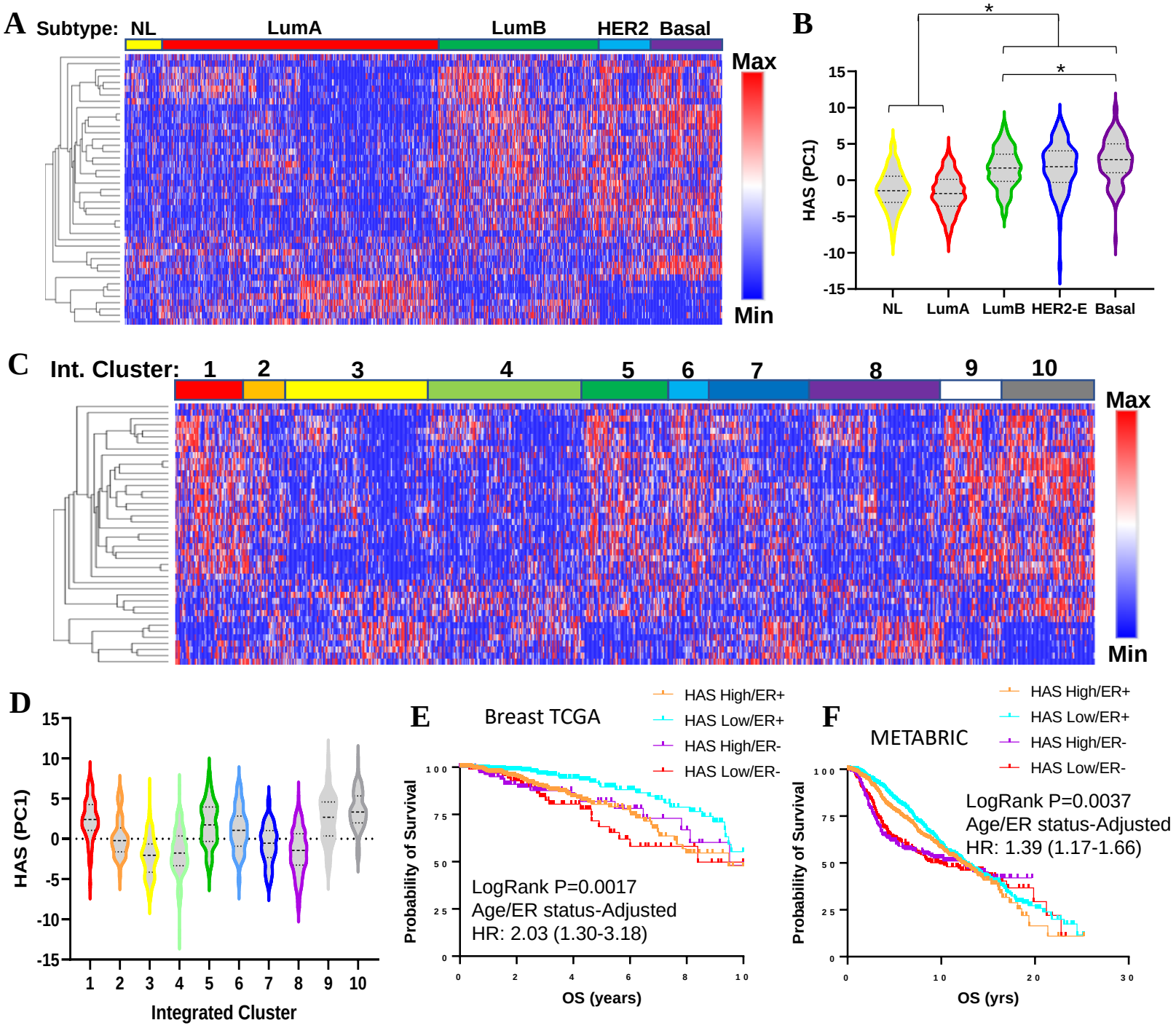
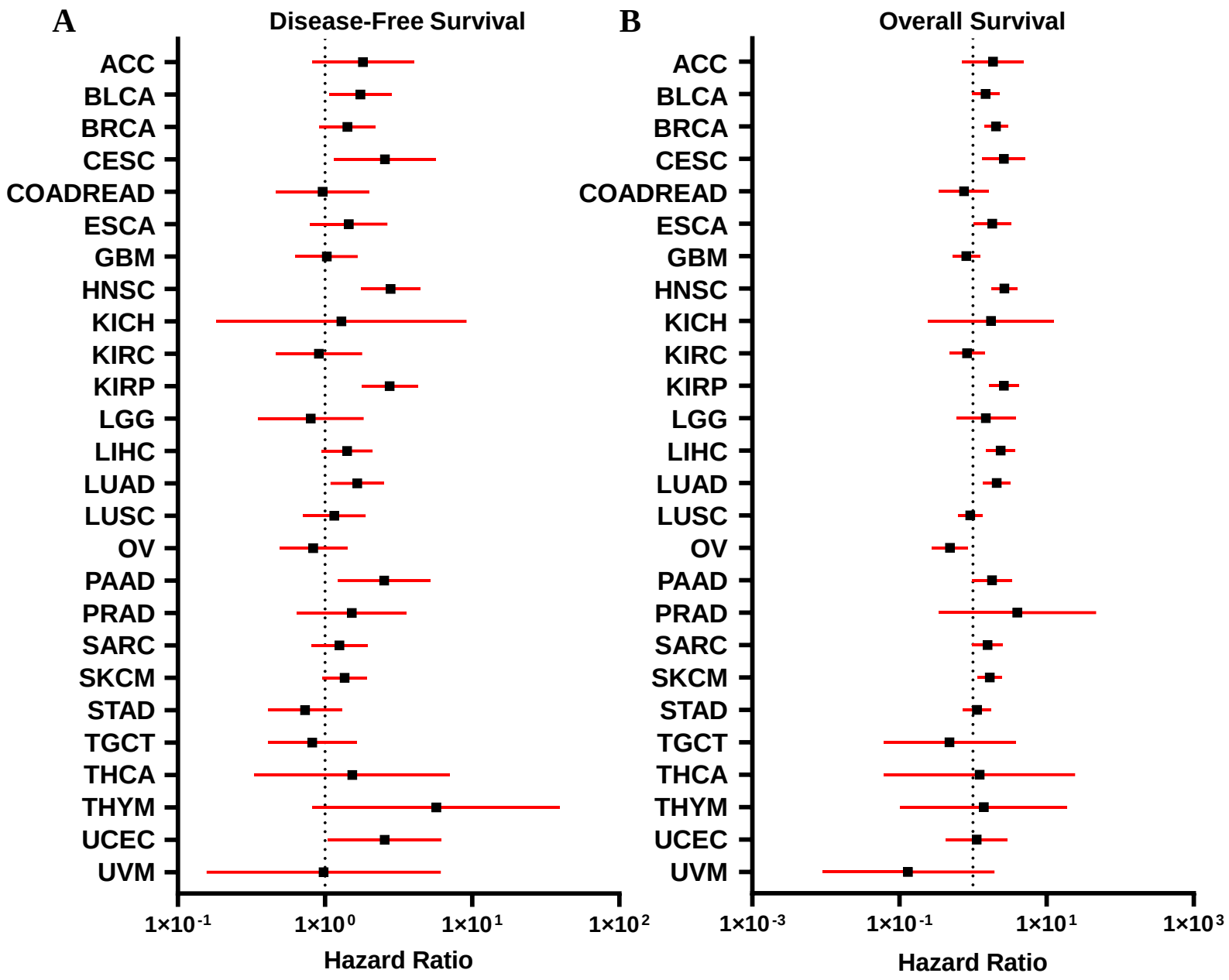
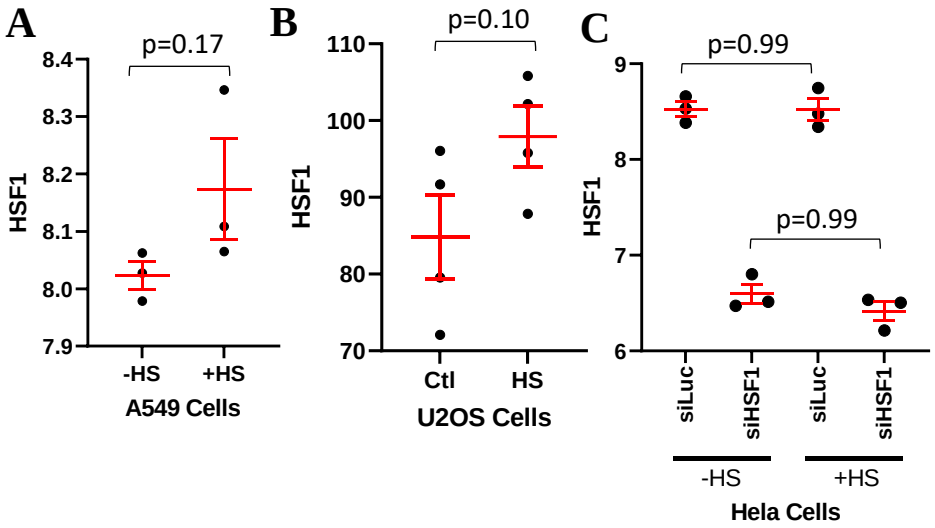
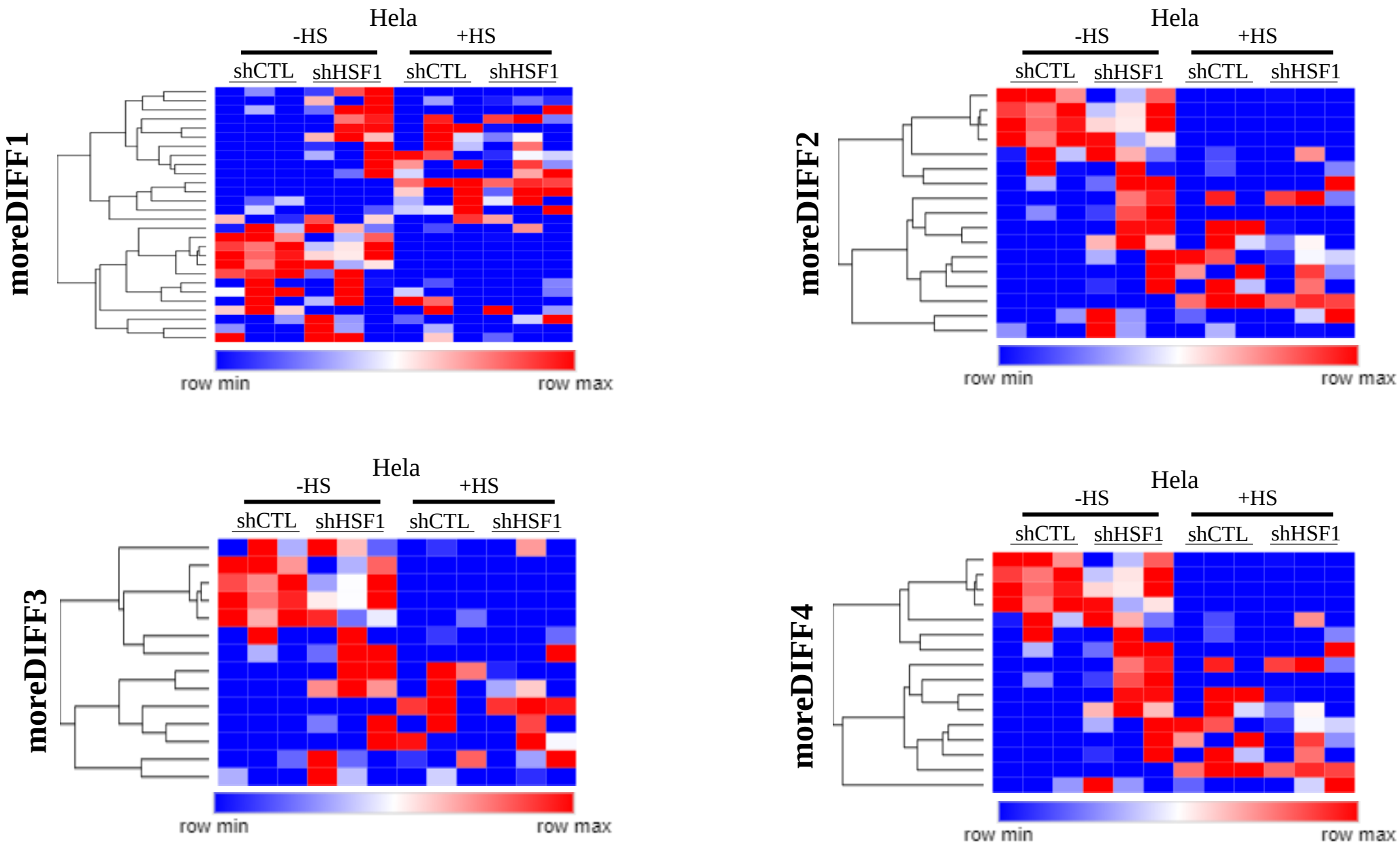


Figure 5: HSF1 Activity Signature is associated with breast cancer molecular subtypes.

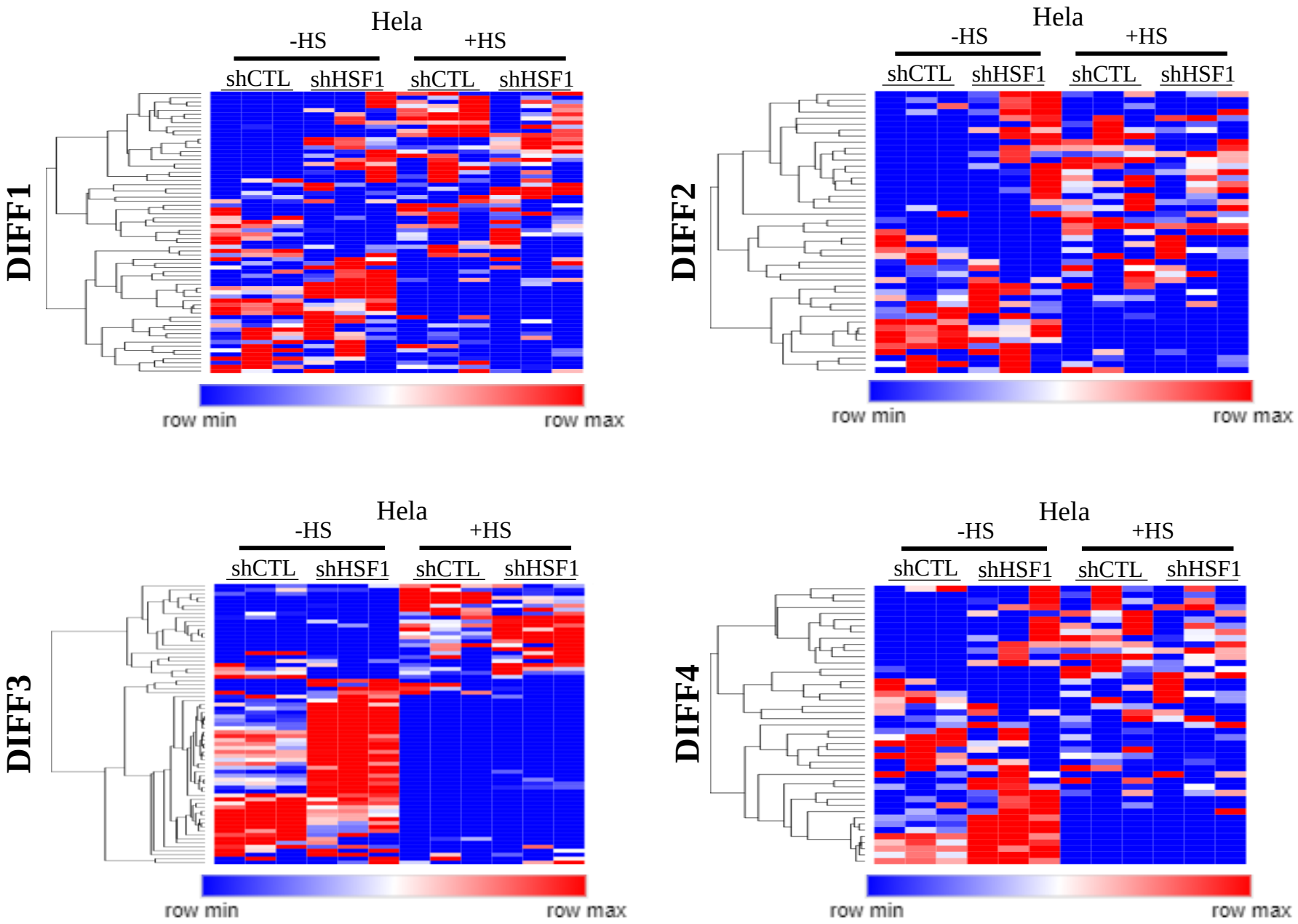

A
Disease-Free Survival
ACC
BLCA
BRCA
CESC
COADREAD
ESCA
GBM
HNSC
KICH
KIRC
KIRP
LGG
LIHC
LUAD
LUSC
OV
PAAD
PRAD
SARC
SKCM
STAD
TGCT
THCA
THYM
UCEC
UVM
1×10-1
1×100
1×101
1×102
Hazard Ratio
B
Overall Survival
ACC
BLCA
BRCA
CESC
COADREAD
ESCA
GBM
HNSC
KICH
KIRC
KIRP
LGG
LIHC
LUAD
LUSC
OV
PAAD
PRAD
SARC
SKCM
STAD
TGCT
THCA
THYM
UCEC
UVM
1×10-3
1×10-1
1×101
1×103
Hazard Ratio



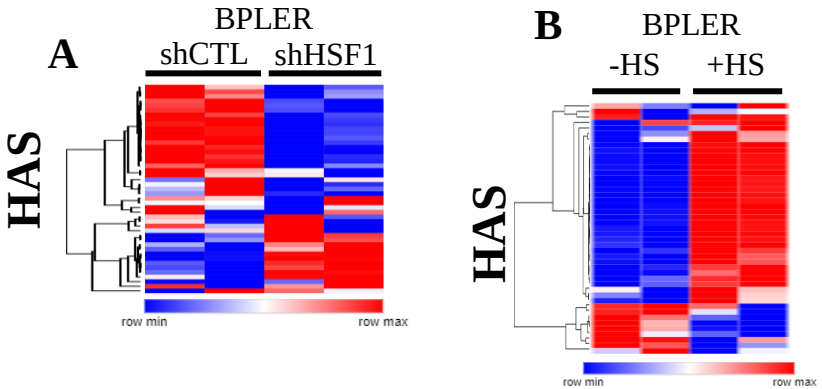
Suppl. Fig. 1 HSF1 expression in response to heat shock



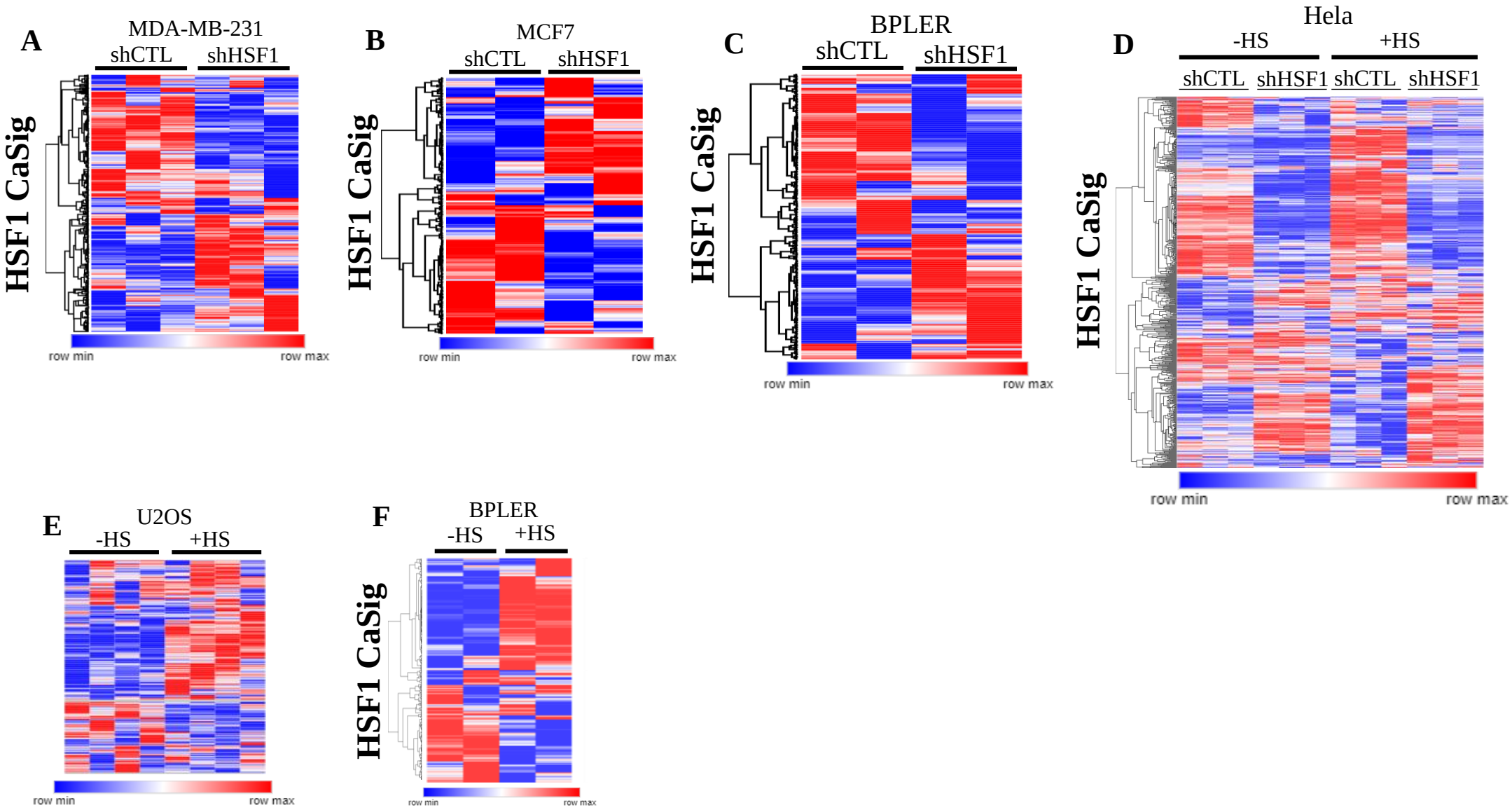
Suppl. Fig. 2 HSF1 Activity Gene Sets in response to HSF1 knockdown and heat shock.



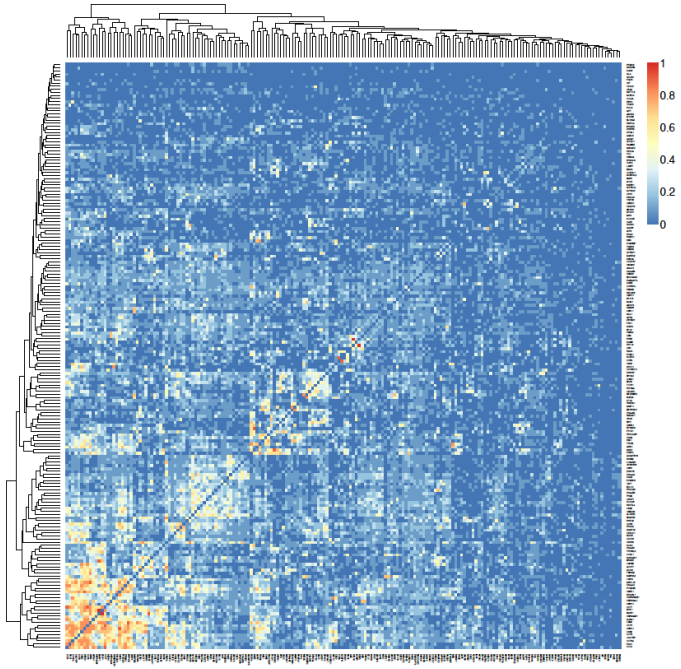
Suppl. Fig. 2 HSF1 Activity Gene Sets in response to HSF1 knockdown and heat shock.



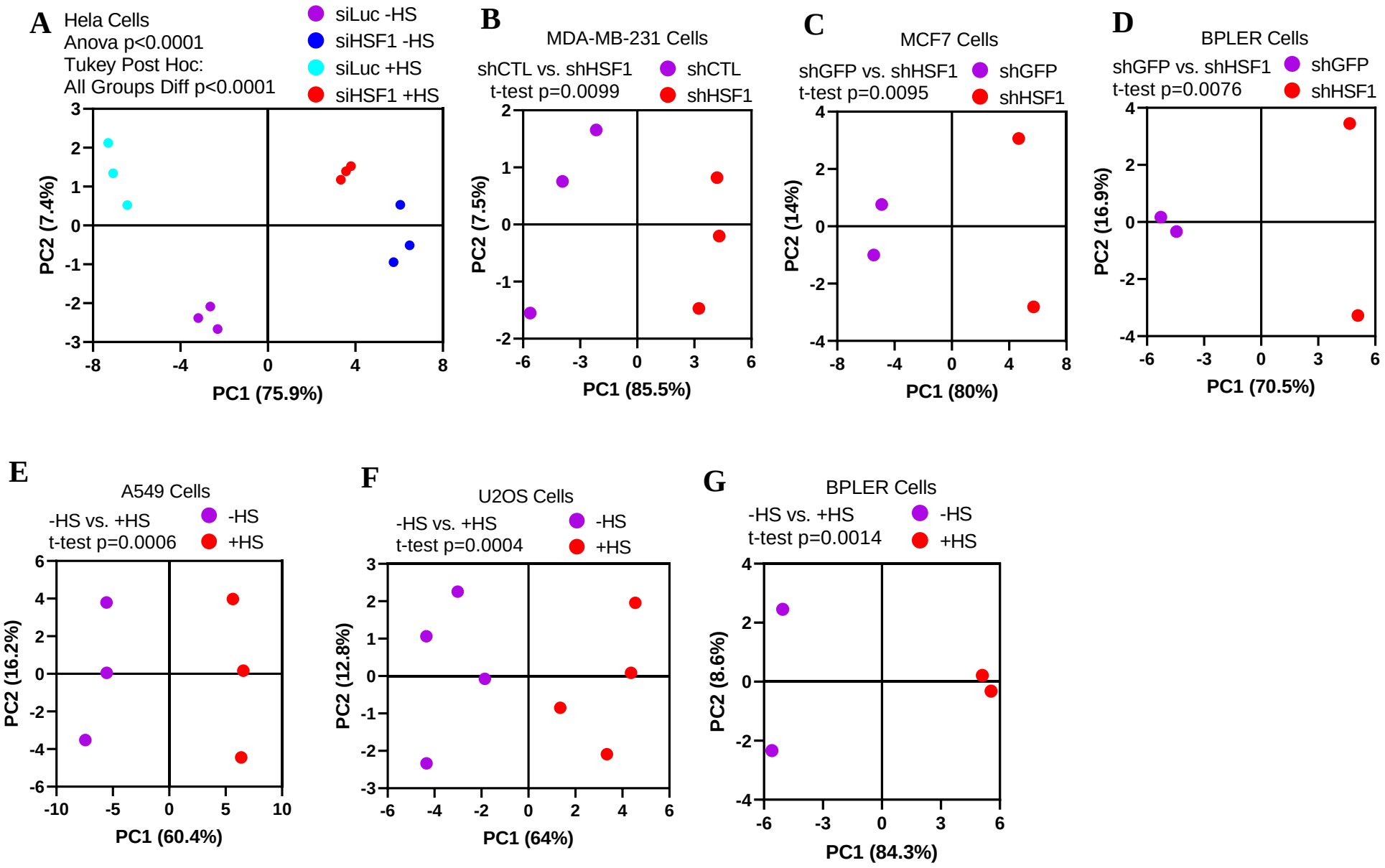
Suppl. Fig. 3 HAS in response to HSF1 depletion and heat shock



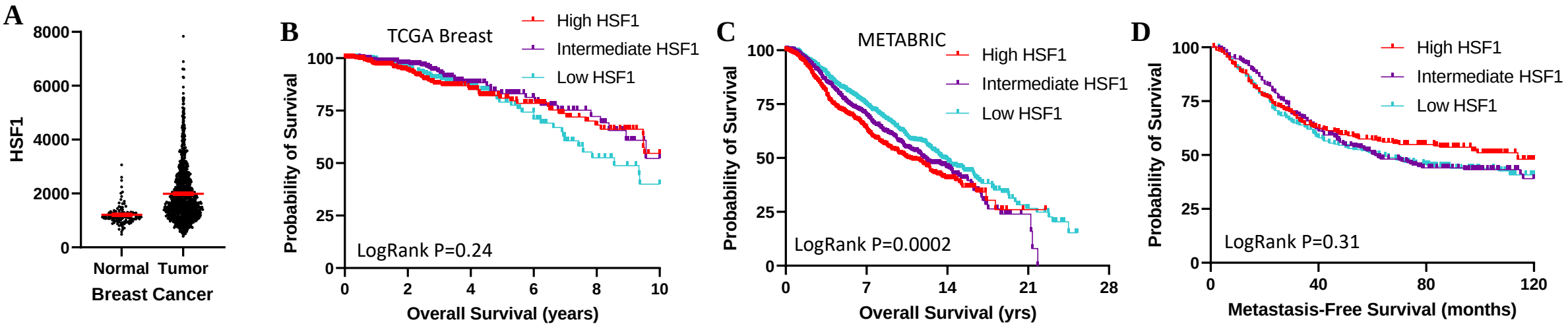
Suppl. Fig. 4 HSF1 CaSig in response to HSF1 depletion and heat shock.



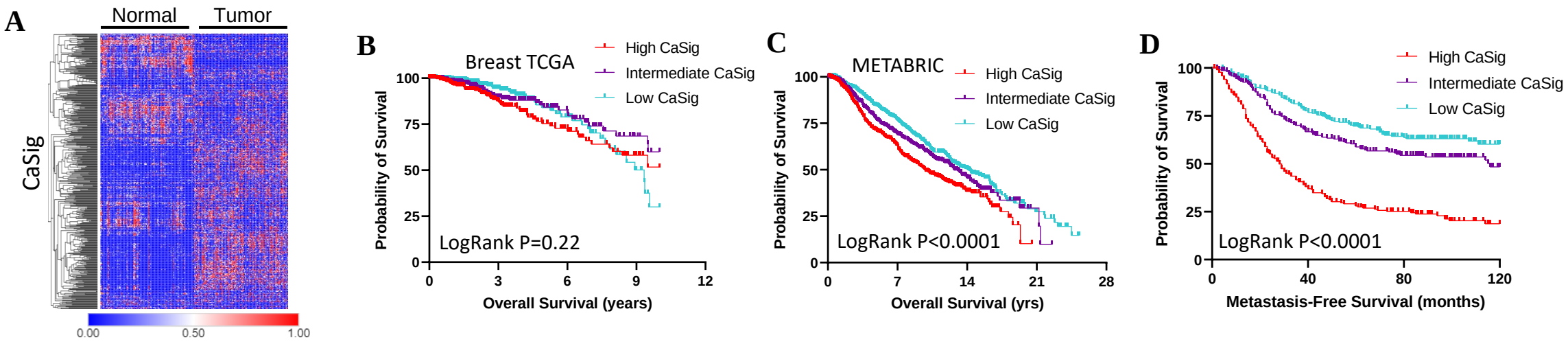
Suppl. Fig. 5 HSF1 CaSig in response to HSF1 depletion and heat shock (cont’d)



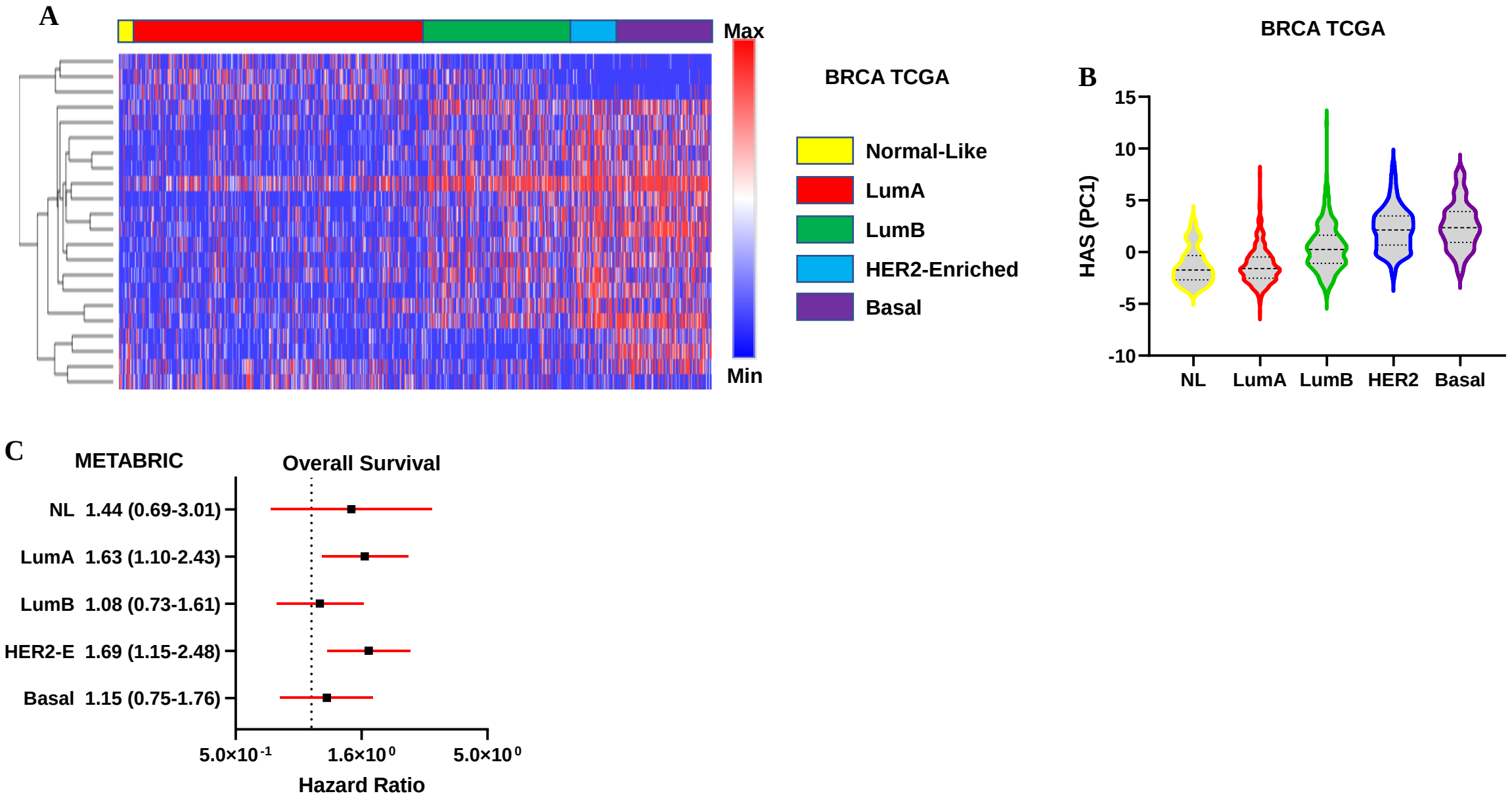
Suppl. Fig. 6 PCA of HAS in response to HSF1 depletion and heat shock



Suppl. Fig. 7 HSF1 Gene expression as a Cancer Biomarker



Suppl. Fig. 8 CaSig Performance as Cancer Biomarker



Suppl. Fig. 9 HAS in BC subtypes