Characterizing the Altered Cellular Proteome Induced by the Stress-Independent Activation of Heat Shock Factor 1

Lisa M. Ryno[1], Joseph C. Genereux[1], Tadasuke Naito[2], Richard I. Morimoto[2], Evan T. Powers[3], Matthew D. Shoulders[1,4], R. Luke Wiseman[1,*]

Page	Contents

SUPPLEMENTAL EXPERIMENTAL PROTOCOLS

Quantitative RT-PCR (qPCR). Transcript expression levels of target genes were determined using qPCR. Cells were treated as described at 37°C, harvested by trypsinization, washed with Dulbecco's phosphate-buffered saline (Gibco), and then RNA was extracted using the RNeasy Mini Kit (Qiagen). qPCR reactions were performed on cDNA prepared from 500 ng of total cellular RNA using the QuantiTect Reverse Transcription Kit (Qiagen). The FastStart Universal SYBR Green Master Mix (Roche), cDNA, and appropriate primers purchased from Integrated DNA Technologies (**Table S3**) were used for amplifications (45 cycles of 2 min at 95°C, 10 sec at 95°C, 30 sec at 60°C) in an ABI 7900HT Fast Real Time PCR machine. Primer integrity was assessed by performing a thermal melt to confirm homogeneity and the absence of primer dimers. All transcripts were normalized to the housekeeping genes *Rplp2* and all measurements were performed in triplicate. Data were analyzed using the RQ Manager and DataAssist 2.0 softwares (ABI). qPCR data are reported as mean ± 95% confidence interval as calculated in DataAssist 2.0.

Sample Preparation, Labeling, and Chip Hybridization for RNAseq Analyses. Total RNA of 3 biological triplicates for each condition (cHSF1 (+/- dox) and FKBP.cHSF1 (+/- Shield-1) were prepared using a Qiagen RNeasy kit supplemented with a DNAse I digestion. Total RNA was quantified using NanoDrop (ND-1000). Samples were run on the Illumina HiSeq system. Read lengths of 50 to 300 bases were generated by hybridizing a library with ligated adapter sequences to the glass surface within the flow cell. The flow cell surface has covalently attached oligonucleotide probes that are complementary to the adapters on the library. Next, a polymerase amplification step was used to amplify the individual hybridized library fragments in spatially distributed "clusters" on the flow cell surface. Fluorescently labeled nucleotides with blocked 3'-OH

were used in a series of polymerase-mediated single base extension steps with the 3' hydroxyl groups de-blocked and fluorescent labels cleaved off after each round of extension. By collecting fluorescent images of the flow cell after each extension reaction, the incorporated nucleotide at each cluster can be identified. After the sequencing run, cluster sequences are determined and quality filtering is applied.

The RNA-Seq data was analyzed using R Bioconductor package DESeq which estimates variance-mean dependence in count data from high-throughput sequencing assays and test for differential expression based on a model using the negative binomial distribution (*1*). Normalization was performed by the TMM (trimmed mean of M) method where the sample whose 75%-ile (of library-scale-scaled counts) is closest to the mean of 75%-iles as the reference. This adjustment is applied to the library size (total sample counts) to account for the compositional bias. In order to estimate the biological or sample-to-sample variability the dispersion parameter is estimated as the overdispersion relative to the Poisson distribution. When a negative binomial model is fitted, the dispersion parameter is calculated by the quantile-adjusted conditional maximum likelihood (qCML) method.

Cluster Analysis of Genes Identified in RNAseq. A matrix of pairwise correlations was generated for the expression patterns of genes that had an abundance > 1.4 RPKM in at least one of the twelve RNAseq datasets generated as described above. This condition was satisfied by 11,691 genes. The sub-matrix corresponding to the pairwise correlation coefficients between a selection of 110 proteostasis genes was then isolated and the genes were clustered hierarchically based on their patterns of correlations using squared Euclidean distance as a distance function and the single linkage algorithm to determine intercluster dissimilarity, as implemented in Mathematica 9.0. These clusters were used to draw the dendrogram in **Figure 3B**.

The network graph in **Figure 3C** was prepared by representing each of the proteostasis genes mentioned above as a vertex and connecting the vertices that have pairwise correlation coefficients > 0.9. This yielded 549 edges (out of a maximum of 5,995 possible), each representing a strong correlation between the expression patterns of a pair of genes. To estimate the number of false positives in this set of edges, we first standardized the data gene-by-gene, and then randomly permuted the data sample-by-sample. This procedure is equivalent to generating a random set of data drawn from the same distributions as the samples in the experimental data set. The pairwise correlation coefficients for the random dataset were calculated and a submatrix of a selection of 110 of the randomly generated "genes" was isolated. This submatrix had only 5 correlation coefficients that were > 0.9, suggesting that the false positive rate in our selection of edges is around 1%. The network graph in **Figure 3E** is a subgraph of **Figure 3C** showing only the vertices that are connected to the vertex that represents DNAJB1.

Sample Preparation, Labeling, and Analysis for TMT-MuDPIT Experiments.

MilliQ-purified water was used for all preparations, and only mass spectrometry grade solvents were employed. Buffer A consisted of 5% acetonitrile and 0.1% formic acid, buffer B consisted of 80% acetonitrile and 0.1% formic acid, and buffer C consisted of 500 mM ammonium acetate. Cells were lysed on ice in RIPA buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1% Triton X100, 0.5% sodium deoxycholate, 0.1% SDS) and centrifuged for 20 min at 10000 x *g*. Protein concentrations of supernatants were determined by BCA (Pierce). For each sample, 100 μg of lysate was washed by chloroform/methanol precipitation. Air-dried pellets were resuspended in 45 mM aqueous triethylammonium bicarbonate (Sigma) with 0.05% SDS. Proteins were reduced with 10 mM Tris(2-carboxyethyl)phosphine hydrochloride (Pierce) for 1 h at 55°C and alkylated with 17 mM iodoacetamide (Sigma) for 30 min at ambient temperature and protected from

light. Proteins were digested for 18 hr at 37°C with 2.5 μg trypsin (Promega) in the presence of 1 mM $CaCl_2$. After digestion, 50 μg peptides from each sample were reacted for 1 h with the appropriate TMT isotopic label (Pierce) in 10 mg/mL anhydrous acetonitrile and quenched with 0.3% hydroxylamine for 15 min. Samples with different TMT labels were pooled and acidified with 5% formic acid. Debris was removed by centrifugation for 30 min at 18000 x *g*. MudPIT microcolumns (*2*) were prepared with Kasil frits in undeactivated 250 mm ID/360 mm OD fused silica capillaries (Agilent). The Kasil frits were prepared by briefly dipping a 20–30 cm capillaries in well-mixed 300 mL Kasil 1624 (PQ Corporation, Malvern, PA) and 100 mL formamide, curing at 100°C for 4 h, and cutting the frits to ~2 mm in length. Strong cation exchange particles (SCX Luna, 5 mm dia., 125 Å pores, Partisphere, GE Healthcare, UK) were packed into the columns under pressure from particle slurries in methanol to 2.5 cm. 2.5 cm reversed phase particles (C18 Aqua, 5 mm dia., 125 Å pores, Phenomenex, Torrance, CA) were then packed into the capillaries using the same method as SCX loading, to generate biphasic columns, onto which the acidified peptides samples were loaded. Analytical reversed phase LC microcolumns were generated by pulling a 100 mm ID/360 mm OD capillary with a P-2000 laser (Sutter Instrument Co., Novato, CA) to a 5 mm ID tip. Reversed phase particles (Aqua C18, 5 mm dia., 125 Å pores, Phenomenex, Torrance, CA) were packed directly into the pulled columns at 800 psi until 15 cm long. MuDPIT microcolumns were connected to the analytical column using a zero-dead volume union (Upchurch Scientific, P-720-01). LCMS/ MS analysis was performed using an Agilent 1200 HPLC pump and a Thermo LTQ-Orbitrap Velos mass spectrometer (Thermo Scientific, San Jose, CA). MuDPIT experiments were performed where each step corresponds to 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100% buffer C being run for 4 min at the beginning of each gradient of buffer B. Electrospray was performed directly from the analytical column by applying the ESI voltage at a tee (150 mm ID, Upchurch

Scientific) directly downstream of a 1:1000 split flow used to reduce the flow rate to 300 nL/min through the columns. Electrospray directly from the LC column was done at 2.5 kV with an inlet capillary temperature of 275°C. Data-dependent acquisition of MS/MS spectra were performed with the following settings: eluted peptides were scanned from 300 to 1600 m/z with resolution 30000 and the mass spectrometer in a data dependent acquisition mode. The top ten peaks for each full scan were fragmented by HCD using a normalized collision energy of 45%, a 100 ms activation time, and a resolution of 7500. Dynamic exclusion parameters were 1 repeat count, 30 ms repeat duration, 500 exclusion list size, 120 s exclusion duration, and exclusion width between 0.51 and 1.51. Protein and peptide identification and protein quantitation were done with the Integrated Proteomics Pipeline - IP2 (Integrated Proteomics Applications, Inc., San Diego, CA. http://www.integratedproteomics.com/). Tandem mass spectra were extracted from raw files using Raw Xtractor 1.9.13 (*3*) and were searched against a database containing 20245 human sequences (longest entry for the IPI database for each protein) with reversed sequences using ProLuCID (*4, 5*). Carbamidomethylation (+57.02146 Da) of cysteine and TMT tagging of N-termini and of lysine residues (+229.1629 Da) were considered as a static modifications. Peptide candidates were filtered using DTASelect2 (version 2.0.27) for a false positive (decoy) peptide ratio of ~ 1% (*3, 6*). Quantitation was performed using Census7 (*7*), followed by deconvolution of isotopic impurity as reported in the lot analysis supplied by Thermo Fisher, and finally normalization of ratio values based on the mode. Redundant peptides were generally assigned to all proteins. For several proteins, increases with HSF1 activation were observed that did not correspond to the RNAseq. In each case, a shared peptide with HSPA6 or HSPA1A was responsible for the entire increase while unique peptides did not change. These proteins were hence not considered to have changed with HSF1 activation.

SUPPLEMENTAL TABLE LEGENDS

Table S1 (Supplement to Figure 2). RNAseq characterization of HEK293T-Rex cells following Shield-1-dependent FKBP.cHSF1 stabilization or dox-dependent cHSF1 overexpression (see Excel Spreadsheet).

Table S2 (Supplement to Figure 3). Table showing the GO annotations enriched in genes induced by Shield-1-dependent FKBP.cHSF1 stabilization (see Excel Spreadsheet).

Table S3 (Supplement to Figure 3). Table showing alterations in mRNA (measured by RNAseq) and protein (measured by TMT-MuDPIT) for 116 primary proteostasis factors induced by Shield-1-dependent FKBP.cHSF1 stabilization or dox-dependent cHSF1 overexpression (see Excel Spreadsheet).

Table S4 (Supplement to Figure 4). TMT-MuDPIT characterization of HEK293T-Rex cells following Shield-1-dependent FKBP.cHSF1 stabilization or dox-dependent cHSF1 overexpression (see Excel Spreadsheet).

Table S5 (Supplement to Extended Experimental Protocols). Compilation of PCR primers used for qPCR.

Name	Forward	Reverse
BAG3	5'-TGGGAGATCAAGATCGACCC-3'	5'-GGGCCATTGGCAGAGGATG-3'
CRYAB	5'-CCTGAGTCCCTTCTACCTTCG-3'	5'- CACATCTCCCAACACCTTAACTT-3'
DNAJB1	5'-TGTGTGGCTGCACAGTGAAC-3'	5'-ACGTTTCTCGGGTGTTTTGG-3'
HSPA1A	5'-GGAGGCGGAGAAGTACA-3'	5'-GCTGATGATGGGGTTAACA-3'
HSP90AA1	5'-GATAAACCCTGACCATTCC-3'	5'-AAGACAGGAGCGCAGTTTCATAAA-3'
HSPE1	5'-ATGGCAGGACAAGCGTTTAGA-3'	5'-CCGATCCAACAGCGACTACT-3'
RPLP2	5′-CGTCGCCTCCTACCTGCT-3′	5′-CATTCAGCTCACTGATAACCTTG-3′

FIGURE LEGENDS

Figure S1 (Supplement to Figure 1) – Shield-1-dependent FKBP.cHSF1 stabilization allows stress-independent activation of HSF1

A. qPCR of *DNAJB1*, and *BAG3* in HEK293T-Rex cells stably expressing tet-inducible cHSF1 or GFP. Cells were treated for 15 h with either dox (2.25 μM) or vehicle, as indicated. The error bars show 95% confidence interval.

B. Representative immunoblot of lysates prepared from HEK293T-Rex cells stably expressing FKBP.YFP, FKBP.cHSF1, tet-inducible GFP, or tet-inducible cHSF1. Cells were treated for 15 h with Shield-1 (1 μM), dox (2.25 μM) or vehicle, as indicated.

C. Quantification of immunoblots as shown in **Figure S1B**. n=3.

D. qPCR of *DNAJB1* in HEK293T-Rex cells stably expressing FKBP.cHSF1 or dox-regulated cHSF1. Cells were treated for 15 h with either Shield-1 (1 μM) or dox (2.25 μM), as indicated. HEK293T-Rex cells were also subjected to a 1 h heat-shock at 42 °C and a 2 h recovery, as indicated to demonstrate levels of *DNAJB1* induction observed during heat stress. The error bars show 95% confidence interval.

E. qPCR of *DNAJB1* in HEK293T-REx cells stably expressing FKBP.cHSF1 (blue) or dox-regulated cHSF1 (orange). Shield-1 (1 μM) or dox (2.25 μM) was added for the indicated time. Error bars show 95% confidence interval. This plot shows the raw fold increase for the data presented in **Figure 1E**.

F. qPCR of *HSP90AA1* in HEK293T-Rex cells stably expressing FKBP.cHSF1 (blue) or dox-regulated cHSF1 (orange). Shield-1 (1 μM) or dox (2.25 μM) was added for the indicated time. The expression of *HSP90AA1* was normalized to the maximal expression observed at 24 h. Error bars show 95% confidence interval.

G. Graph showing the survival of HEK293T-Rex cells stably expressing tet-inducible cHSF1 treated for 15 h with dox (2.25 μM) or acute heat shock (42 °C; 1h). n = 6; *p<0.01, **p<0.001.

Figure S2 (Supplement to Figure 2) - RNAseq transcriptional profiling of stress-independent HSF1 activation

A. Venn diagram showing the overlap of genes induced by stress-independent FKBP.cHSF1 activation in HEK293T-Rex (fold-increase >2; FDR <0.01; this study), heat shock (43 °C, 2 h) in Hela cells (fold-increase >1.4, FDR<0.1; GSM84590, GSM84602, GSM84603, GSM84591, GSM84592, GSM84593; (*8*)) or heat shock (44 °C, 2 h) in HEK293T (fold-increase >2; rpkm in controls >1; GSM1063575, GSM1063576; (*9*)). All of the published data used for this plot was reanalyzed prior to preparing the Venn diagram. The overlapping genes are shown.

B. qPCR of *BAG3* and *HSPE1* in HEK293T-Rex cells stably expressing FKBP.cHSF1 or dox-regulated cHSF1. Cells were treated for 15 h with either Shield-1 (1 μM) or dox (2.25 μM), as indicated. The error bars show 95% confidence interval.

Figure S3 (Supplement to Figure 3) – Stress-Independent HSF1 activation remodels cytosolic proteostasis pathways.

A. Graph depicting the complete cluster analysis of primary proteostasis factors (see **Table S3**) whose expression is highly correlated following HSF1 activation. The color indicates the fold change mRNA level observed for specific highly correlated proteostasis factors following Shield-1-dependent FKBP.cHSF1 stabilization, as indicated. *indicates gene only identified in cHSF1 overexpression RNAseq.

B. Graph showing the genes whose expression is highly-correlated with *DNAJA1* following HSF1 activation. The color indicates the fold-change mRNA induced by FKBP.cHSF1 stabilization, as indicated in **Figure S2A**.

C. Graph showing the genes whose expression is highly-correlated with *DNAJB4* following HSF1 activation. The color indicates the fold-change mRNA induced by FKBP.cHSF1 stabilization, as indicated in **Figure S2A**.

FIGURE S4 (SUPPLEMENT to FIGURE 6) – Stress-independent FKBP.cHSF1 activation attenuates intracellular aggregation of an aggregation-prone model protein.

Images depicting HEK293T cells transfected with either Q_0-tdTomato (left panels) or Q_{67}-tdTomato (right panels). Top panels are Cy5-filter set fluorescence images, bottom panels are phase contrast images.

FIGURE S1.

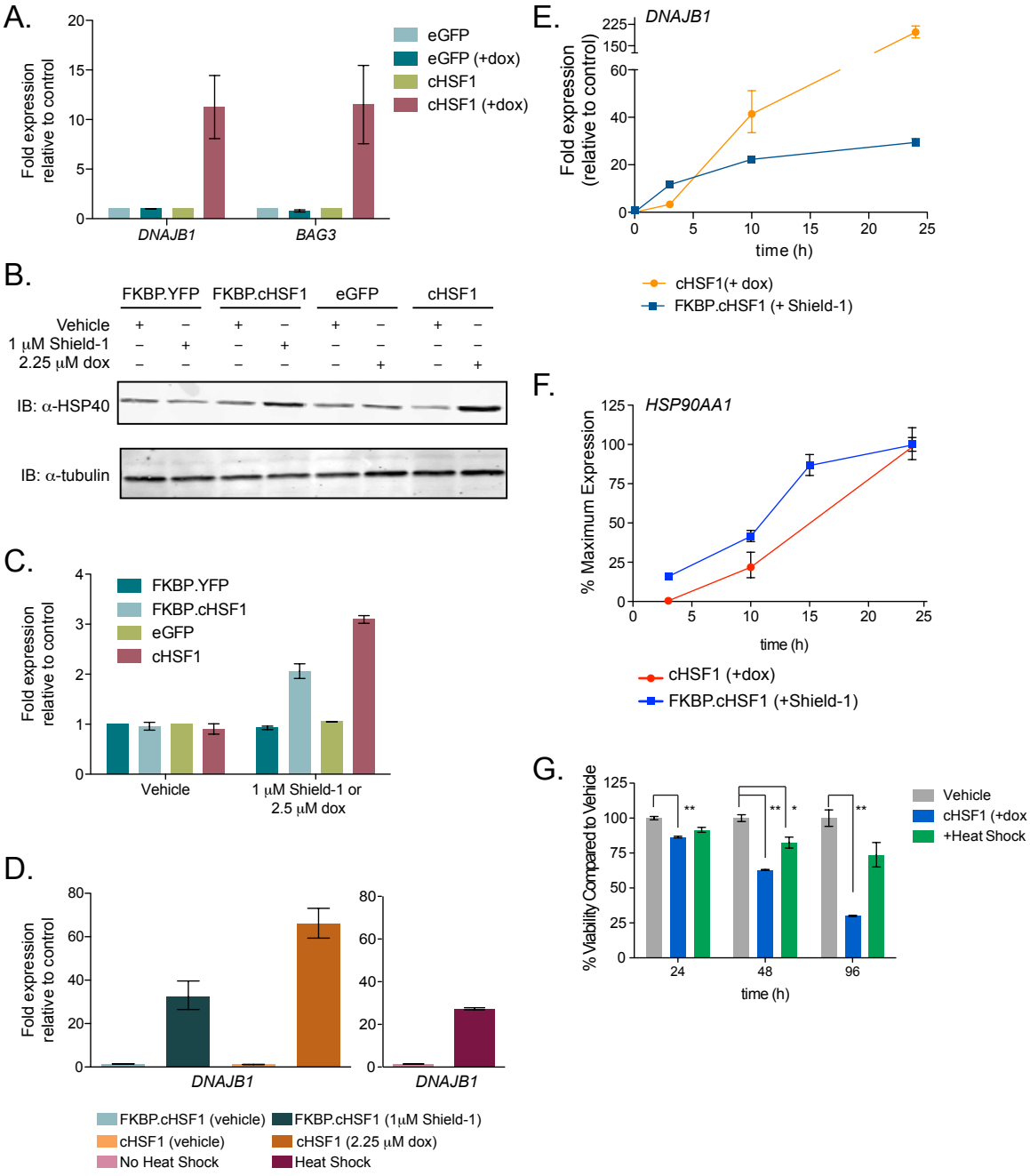
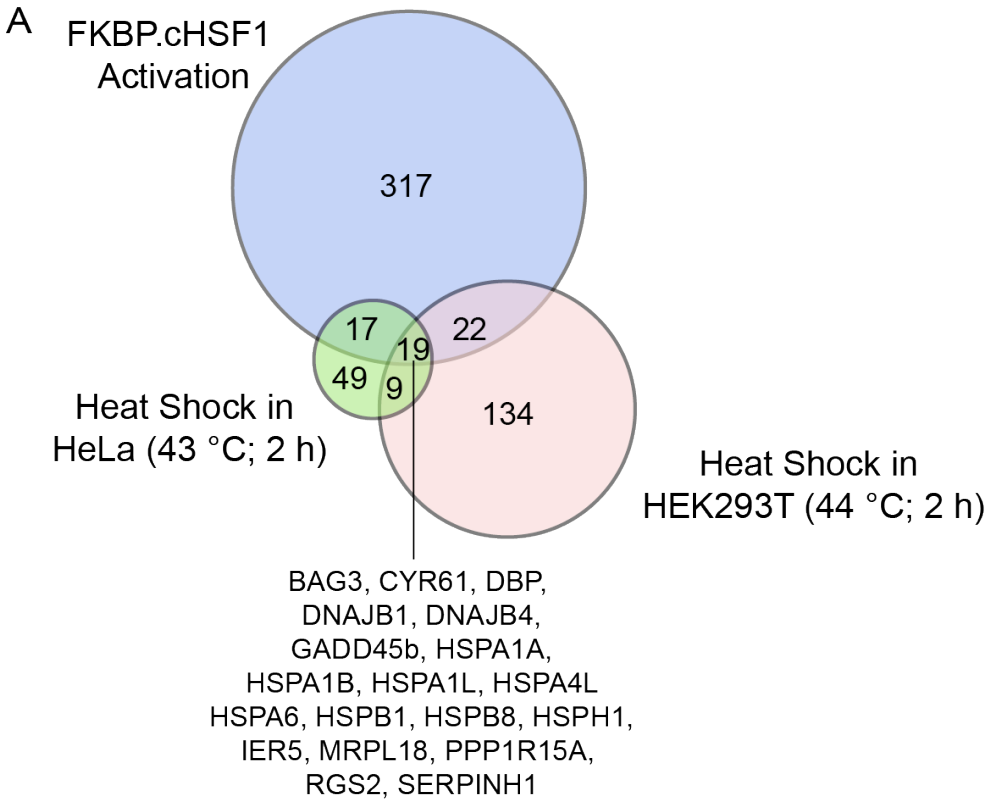

A.
Fold expression relative to control
eGFP
eGFP (+dox)
cHSF1
cHSF1 (+dox)
DNAJB1
BAG3
B.
FKBP.YFP
FKBP.cHSF1
eGFP
cHSF1
Vehicle
1 μM Shield-1
2.25 μM dox
IB: α-HSP40
IB: α-tubulin
C.
Fold expression relative to control
FKBP.YFP
FKBP.cHSF1
eGFP
cHSF1
Vehicle
1 μM Shield-1 or 2.5 μM dox
D.
Fold expression relative to control
DNAJB1
DNAJB1
FKBP.cHSF1 (vehicle)
FKBP.cHSF1 (1μM Shield-1)
cHSF1 (vehicle)
cHSF1 (2.25 μM dox)
No Heat Shock
Heat Shock
E.
DNAJB1
Fold expression (relative to control)
time (h)
cHSF1(+ dox)
FKBP.cHSF1 (+ Shield-1)
F.
HSP90AA1
% Maximum Expression
time (h)
cHSF1 (+dox)
FKBP.cHSF1 (+Shield-1)
G.
% Viability Compared to Vehicle
time (h)
Vehicle
cHSF1 (+dox
+Heat Shock

FIGURE S2


A
FKBP.cHSF1
Activation
317
17
19
22
49
9
134
Heat Shock in
HeLa (43 °C; 2 h)
Heat Shock in
HEK293T (44 °C; 2 h)
BAG3, CYR61, DBP,
DNAJB1, DNAJB4,
GADD45b, HSPA1A,
HSPA1B, HSPA1L, HSPA4L
HSPA6, HSPB1, HSPB8, HSPH1,
IER5, MRPL18, PPP1R15A,
RGS2, SERPINH1

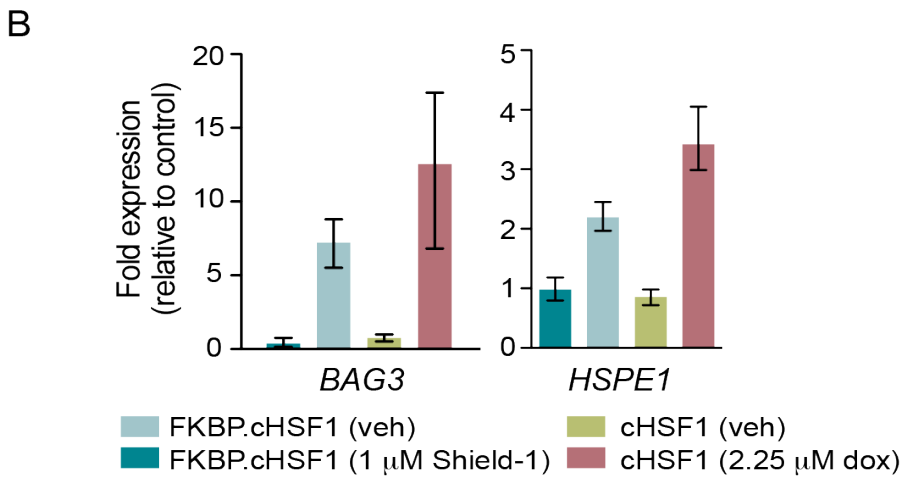

B
Fold expression
(relative to control)
20
15
10
5
0
5
4
3
2
1
0
BAG3
HSPE1
FKBP.cHSF1 (veh)
FKBP.cHSF1 (1 μM Shield-1)
cHSF1 (veh)
cHSF1 (2.25 μM dox)

FIGURE S3.

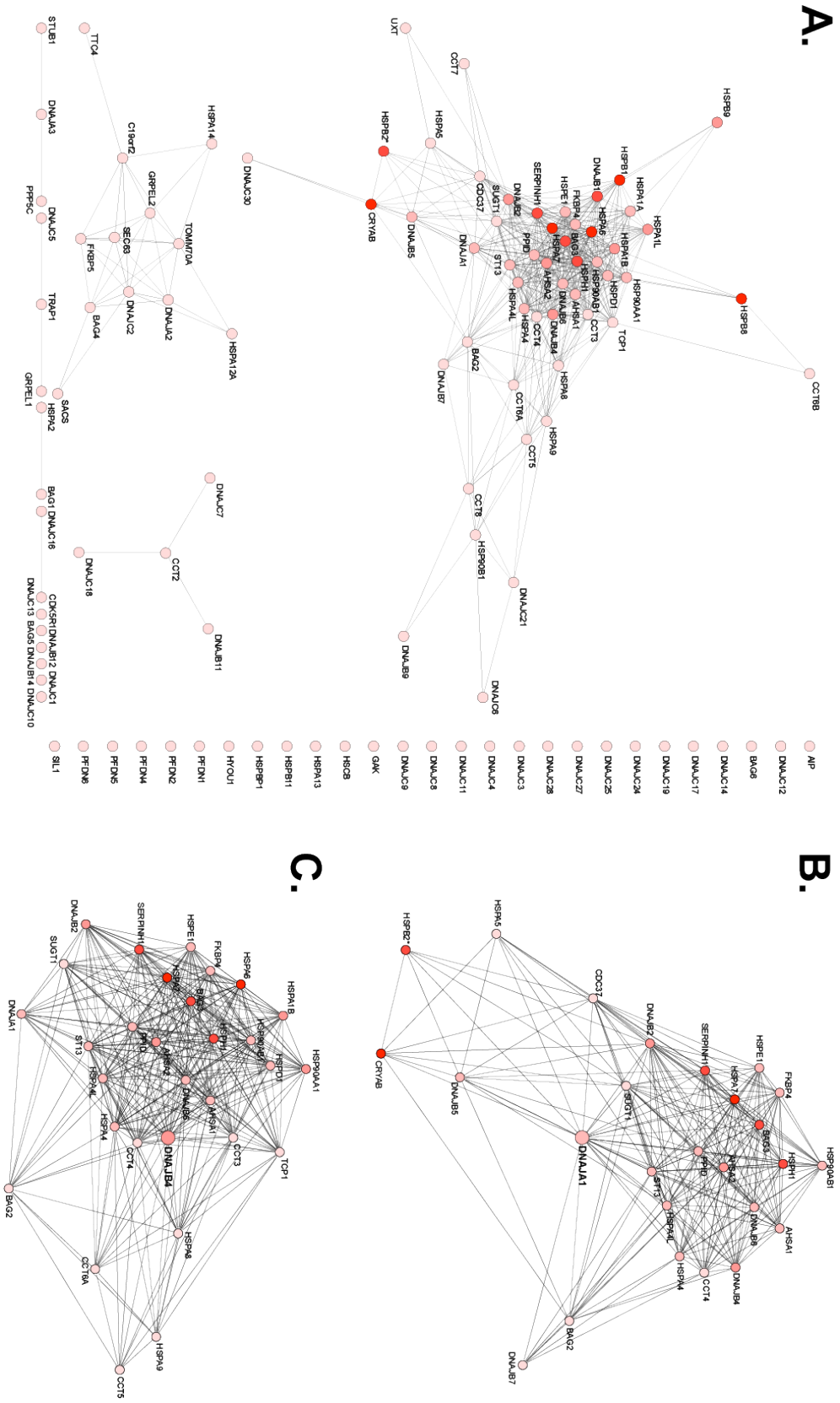
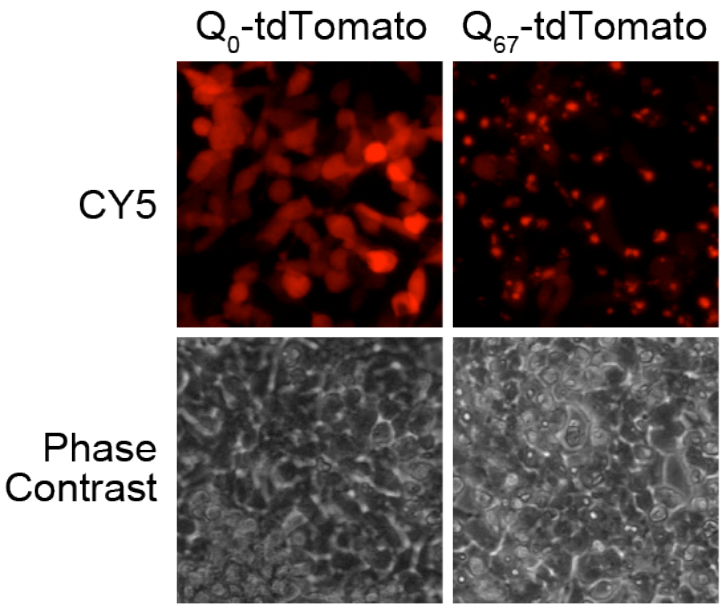

A.
B.
C.
STUB1
TTC4
DNAJA3
C19orf2
HSPA14
DNAJC30
PPP5C
GRPEL2
DNAJC5
SEC63
TOMM70A
FKBP5
TRAP1
DNAJC2
DNAJA2
BAG4
HSPA12A
GRPEL1
HSPA2
SACS
BAG1
DNAJC16
DNAJC7
DNAJC18
CCT2
DNAJB11
CDK5RI
DNAJB12
DNAJC1
DNAJC13
BAG5
DNAJB14
DNAJC10
UXT
CCT7
HSPB9
HSPB2
HSPA5
HSPB1
HSPA1A
CRYAB
DNAJB5
SERPINH1
DNAJB1
CDC37
SUGT1
HSPE1
FKBP4
DNAJB2
HSP90AB1
DNAJA1
PPID
HSPA1L
ST13
HSPA8
BAG3
HSPA1B
AHSA2
HSPH1
HSPD1
HSP90AA1
HSPA4L
DNAJB6
AHSA1
HSPB8
CCT3
HSPA4
CCT4
TCP1
DNAJB4
BAG2
CCT8B
DNAJB7
HSPA8
CCT6A
HSPA9
CCT5
CCT8
HSP90B1
DNAJC21
DNAJB9
DNAJC8
AIP
SIL1
PFDN6
PFDN5
PFDN4
PFDN2
PFDN1
HYOU1
HSPBP1
HSPB11
HSPA13
HSC8
GAK
DNAJC9
DNAJC8
DNAJC11
DNAJC4
DNAJC3
DNAJC28
DNAJC27
DNAJC25
DNAJC24
DNAJC19
DNAJC17
DNAJC14
BAG6
DNAJC12

FIGURE S4.



References

1. Anders, S., and Huber, W. (2010) Differential expression analysis for sequence count data, *Genome biology 11*, R106.
2. Wolters, D. A., Washburn, M. P., and Yates, J. R., 3rd. (2001) An automated multidimensional protein identification technology for shotgun proteomics, *Anal Chem 73*, 5683-5690.
3. McDonald, W. H., Tabb, D. L., Sadygov, R. G., MacCoss, M. J., Venable, J., Graumann, J., Johnson, J. R., Cociorva, D., and Yates, J. R., 3rd. (2004) MS1, MS2, and SQT-three unified, compact, and easily parsed file formats for the storage of shotgun proteomic spectra and identifications, *Rapid communications in mass spectrometry : RCM 18*, 2162-2168.
4. Peng, J., Elias, J. E., Thoreen, C. C., Licklider, L. J., and Gygi, S. P. (2003) Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome, *J Proteome Res 2*, 43-50.
5. Xu, T., Venable, J. D., Park, S. K., Cociorva, D., Lu, B., Liao, L., Wohlschlegel, J., Hewel, J., and Yates, J. R. (2006) ProLuCID, a fast and sensitive tandem mass spectra-based protein identification program, *Mol Cell Proteomics 5*, S174-S174.
6. Tabb, D. L., McDonald, W. H., and Yates, J. R. (2002) DTASelect and contrast: Tools for assembling and comparing protein identifications from shotgun proteomics, *J Proteome Res 1*, 21-26.
7. Park, S. K., Venable, J. D., Xu, T., and Yates, J. R. (2008) A quantitative analysis software tool for mass spectrometry-based proteomics, *Nat Methods 5*, 319-322.
8. 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.
9. Shalgi, R., Hurt, J. A., Krykbaeva, I., Taipale, M., Lindquist, S., and Burge, C. B. (2013) Widespread regulation of translation by elongation pausing in heat shock, *Molecular cell 49*, 439-452.