Coessential Genetic Networks Reveal the Organization and Constituents of a Dynamic Cellular Stress Response

David R. Amici[1,2,3,4], Jasen M Jackson[1,2,3], Kyle A. Metz[1,2,3], Daniel J. Ansel[1,2,3], Roger S. Smith[1,2,3,4], Sonia Brockway[1,2,3], Seesha R Takagishi[1,2,3], Shashank Srivastava[1,2,3], Brendan P. O'Hara[1,2,3], Byoungkyu Cho[5], Young Ah Goo[1,5], Neil L. Kelleher[1,5], Issam Ben-Sahra[1,2,3], Daniel R. Foltz[1,2,3], Marc L. Mendillo[1,2,3]

Corresponding Author: Marc L. Mendillo, mendillo@northwestern.edu

1. Dept. of Biochemistry and Molecular Genetics, Northwestern University, Chicago, IL, USA
2. Simpson Querrey Center for Epigenetics, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
3. Robert H. Lurie Comprehensive Cancer Center, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
4. Medical Scientist Training Program, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
5. Northwestern Proteomics Center of Excellence Core Facility, Northwestern University, Evanston, IL, USA

Summary

The interrelated programs essential for cellular fitness in the face of stress are critical to understanding tumorigenesis, neurodegeneration, and aging. However, modelling the combinatorial landscape of stresses experienced by diseased cells is challenging, leaving functional relationships within the global stress response network incompletely understood. Here, we leverage genome-scale fitness screening data from 625 cancer cell lines, each representing a unique biological context, to build a network of "coessential" gene relationships centered around master regulators of the response to proteotoxic, oxidative, hypoxic, and genotoxic stress. This approach organizes the stress response into functional modules, identifies genes connecting distinct modules, and reveals mechanisms underlying cellular dependence on individual modules. As an example of the power of this approach, we discover that the previously unannotated HAPSTR (C16orf72) promotes resilience to diverse stressors as a stress-inducible regulator of the E3 ligase HUWE1. Altogether, we present a broadly applicable framework along with an interactive tool (http://fireworks.mendillolab.org/) to interrogate biological networks using unbiased genetic screens.

INTRODUCTION

All living organisms retain an evolutionarily conserved capacity to adapt to physiological stress. Highly relevant to human disease, cellular stress response pathways empower aging neurons to clear protein aggregates, sun-exposed melanocytes to repair damaged DNA, and large tumors to proliferate despite having outgrown their blood supply. Certain types of stress, such as thermal stress, are easily studied. Others, however, such as the landscape of misfolded proteins in tumors, are more difficult to model experimentally. Moreover, diseased cells rarely face one stress in isolation. Do cells grown in a hypoxic incubator in pH-buffered, nutrient-rich media adequately model the stresses faced by an ischemic cardiomyocyte after vessel occlusion? Isolated stress perturbations are invaluable to provide mechanistic insight into individual stress response programs, but they fail to capture the complexity that may be found in disease states where these pathways function in a global, integrated network. Unfortunately, it is intractable to comprehensively study all possible combinations of different stresses and stress doses. Alternative strategies are needed to probe stress response pathways in disease-relevant contexts and to identify the mechanisms by which these pathways are integrated.

One approach to investigate context-specific biological relationships is to leverage the heterogeneity of cancer cell lines, each representing a unique genomic and epigenomic landscape obtained through rapid evolution from diverse cells of origin. In studying stress biology, cancer cells may be a particularly suitable model of physiologically relevant states, as tumors routinely co-opt stress response systems to

withstand the physiologic challenges imposed by carcinogenesis (Luo et al., 2009). For example, individual tumors may suppress the Hippel-Lindau tumor suppressor (VHL) to induce constitutive hypoxia signaling (Maxwell et al., 1999). Other tumors may alternatively splice pyruvate kinase to promote a metabolic shift toward aerobic glycolysis (Christofk et al., 2008). Still others may activate Heat Shock Factor 1 (HSF1) to coordinate protein homeostasis (proteostasis) during anabolic proliferation (Santagata et al., 2013). Thus, the contextual breadth of cancer cells and the multitude of mechanisms by which they activate stress response pathways may provide unique insights into the regulation and integration of these pathways in diseased cells.

Large-scale fitness screening efforts have now quantified the essentiality of most protein-coding genes across hundreds of different cancer cell lines (Hart et al., 2015; Meyers et al., 2017; Tsherniak et al., 2017). These studies reveal that, even in cancer cells, most critical regulators of the stress response are not universally essential for cell viability. But are all cancer cell lines similarly affected by the loss of individual stress response factors? If not, what mechanisms underlie the requirement for these factors in some cell lines but not others, and which other genes share this context-specific phenotype? Regarding the latter question, recent work has demonstrated that correlated patterns of gene essentiality ("coessentiality") across many cancer cell lines can indeed identify functional relationships between genes (Boyle et al., 2018; Kim et al., 2019; Pan et al., 2018; Wang et al., 2017). These studies utilize a top-down approach, organizing the genome into coessential clusters which have already proven to yield valuable insights. However, due to technical limitations, such as spurious correlations between genes within genomic regions subject to copy number variation and the relatively modest signal for

genes which function outside of obligate protein complexes, much of the genome is left incompletely explored by these methods. Overcoming these technical limitations may thus expand the reach of this approach to biological systems of greater regulatory complexity, such as the cellular stress response network.

In this article, we develop a framework to use genome-scale, CRISPR-Cas9 fitness screening data from 625 cancer cell lines to build a bottom-up network of coessential relationships centered around critical regulators and effectors of the stress response. Our approach organizes functional modules within the stress response, corresponding to both canonical and novel functions of these factors, and identifies genes which represent points of crosstalk between distinct modules. We demonstrate that context, in tumor subtype and cell lineage, is associated with dynamic changes in the organization and utilization of stress response modules in a manner reflecting therapeutically-targetable biology. Finally, we identify novel genes in a global stress network, validating that the previously unannotated HAPSTR (C16orf72) promotes cellular adaptation to diverse stressors as a stress-inducible regulator of HUWE1, an E3 ligase which modulates genotoxic, proteotoxic, and hypoxic stress responses (Kao et al., 2018). More broadly, we demonstrate the power of our approach using the global stress response network as proof of principle and provide an interactive web-based tool (http://fireworks.mendillolab.org/) to facilitate studies of other biological networks.

RESULTS

Assembling a coessentiality network around the master regulator of cytosolic proteostasis

Previous reports have demonstrated that genes with shared functions, particularly those which encode components of obligate protein complexes, have convergent effects on fitness when deleted in cancer cell lines (Kim et al., 2019; Pan et al., 2018). Consistent with prior analyses performed in few cell lines, network analysis of the strongest coessential relationships in the genome (r>0.6; 1532 pairings) successfully identifies modules of biological relevance with highly correlated fitness profiles (Figure S1B). However, these clusters are dominated by few pathways, many of which center on major protein complexes – primarily, the ribosome, oxidative phosphorylation, and spliceosome machinery. Indeed, the gene pairs with the most correlated (coessential) knockout fitness effects overwhelmingly represent protein complex members (86% of top gene pairs; 23% expected by chance; Figure S1C). For example, the tuberous sclerosis complex proteins TSC1 and TSC2 represent the strongest coessential relationship in the genome (r=0.92, p=1e-269). Indeed, despite remaining highly statistically significant, the average magnitude of the top-ranked correlation for genes encoding transcription factors is substantially less than that for genes in the CORUM human protein complex database (Figure S1D). Moreover, in this network, non-informative clusters emerge which correspond to genes in the same chromosomal region, likely reflecting the confounding fitness effect of double-stranded breaks in regions with copy number alteration (CNA) (Aguirre et al., 2016) (Figure S1B). CNA-based corrections, now standard in defining

gene essentiality from CRISPR-Cas9 screens (Meyers et al., 2017), address the magnitude of signal attributable to a gene at an amplified locus. However, this uniform correction does not address the shared patterns of relative essentiality between genes at that locus, fundamentally biasing correlation-based analyses. Thus, in a genome-scale, top-down coessentiality analysis, true biological signal is lost for genes encoding proteins which function independently from large molecular assemblies and/or have variable copy number across tumors – such as master regulator transcription factors of the cellular stress response (Figure S1E).

To address the limitations of existing coessentiality approaches, we developed a bottom-up approach to generate a locus-adjusted, rank-based network from central source node(s) of interest (Figure 1A). We first investigated HSF1, a transcription factor considered the master regulator of cytosolic proteostasis. Across cancer cell lines of diverse cell lineage and tumoral subtype, HSF1 loss negatively impacts cellular fitness, with a spectrum of relative essentiality within each lineage (Figure 1B). Notably, HSF1 is located on chromosome 8q, a region commonly amplified in cancer. In contrast to HSF1 transcript levels (Figure 1C), HSF1 essentiality was not correlated with copy number (Figure 1D). This reflects the copy number adjustment performed to obtain the CERES score, a continuous measure of the fitness effect of individual gene loss (Meyers et al., 2017). However, the genes most coessential with HSF1 still showed substantial enrichment for genes in the same chromosomal neighborhood ($p = 7e-5$; Figure 1E). We found that correcting for the median fitness effect of neighbor genes prior to performing genome-wide correlations eliminated this locus bias (Figure 1E).

We next profiled the top ranked HSF1-coessential genes using our locus-adjusted method. Remarkably, despite lower magnitudes of correlation as compared with protein complexes, the genes with fitness profiles most similar to HSF1 included genes having known functional relationships with HSF1, such as Heat Shock Factor 2 (HSF2), HSP70 and HSP110 family members (HSPA4, HSPA14), and HSP90 ligands (ANKRD49, FKBPL). Gene set enrichment analysis (GSEA) of top HSF1-coessential genes confirmed enrichment for canonical HSF1-related processes (i.e. chaperone-mediated protein folding; Figure 1F). GSEA also corroborated pathways with emerging relationships to HSF1, such as ribosome biogenesis and RNA processing, where HSF1 senses ribosome activity during translation (Santagata et al., 2013), and mitochondrial protein import, where HSF1 drives the multifaceted response to cytosolic accumulation of non-imported mitoproteins (Boos et al., 2019).

To better understand functional modules within the HSF1 coessential gene set, we generated an HSF1 coessentiality network containing the top 100 HSF1-coessential genes and up to 5 secondary connections per gene (Figure 1G). Like GSEA, the network approach reveals functional modules connecting to HSF1, with the additional advantage of illuminating the interplay within and between modules. For example, enzymes critical to protein glycosylation or UFMylation connect only to HSF1 and other genes in the same functional module. On the other hand, nucleolin (NCL) – a heat-shock responsive nucleolar protein which functions in the first step of rRNA processing (Daniely et al., 2002; Ginisty et al., 1998) – is central to a subnetwork which bridges the heat shock and RNA processing moedules.

We next investigated whether the HSF1 coessentiality network would vary by cancer subtypes, indicating lineage-specific biological relationships which did not emerge in pan-cancer analyses. As expected, the correlation of HSF1 with heat shock module genes was the most consistent correlation lineages, indicating that this canonical role of HSF1 is highly conserved across tissue types and genetic backgrounds (Figure 1H). Gene sets corresponding to other biological processes which have been previously linked to HSF1, such as glycolysis (Dai et al., 2007; Santagata et al., 2013; Zhao et al., 2009), had similar knockout phenotypes to HSF1 only in some cancer subsets, suggesting that context may drive these relationships (Figure 1H).

A notable exception to the otherwise subtype-conserved correlation between HSF1 and heat shock genes was found in acute myeloid leukemia (AML), where no relationship was observed between HSF1 and the heat shock module (Figure 1H). Rather, the most correlated module comprised genes involved in ribosome biogenesis. To better understand the contexts driving HSF1 essentiality in AML, we compared the transcriptome, metabolome, and proteome for the cell lines most dependent on HSF1 with the cell lines least dependent on HSF1. Strikingly, nearly every gene overexpressed in the HSF1-dependent AML lines encoded a protein involved in translation (enrichment $p = 6e\text{-}25$). Fittingly, protein levels of eukaryotic elongation factor 2 (EEF2) and p70s6k were elevated in these cell lines (Figure 1I), together indicating a high-translation phenotype which does not correspond to any known subtype of AML. Remarkably, a previous report described that inhibiting translation initiation with rocaglates potently inactivates HSF1, suppressing the high-translation malignant state in a manner which most potently impacted AML cells (Santagata et al., 2013). Notably, HSF1 dependence

does not strongly correlate with this RNA signature of translation in other cancer subtypes (Figure 1J). Moreover, the observation that this phenotype exists in a spectrum within AML garners further consideration, given the ongoing exploration of rocaglates/translation initation inhibitors for cancer care (Cunningham et al., 2018). Gene expression data from AML patient tumors reveals that the protein synthesis genes identified in our cell line analysis above stratify patients into distinct prognostic populations (Figure 1K), suggesting that translation initiation inhibitors may be most efficacious in this patient subset. Moreover, these data suggest the existence of a low-translation, poorer prognosis group which may require alternative therapeutic approaches. Altogether, these data identify canonical and context-specific functions for HSF1, integrate genetic fitness data in a manner relevant to targeted cancer therapeutics, and serve as proof of principle for bottom-up, locus-adjusted coessentiality network analysis.

A chaperone coessentiality network resolves functionally and spatially differentiated proteotoxic stress responses

While HSF1 is a critical upstream regulator of the proteotoxic stress response, the cellular proteostasis network also includes hundreds of effector chaperones and co-chaperones, many with distinct clients, patterns of activation, and subcellular localization. Reflecting the functional nature of chaperones to physically interact with client proteins, our understanding of chaperone biology is highly influenced by physical interaction studies, which are often constrained by the overexpression of tagged proteins, limited numbers of cell lines, and chemical inhibitors which fail to distinguish highly similar family

members. Our genetic approach is not subject to these limitations and can identify functional relationships which do not require physical interaction. Thus, we expanded our network from HSF1 to the major chaperone families to provide functional insights into the broader cellular protein homeostasis network.

We first investigated the essentiality of the HSP40/J-domain protein (JDP), HSP70, HSP90, HSP110, HSP60/10, and small HSP chaperone family members (Figure 2A). Even within these families comprising highly related members, we observe great variability in the essentiality of individual chaperones. For example, within the HSP70 family, HSPA9 (Mortalin) and HSPA5 (BiP/Grp78) are broadly required for cellular proliferation (mean essentiality, 1.53 and 1.32; essentiality rank, 203 and 436 of 18333), whereas targeting either HSPA12A or HSPA12B has no deleterious effect on cellular fitness whatsoever (mean essentiality rank, 16738 and 13407). While most HSP70 family members are cytosolic, the two highly essential HSP70s represent the primary HSP70 in the mitochondria (HSPA9) and endoplasmic reticulum (HSPA5). Likewise, the ER-localized HSP90B1 (Grp94) is more essential than the cytosolic HSP90AA1 and HSP90AB1, suggesting that the specialization of these factors limits the capacity of the cell to buffer their loss via the expression of more closely related paralogs. Supporting this model, we observe that dependence on individual cytosolic HSP90 isoforms is predicted by relative expression of that isoform (Figure S2A,B). For example, leukemias preferentially express and depend on HSP90AB1, whereas cervical cancers preferentially express and depend on HSP90AA1 (Figure S2C). Because therapeutic modulation of chaperones remains a goal in many areas of patient care, these data argue for the

consideration of paralog redundancy/buffering for any therapy targeted to a particular chaperone.

Next, to investigate interplay between chaperone genes, we constructed a coessentiality network using these chaperone family members as source nodes (Figure 2B). Topologically, the chaperone network is more interconnected than would be expected by chance ($p<0.001$ vs. randomly permuted network), with several canonical relationships recapitulated. These included strong reciprocal connections between specific HSP70s and the JDP cochaperones with which they function, such as HSPA14 and DNAJC2. To further analyze the network, we performed Markov clustering, which revealed several distinct modules containing clear biological connections (Figure 2C). Cluster 1 contains the chaperones most correlated with core cellular fitness processes, such as DNA replication (PCNA), translation (RPL11, RPL37, RPS13), and transcription (POLR2J). Cluster 5 contained five member genes, four of which have known functions at the inner mitochondrial membrane, strongly suggesting that the outlier node (DNAJC24) also plays a role in mitochondrial protein maintenance. Remarkably, the 11 genes in Cluster 2 all localize to the ER lumen and have known functional connections to the ER stress response. On the other hand, multiple clusters emerged containing primarily to cytosolic proteins. To support the biological segregation of these cytosolic and ER lumen gene sets beyond localization, we investigated the transcriptional response of MDA-MB-231 cells to either 2-deoxyglucose (2DG; an inducer of protein misfolding in the ER lumen) or heat shock (HS; an inducer of protein misfolding in the cytosol). We found striking induction of ER cluster genes by 2DG, but not HS, and induction of cytosolic cluster genes by HS but not 2DG (Figure 2C, rightmost panels), indicating a functional

differentiation of these modules in the response to distinct proteotoxic stressors. More broadly, these data demonstrate that bottom-up coessentiality network approach identifies functional modules, points of functional redundancy, and putative novel interactions within a highly interconnected and compartmentalized proteostasis network.

Identifying topology and crosstalk between functional modules in a global stress response network

As disease states often impose multiple simultaneous stressors, we next sought to expand beyond proteostasis to investigate the crosstalk between diverse stress response programs. Thus, we used the approach outlined above to generate a global stress response coessentiality network centered around master regulators of the response to genotoxic, proteotoxic, hypoxic, nutrient, and oxidative stresses (Figure 3A). For this network, we also included highly-ranked negative correlations to capture antagonistic relationships (Figure 3B).

The stress network (Figure 3B), comprising proteins with diverse molecular activities (Figure 3C), has far greater connectivity than would be expected by chance alone. Moreover, this network is replete with examples of canonical signaling relationships. For example, within the genotoxicity cluster, TP53 and ATM share connections with many genes critical to the response to double-stranded DNA breaks (e.g. CHEK2, CDKN1A/p21, TP53BP1). NFE2L2/NRF2 was highly anticorrelated with its canonical negative regulator, KEAP1, sharing multiple divergent nodes – i.e. nodes correlated to one source node and anticorrelated with the other. These divergent nodes included PDCD10/CCM3, which phosphorylates the Ezrin/Radixin/Moesin family proteins

to promote fitness under oxidative stress (Fidalgo et al., 2012). The HIF family of transcription factors were also tightly connected, with the exception of the atypical HIF3A. Furthermore, VHL anticorrelated with both HIF1A and EPAS1 (HIF2A), but not ARNT (HIF1B), consistent with the known function of VHL to stimulate degradation of HIFα but not HIFβ proteins. The heat shock factor paralogs HSF1 and HSF2 were strongly connected, both displaying connections to the genotoxic stress cluster via genes such as NCL, which is known to mobilize from the nucleolus to nucleoplasm after heat shock or DNA damage in a p53-dependent process (Daniely et al., 2002).

Other clear modules in the stress network included a large cluster corresponding to ERN1 and XBP1. In response to proteotoxic stress in the ER lumen, ERN1 cleaves XBP1 to initiate one of the three branches of the unfolded protein response (UPR). Indeed, we find that ERN1 and XBP1 share a dense cluster of coessential genes enriched for UPR/ER stress signaling (e.g. HERPUD1, DNAJC3; enrichment p = 1e-9). Interestingly, EIF2AK3 and ATF6 – the sensors for the other two branches of the UPR – did not share many coessential genes with ERN1 and XBP1, consistent with recent data suggesting that specific contexts confer differential dependence on different branches of the UPR (Adamson et al., 2016). An exception to the lack of connectivity between UPR branches is MANF, which correlates with XBP1, but anticorrelates with EIF2AK3/PERK. Consistent with our network, MANF has conserved, divergent genetic interactions with XBP1 and EIF2AK3, but not ATF6 (Lindstrom et al., 2016). Moreover, MANF was recently identified to stabilize a subset of HSPA5-client complexes in the ER lumen (Yan et al., 2019). Indeed, HSPA5 is highly coessential with MANF in addition to correlating with

XBP1 and anticorrelating with EIF2AK3, supporting a model in which MANF is a point of regulation between the functionally differentiated branches of the UPR.

Another highly interconnected cluster comprised genes involved in nutrient stress/autophagy, including the source nodes EIF2AK4 (GCN2) and GCN1 (nutrient sensors), ATG5 and ATG7 (critical genes for autophagosome formation), TSC1 and TSC2 (negative regulators of mTORC1), and ATF4. ATF4 is perhaps best known for its role downstream of EIF2AK3/PERK in the UPR, and does indeed connect to EIF2AK3 in the network via Thrombospondin 1 (THBS1), an ER-localized protein which promotes a protective ER stress response (Lynch et al., 2012). However, ATF4 clusters much more closely with the nutrient stress/autophagy genes, consistent with a more recently described role for ATF4 in mTORC1 signaling (Ben-Sahra et al., 2016). More globally, the autophagy cluster connected strongly with the oxidative stress module, featuring crosstalk genes such as TXNDC17 (TRP14), a disulfide reductase which has both peroxidase (Jeong et al., 2004) and autophagy-inducing (Tan et al., 2019; Zhang et al., 2015) activities.

While the stress network readily identifies canonical stress response signaling relationships, not all of the genes in the stress network have established connections to stress response biology. This may be because many of these genes have not been studied directly (Figure 3E). To test whether the genes in our network are regulated by stress, we treated MDA-MB-231 cells with five different perturbations which capture the breadth of the stress network (CoCl2: hypoxia mimetic, Cyclophosphamide/CPA: DNA damaging agent (CoCl2: hypoxia mimetic, Cyclophosphamide/CPA: DNA damaging agent, 2DG: ER stress inducer and metabolic stress, serum starvation (SS): nutrient

stress, and HS: cytosolic proteotoxic stress). We observed that many genes in our stress network are stress-responsive at the transcript level, including three genes induced by multiple stressors which have not been studied directly (CCDC113, C16orf72, UBALD1; Figure 3F). While most genes were regulated by multiple stressors, certain genes corresponding precisely to one module were primarily induced by the stressor most associated with that module, such as TSC2 in serum/amino acid starvation, EGLN1 in CoCl2/pseudohypoxia, and the previously discussed MANF in 2DG/ER stress (Figure 3F). Altogether, the stress network recapitulates positive and negative regulatory interactions between distinct stress response modules, identifies genes which represent points of crosstalk between modules, and identifies a set of minimally-studied genes as putative members of an integrated stress response.

Cancer cell dependence on dynamic stress modules reflects targetable vulnerabilities

While pan-cancer coessentiality analysis allows for the greatest sampling of distinct cell states, there are likely lineage-specific relationships within the stress network which might be obscured at the pan-cancer level. To identify such relationships, we constructed a stress response network using cancer cell lines from each lineage containing at least 10 cell lines and systematically assessed connectivity between source nodes (Figure 4A).

Across lineages, the most highly connected stress response subnetwork corresponded to the TP53-ATM genotoxicity cluster, which had members in nearly every lineage, but was most predominant in skin cancers (18 members in skin; mean members

in other lineages, 5.4) (Figure 4B). This is perhaps unsurprising, as melanoma classically contains the largest mutational burden of major cancers (Zhang et al., 2018). In contrast to the relatively conserved genotoxicity subnetwork, we identified an ATF3-EIF2AK3 network highly specific to neuroblastoma (17 members; one total member in other lineages). This network was comprised entirely of divergently correlated nodes – i.e. correlated with ATF3 and anticorrelated with EIF2AK3/PERK, or vice versa (Figure 4C). Notably, in neuroblastoma, ATF3 and EIF2AK3 are each the most anticorrelated gene for the other (r= -0.87, p=6e-6) while they are unrelated in the pan-cancer analysis (r=-0.04, p=0.3). ATF3 functions as a transcriptional repressor, and is induced by EIF2AK3 downstream of the ER stress response (Jiang et al., 2004). Querying ENCODE ChIP-seq data, we find that ATF3 can bind an EIF2AK3 enhancer (Figure S4). Taken together, these data suggest that ATF3 and EIF2AK3 operate in a negative feedback loop particularly relevant to neuroblastoma.

Another subnetwork with lineage-enriched membership was the ERN1-XBP1 UPR subnetwork, which had 15 members in the hematopoietic lineage network (mean membership in other lineages, 0.75) (Figure 4D). Interestingly, unlike the skin network (Figure 4B), which featured crosstalk between all three branches of the UPR, the hematopoietic lineage network was enriched only for the ERN1-XBP1 branch. We find that that this ERN1-XBP1 relationship in hematopoietic cells is primarily driven by multiple myeloma (Figure 4E), a malignancy characterized by constitutive UPR signaling due to high protein secretion load (Obeng et al., 2006). (Figure 4E). Strikingly, the XBP1-ERN1 network in myeloma was enriched for both network connectivity (highly correlated within the lineage) and network essentiality (selective essentiality within the lineage) (Figure

4F). Thus, the ERN1-XBP1 UPR branch is an exquisite vulnerability specific to a subset of multiple myeloma which could, in principle, be exploited by targeting any individual member of the network.

To further investigate this ERN1-XBP1-dependence in multiple myeloma, we compared available transcriptome, metabolomic, and drug sensitivity data for the myeloma lines most dependent on this network with those least dependent on the network. We found that this high-dependency subset of myeloma was characterized by a transcription and metabolic signature suggesting diminished oxidative metabolism (Figure 4G). Interestingly, these cell lines were more sensitive to Bortezomib (Figure 4I), a proteasome inhibitor used clinically in multiple myeloma which induces pro-apoptotic UPR signaling (Obeng et al., 2006). In contrast, the low-UPR-dependent, relatively proteasome inhibitor-resistant subset of myeloma was particularly sensitive to Elesclomol, an inhibitor of oxidative metabolism recently identified as a selective vulnerability in proteasome inhibitor-resistant myeloma (Tsvetkov et al., 2019). To investigate the relevance of these data to clinical care, we queried tumor microarray data from a clinical trial of Bortezomib, finding that the patients whose tumors shared this low-OXPHOS signature (>1 standard deviation from mean; n=29) were more likely to respond to Bortezomib therapy as compared with the other enrolled patients (n=159; Figure 4J). Altogether, these data demonstrate the utility of intra-lineage coessentiality network analysis to identify regulatory relationships and therapeutic vulnerabilities particularly relevant to cells of a specific lineage.

C16orf72 is a stress-inducible modulator of the cellular adaptation to diverse stressors

To demonstrate the power of our approach to identify a novel gene critically relevant to the stress response, we chose to study C16orf72, a multi-stress-inducible node in the stress network which remains entirely unannotated (Figure 3C,E-F). Mining publicly available datasets, we found that C16orf72 is indeed expressed across cancer cell lines, normal tissue, and tumors (Figure S7), with its expression in tumors being prognostic in many cancers (Figure 5A). C16orf72 loss is broadly deleterious to the fitness of cancer cell lines particularly those derived from kidney tumors (Figure 5B). Evolutionarily, C16orf72 is highly conserved through C. elegans (Figure 5C), with CRISPR-Cas9 fitness screening data in Drosophila cells (Viswanatha et al., 2018) indicating that its importance for cellular viability is also conserved (Figure 5D).

To validate the essentiality data for C16orf72 using an orthogonal approach, we tested the effect of C16orf72 depletion with siRNA on the short-term proliferation of three breast cancer cell lines with varying dependence on C16orf72: SK-BR-3 (C16orf72 essentiality score, 1.65; cell line rank, 1/625), ZR-75-1 (essentiality score, 1.09; cell line rank, 38/625), and MDA-MB-231 (essentiality score, 0.05; cell line rank 581/625). As expected, proliferation of SK-BR-3 cells was most impacted by C16orf72 depletion, followed by ZR-75-1, with no fitness effect at 72 hours in MDA-MB-231 cells (Figure 5E). Interestingly, C16orf72 depletion produced a similar overall transcriptional effect in each of these three cell lines (Figure 5F), consistently inducing genes related to DNA damage (e.g. CDKN1A/p21) and suppressing genes related to inflammatory, hypoxic, and other stress response programs.

C16orf72 emerged in our stress network, is induced by multiple stressors, and alters the transcription of stress response genes. Moreover, the C16orf72-coessentiality network contained clear modules related to genotoxic, hypoxic, proteotoxic, and nutrient stress (Figure 5G). To test the hypothesis that C16orf72 plays a broad role in the cellular response to stress, we designed a targeted screen to test the effects of C16orf72 depletion on cancer cells subjected to perturbagens which induce these diverse stress response pathways. We found that C16orf72-silenced cells were strikingly less tolerant to these stressors, except for the proteasome inhibitor MG132 (Figure 5H). Of note, MDA-MB-231 cells – which did not exhibit a proliferation defect upon C16orf72 depletion in the non-stressed, baseline condition – had a marked fitness defect when C16orf72 depletion was combined with stressors.

Because C16orf72 depletion most sensitized MDA-MB-231 cells to CoCl2 (hypoxia mimetic), serum starvation (nutrient stress), and CPA (DNA-damaging agent), we chose to further investigate the interplay between C16orf72 and these three stressors. Corroborating the ATP-based viability assay (Figure 5H), live cell imaging revealed reduced proliferation of C16orf72-depleted MDA-MB-231 cells in these conditions as compared with controls (Figure 5I-K). Moreover, knockdown of C16orf72, which was induced from baseline levels by each of the three stressors in the siNT control group, dramatically altered the transcriptional response to these stressors (Figure 5I-K). For example, 1367 genes were differentially expressed in serum starved cells without C16orf72 as compared to serum starved cells with C16orf72 (Figure 5I-K). Across all three stress conditions, knockdown of C16orf72 suppressed expression of genes corresponding to stress and inflammatory response pathways. Additionally, context-

specific patterns of regulation emerged, whereby C16orf72 knockdown resulted in misregulation of amino acid metabolism genes under amino acid starvation, replication stress genes under interstrand crosslinking conditions, and oxygen response genes under hyperactive oxidative radical production. Altogether, our data identify C16orf72 as a conserved, stress-inducible gene which promotes cellular fitness in the face of diverse stressors.

C16orf72/HAPSTR modulates the stress response by interacting with the stress-responsive E3 ligase HUWE1

To investigate the mechanism by which C16orf72 enables this striking resilience to stress, we first expressed a 3xFLAG-tagged C16orf72 construct in HEK293T cells. C16orf72 predominantly localized to the nucleus (Figure 6A), driven by a conserved nuclear localization signal at the protein's N-terminus (Figure 5C, S6). Next, we performed affinity purification coupled with mass spectrometry to identify proteins which interact with C16orf72. The most prominent hit was the Hect ubiquitin E3 ligase HUWE1 (Figure 6B), which has known functional roles in the response to diverse stress conditions, such as genotoxicity, proteotoxicity, and hypoxia (Choe et al., 2016; Kao et al., 2018; Maghames et al., 2018). Confirming a physical interaction, we co-immunoprecipitated endogenous HUWE1 with C16orf72 using antibodies targeting both HUWE1 and FLAG in 293T cells (Figure 6C,D).

Interestingly, HUWE1 is also C16orf72's most coessential gene (r=0.49; Figure 6E). Several other proteins with which C16orf72 physically interacts also shared correlated fitness profiles with C16orf72 (Figure 6F). In contrast with the anticorrelated

fitness profiles often observed between E3s and the targets they mark for degradation (e.g. MDM2 and TP53, r = -0.704), the highly convergent fitness profiles of C16orf72 and the E3 ligase HUWE1 suggest a cooperative function for these interacting proteins. Supporting the hypothesis that C16orf72 is not simply a HUWE1 substrate, we observe no significant change in C16orf72-FLAG expression or ubiquitination in response to HUWE1 knockdown (Figure 6G). Moreover, treatment of MDA-MB-231 cells with siRNAs targeting C16orf72, HUWE1, or both genes resulted in strikingly similar transcriptional profiles (Figure 6H). Taken together, these data strongly suggest a direct, cooperative physical interaction between C16orf72 and HUWE1.

To next address whether this putative cooperative function of C16orf72 may be related to the E3 ligase activity of HUWE1, we stably expressed FLAG-tagged ubiquitin under a doxycycline-inducible promoter in 293T cells and used FLAG affinity purification-mass spectrometry to identify changes in the ubiquitinated proteome after knockdown of either HUWE1 or C16orf72. Across three separate experiments, including baseline conditions (48h and 72h of knockdown) and 24h of serum starvation (72h of knockdown), C16orf72 and HUWE1 had nearly identical effects on the ubiquitinated proteome (Figure 6I). Specifically, C16orf72 and HUWE1 depletion independently resulted in under-ubiquitination of many proteins (Figure 6J). These included known HUWE1 substrates, such as TP53 (notably, C16orf72's second-ranked fitness anticorrelation; r = -0.32, p = 6e-17), as well as putative HUWE1 substrates suggested from prior high throughput studies, such as GNL3/nucleostemin, a protein which promotes genome integrity during DNA replication (Tsai, 2014). Remarkably, of the 20 proteins which had highly significant changes in ubiquitination after C16orf72 and HUWE1 depletion, only UBE2L3 had

increased ubiquitination. UBE2L3 is an E2 ligase which functions to transfer ubiquitin directly to the active-site residue of HUWE1 (a unique feature of Hect E3s). Together, increased ubiquitination of HUWE1's E2 ligase and decreased ubiquitination of its substrates demonstrate that C16orf72 depletion is sufficient to markedly suppress HUWE1 ubiquitin ligase activity in a manner independent of a change in HUWE1 expression (Figure 6J,K). Moreover, C16orf72 depletion recapitulates the effect of HUWE1 loss on protein levels of established HUWE1 substrates (Mcl1, DDIT4, p53) as well as the putative substrate GNL3/nucleostemin (Figure 6K). Altogether, these data demonstrate a critical role for C16orf72 as a a stress-responsive cofactor for HUWE1, a fitness-essential gene known to modulate the stress response (Kao et al., 2018). Considering these data, we propose that this Huwe1-associated modulator of proliferation and the stress response be named HAPSTR. More broadly, the framework we use to identify HAPSTR provides a paradigm for the identification of other uncharacterized genes.

DISCUSSION

Most genes in the human genome are not essential for cellular proliferation in standard culture conditions (Hart et al., 2015; Meyers et al., 2017; Tsherniak et al., 2017). Yet, many of these "nonessential" genes are critical to cell fitness in the context of physiological stress. Foundational work has demonstrated that, in yeast, only ~17% of the genome is essential for growth in rich media (Winzeler et al., 1999), but nearly every gene (~97%) becomes required for optimal growth in at least one chemical or environmental stress condition (Hillenmeyer et al., 2008). Here, we use the extraordinary

heterogeneity of hundreds of cancer cell lines as a model for diverse cellular stress contexts, demonstrating that the contexts in which stress response genes become essential reflects the underlying biology of those factors. We exploit these context-specific essentiality patterns of known stress response genes to build a network of genes having similar context-specific fitness effects, effectively recapitulating canonical regulatory relationships, detailing treatment-relevant tumor dependencies, and identifying novel genes relevant to a global stress response network. Because an imbalance between stressors and the cellular capacity to adapt to those stressors underlies a broad spectrum of human disease, the data uncovered using this approach has broad therapeutic and prognostic implications.

The principle that convergent or epistatic knockout phenotypes may identify functional relationships between genes is not new (Dobzhansky, 1946). Great insights have been gained from pairwise genetic screens in model organisms, most prominently budding yeast (Costanzo et al., 2016), but the larger human genome has proved challenging for genome-wide combinatorial study of genetic perturbations. With the advent of high quality CRISPR-Cas9 screening libraries in recent years, the coessentiality approach has emerged as an alternative to pairwise genetic perturbations for the discovery of novel genes and genetic interactions (Boyle et al., 2018; Kim et al., 2019; Pan et al., 2018). These first coessentiality studies detail a top-down approach which effectively resolves major protein complexes and identifies functional clusters in the genome. Our locus-adjusted, rank-based approach addresses certain limitations of these studies, which include bias from copy number variable regions, overrepresentation of obligate protein complexes, and limited capacity to resolve genes which have dynamic

functions across many cellular states – such as the transcription factors and chaperones of the cellular stress response. For example, HSF1 drives distinct transcriptional programs depending on cellular context (Filone et al., 2014; Mendillo et al., 2012; Scherz-Shouval et al., 2014), and both the chaperone (Joshi et al., 2018; Rizzolo et al., 2017; Rodina et al., 2016) and DNA damage response networks (Bandyopadhyay et al., 2010) are known to rewire under stress conditions.

In each of the networks elucidated in this study, canonical signaling relationships were recapitulated and modules of functionally related genes emerged. For example, within the chaperone network, we observe clusters distinguished both by subcellular localization and by patterns of induction in response to mechanistically distinct proteotoxic stresses. Of note, many chaperone and co-chaperone genes still have unresolved functional partners and localization, particularly within the HSP40/JDP family (Kampinga et al., 2019). Our network corroborates recent reports revealing the localization of individual JDPs, such as DNAJB11 in the ER lumen (Chen et al., 2017) and DNAJC11 at the inner mitochondrial membrane (Ioakeimidis et al., 2014). Thus, the novel relationships observed in our network, such as DNAJC24 at the inner mitochondrial membrane, are valuable targets for future study. Beyond chaperones, in the global stress network, we identify not only genes corresponding to individual stress pathways but also the genes which connect distinct stress programs. These genes, such as C16orf72/HAPSTR and NCL (bridging cytosolic proteostasis and DNA damage) and TSC1/2 and TXNDC17 (bridging oxidative stress and nutrient stress) represent points of crosstalk which likely play critical roles in diseased cells facing combinatorial stresses.

An overarching theme in these data is that cancer cells represent a breadth of disease-relevant cellular states which may offer functional insight not obtained from studies using limited numbers of cell lines or stress perturbations without a direct physiological correlate. For example, we identify a small subset (~1%) of cancer cell lines particularly sensitive to the loss of genes involved in the ERN1-XBP1 branch of the UPR. Remarkably, these cell lines comprise multiple myeloma, a malignancy derived from plasma cells, which are characterized by a dramatically expanded ER network to accommodate the secretion of thousands of immunoglobulin molecules each second (Calame et al., 2003). Myeloma cells are burdened by a specific stress encountered in many disorders of secretory cells (e.g. pancreatic β cells in hereditary diabetes mellitus), and may thus serve as a more disease-relevant ER stress model than tool compounds such as thapsigargin (which induces ER stress by depleting ER calcium stores). Indeed, while all three branches of the UPR critically modulate the response to thapsigargin (Adamson et al., 2016), the ERN1-XBP1 branch is selectively required in these hypersecretory cells. In the clinical care of multiple myeloma patients, proteasome inhibitors are already employed to exploit the ER stress phenotype of malignant plasma cells. However, as the proteasome is a core essential component of all cells, our data suggest direct targeting of ERN1-XBP1 – for example, with available ERN1 inhibitors – may retain therapeutic efficacy while minimally affecting non-tumor cells.

Importantly, our data also indicate that there exists a spectrum of individual stress subnetwork utilization, even within a given lineage. Returning to the example of ERN1-XBP1, while this subnetwork was generally more essential in multiple myeloma, not all myeloma lines were equally impacted by ERN1-XBP1 loss. Thus, while targeting ERN1-

XBP1 may have utility in the treatment of multiple myeloma, our data indicate that some tumors will be more susceptible than others to this modulation. Moreover, we find that the myeloma subset less likely to respond UPR modulation has a multiomic phenotype of increased oxidative metabolism and is, in turn, more susceptible to Elesclomol. Similarly, in AML, we observe variable dependence on the translation-HSF1 link, identify a transcriptional signature reflective of this phenotype, and demonstrate that this signature stratifies AML patients in a meaningful fashion. It is worth noting that this approach to identify multiomic signatures of subnetwork dependency is unlikely to provide meaningful insight in pan-cancer analyses due to the powerful confounding effects of cell lineage. Thus, these data suggest a paradigm by which intra-lineage dependence on individual genes or gene networks can be translated into potentially actionable insight about an individual patient's tumor.

Finally, we demonstrate the potential of our approach to identify novel genes, such as C16orf72/HAPSTR, relevant to a complex biological network. HAPSTR's coessentiality network successfully predicted its cooperative relationship with HUWE1, as well as its role in the response to genotoxic, nutrient, proteotoxic, and oxidative stress. Because HAPSTR impacts the response to diverse stress states, promotes cancer cell growth, and is prognostic of patient outcomes in diverse cancer diagnoses, HAPSTR may be a promising therapeutic target in cancer. Additionally, it is worth noting that HAPSTR is expressed at high levels in the developing brain (Figure S7B), HAPSTR copy number alterations have been associated with autism and schizophrenia (Levinson et al., 2011; Sanders et al., 2011), and mutations of its functional cofactor HUWE1 are associated with

multiple neurodevelopmental conditions (Bosshard et al., 2017). Thus, HAPSTR misregulation may also play a role in human pathology beyond neoplasia.

Far beyond HAPSTR and the response to stress, we emphasize that the approach detailed here is easily adapted to facilitate targeted study of other genes and pathways. To support such efforts, we make available an interactive web application (http://fireworks.mendillolab.org/) where individuals may input gene(s) of interest and quickly visualize a locus-adjusted coessentiality network. Moreover, this tool facilitates the integration of multiomic data to define the mechanisms underlying context-specific relationships. It is our hope that this resource will broadly augment efforts to identify functional interactions within biological networks.

METHODS

Curation of gene essentiality data in cancer cell lines

Data from CRISPR-Cas9 genome scale loss of function screening of 625 cancer cell lines was obtained from DepMap (https://depmap.org/portal/download/). CERES scores were used to quantify the fitness effect of individual gene loss, with "essentiality" in this paper represented as the inverse CERES score (i.e. more positive scores for genes which cause greater dropout of cells with guides targeting that gene. The screen performed in Drosophila melanogaster was identified through a search of BioORCs for any screens performed in Drosophila or C. elegans which targeted that species' C16orf72 homolog.

Integration and analysis of Cancer Cell Line Encyclopedia Multiomic Data

Processed RNA-seq, reverse phase protein array, copy number, and metabolomic data were obtained from the DepMap data portal (https://depmap.org/portal/download/). These data are described in detail in (Ghandi et al., 2019). For descriptive comparisons of different cell lines stratified by dependency signatures, the cell lines with 75th percentile or higher dependency on that signature were compared with cell lines having 25th percentile of lower dependency.

Determination of coessential genes

To determine a locus-corrected coessentiality value for each gene pair, gene essentiality scores for each gene were subtracted from the median essentiality score for that gene's nearest 10 neighbor genes (5 downstream, 5 upstream) on the same chromosome. Strength of coessential relationships are represented as Pearson r, with coessentiality rank used for all network analyses to mitigate bias from p-value inflation or differing numbers of cell lines in cancer cell line subsets.

Coessentiality network visualization and clustering

Rank-based networks were constructed from a single or set of input genes, using a soft rank threshold for each analysis – i.e. correlations below the specific rank were not included. Edges are not weighted by correlation strength or rank. To remove potentially spurious correlations from genes only expressed in certain lineages, networks excluded transcripts where the median expression across cell lines was equal to 0. All networks were visualized in Cytoscape v3.7.2 (https://cytoscape.org/). Unless otherwise specified,

networks used a force-directed layout with modest adjustments made by hand to improve legibility. The circular layout for the stress network was manually arranged following the groupings obtained by hierarchical clustering of source nodes.

Data analysis and visualization

Unless otherwise specified, data were analyzed with Python (version 3.6.4, Anaconda Inc.) using the modules Pandas (v0.23.4) and Numpy (v1.14.2) and visualization used the modules Matplotlib (v2.2.2) and Seaborn (v0.9.0). For heatmaps visualizing gene essentiality patterns across cell lines and cancer subsets, essentiality scores for a given gene are standard scaled to the minimum and maximum essentiality of that gene across all cancer cells. P-values were considered significant at an alpha of 0.05 or lower, as specified for each analysis. Benjamini-Hochberg false discovery rate correction was performed where indicated to adjust for multiple comparisons (Benjamini and Hochberg, 1995).

Gene Set Enrichment Analysis

GSEA was performed using the Molecular Signature DataBase as accessible at http://software.broadinstitute.org/gsea/msigdb/annotate.jsp (Liberzon et al., 2015; Subramanian et al., 2005). The gene sets queried were as follows: hallmark (H), positional (C1), KEGG pathways (C2), REACTOME (C2), GO Biological Process (C5), and GO Molecular Function (C5).

Protein conservation analysis

C16orf72 protein sequences for multiple species were obtained from ENSEMBL (http://useast.ensembl.org/index.html). Multi-sequence alignment was performed using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/), with visualization of the alignment in Jalview v2.11 (https://www.jalview.org/).

Analysis of Patient Survival and Drug Response

For gene expression-stratified survival analysis, hazard ratios and log-rank test p-values were obtained through KM-plotter (https://kmplot.com/analysis/), using standard settings and the pan-cancer TCGA RNA-seq transcription dataset (Nagy et al., 2018). For AML analyses, we employed BloodSpot (http://servers.binf.ku.dk/bloodspot/) (Bagger et al., 2019). Bortezomib (Velcade) drug response likelihood was assessed using pre-treatment microarray data of multiple myeloma patients (GEO: GSE9782) (Mulligan et al., 2007).

Functional classification and annotation of proteins

Grouping of proteins into functional classes was performed using PANTHER (pantherdb.org) to analyze a list of all genes targeted in the AVANA sgRNA library (Mi et al., 2019). NCBI Gene References Into Function (GeneRIFs; data downloaded from ftp://ftp.ncbi.nih.gov/gene/GeneRIF/), manually annotated blurbs summarizing findings in individual papers, were queried to bin the stress network genes by degree of prior existing knowledge.

Experimental model and subject details

HEK293T and HeLa cells were grown in DMEM media supplemented with 10% FBS and 1% pen/strep. MDA-MB-231, ZR-75, and SK-BR3 cells were grown in RPMI media supplemented with 10% FBS and 1% pen/strep. Doxycycline inducible FLAG-Ub expressing HEK-293T cells were generated by transducing lentivirus containing pCW57.1-FLAG-Ub and selected in 2 μg/ml puromycin. All cells were passaged with Accumax unless otherwise specified.

Plasmids, Lentivirus Generation and Infection

FLAG tagged ubiquitin was cloned in pCW57.1 vector (Addgene # 41393) between BsrG1 sites and verified through sequencing. PLenti6 vectors containing DNA constructs of interest were co-transfected with pMD2.G and psPAX2 into 293T cells using Lipofectamine 3000. After 48 hours, media was removed, filtered with an 0.45 μm filter, and centrifuged at 21000xg for 10 min to yield packaged lentivirus in the supernatant. Lentivirus was then added directly to cells for transduction. The C16orf72-FLAG overexpression vector contained a blasticidin resistance cassette, and stable overexpressing cell lines were selected for in blasticidin at 10 ug/mL (293T) or 20 ug/mL (MFC7, MDA-MB-231) for 5 days.

Gene silencing and co-transfections

Smart-pool siRNAs were obtained for each target gene of interest, as well as a non-targeting sequences, and transfected using RNAimax (Thermo Fisher) using a standard protocol. Unless otherwise specified, cells were harvested 72h after siRNA

transfection. Knockdown was confirmed for each siRNA experiment by qPCR or immunoblot.

Drug sensitivity in C16orf72-knockdown cells

MDA-MB-231, SKBR3, and ZR75 cells were reverse transfected with C16orf72 or non-targeting siRNA at 1000 cells/per well in a 384-well plate. After 1 day, RPMI media was replaced and drugs were arrayed at half-log intervals into wells using a Tecan D300E drug printer. Drugs were selected based on targeting of pathways identified in the C16orf72 coessentiality network. Drug concentrations were selected based on a pilot experiment and prior literature to avoid complete cytotoxicity in the wild-type cell background.

Proliferation and viability assays

Proliferation was assessed using live-cell images obtained via Incucyte. Cells were grown in 384-well plates, in triplicate for each condition, with whole-well images being taken every 2 hours. Area under the proliferation curve represented endpoint confluency subtracted from initial confluency for each well. Viability was assessed using an adapted CellTiterGlo (Promega) protocol; briefly, 384-well plates were brought to room temperature for 15 min, 12 uL of CellTiterGlo reagent mix was added to each well containing 50 uL of media, and the plate was agitated for 2 min. Luminescence was read out on a Tecan infinite M1000 pro platereader.

Immunoblot

Protein samples were lysed in RIPA buffer or ubiquitin lysis buffer (as specified below) containing 1mM PMSF and a Roche Protease Inhibitor Cocktail tablet and passed through a 21 gauge syringe 15 times per sample. Protein concentration was assessed by standard BCA assay (Pierce, #23255), denatured in 4X Laemmli sample buffer containing beta-mercaptoethanol, and heated at 95C for 10 min. Electrophoresis was performed using 4-20% Bis-Tris gradient gels unless otherwise specified, with transfer to PVDF membranes using a 7-minute protocol on an iBlot machine. Membranes were blocked for 1 hour at room temperature in 5% fat-free milk. Primary and secondary antibodies were diluted in 5% fat free milk and exposed to membranes overnight at 4C and for 1 hour at room temperature, respectively. Imaging was performed with the BioRad ChemiDoc Touch Imaging System (732BR0783) after incubation for 2 min in HRP substrate (Immobilon, Millipore). Blots were analyzed using ImageLab v6.0.1 (BioRad).

Immunofluorescence

Cells were grown on poly-D-lysine treated sterile coverslips in a 24 well-plate. Steps were performed at room temperature unless otherwise specified. Cells were washed three times with cold PBS, fixed with 4% paraformaldehyde for 10 min, and permeabilized with 0.2% Triton X100 for 5 min. Blocking encompassed incubation in 2% FBS for 30 min. Primary antibodies were diluted at 1:500 and secondary antibodies were diluted at 1:1000 in 2% FBS and exposed to cells for 1 hour each, with 3 PBST washes between. Coverslips were mounted to a slide using a DAPI/mounting mixture and allowed to dry overnight before imaging.

Microscopy and Image Analysis

Images were acquired at 63x magnification using a Zeiss LSM800 confocal microscope. A z-stack slicing distance of 0.7uM was used, with final images visualized as an orthogonal projection of the maximal value per pixel across stacks. No non-linear adjustments were performed. ImageJ/Fiji (https://imagej.net/Fiji#Downloads) was used for all microscopy analyses.

Immunoprecipitation

Cells were rinsed twice with cold PBS, removed from plates by scraping, and centrifuged for 4 min at 1000g and 4° C before pellet resuspension in cold lysis buffer. For immunoprecipitations of ubiquitin or of targets for downstream ubiquitination analysis, cells were vortexed and passed through a syringe in ubiquitin lysis buffer (2% SDS, 150mM NaCl, 10 mM Tris HCl pH 8, 1 Roche protease inhibitor tablet, 5mM N-ethylmaleide, and 1mM PMSF) and heated at 95° C for 10 min before dilution in ubiquitin lysis buffer with 1% triton and no SDS, resulting in a final SDS concentration of 0.2%. For all other immunoprecipitations, the lysis buffer was 1% NP40, 100 mM NaCl, 50 mM Tris pH 7.5, 0.2 mM EDTA, 5% glycerol, and 1mM PSMF and lysis was achieved by sonication in a 4C water bath (10 cycles of 30 sec on, 1 min off). After lysis, cells were spun at 21000xg and 4C for 10 minutes and the supernatant kept for input and immunoprecipitation. FLAG-immunoprecipitations were performed using M2 affinity agarose (Thermo Fisher).

Mass Spectrometry

A protein gel band was submitted to the Northwestern University Proteomics Core Facility for an in-gel digestion. Peptides were analyzed by LC-MS/MS using a Dionex UltiMate 3000 Rapid Separation nanoLC coupled to a Orbitrap Elite Mass Spectrometer (Thermo Fisher Scientific Inc, San Jose, CA). Samples were loaded onto the trap column, which was 150 µm x 3 cm in-house packed with 3 um ReproSil-Pur® beads. The analytical column was a 75 um x 10.5 cm PicoChip column packed with 3 um ReproSil-Pur® beads (New Objective, Inc. Woburn, MA). The flow rate was kept at 300nL/min. Solvent A was 0.1% FA in water and Solvent B was 0.1% FA in ACN. The peptide was separated on a 120-min analytical gradient from 5% ACN/0.1% FA to 40% ACN/0.1% FA. MS^1 scans were acquired from 400-2000m/z at 60,000 resolving power and automatic gain control (AGC) set to $1x10^6$. The 15 most abundant precursor ions in each MS^1 scan were selected for fragmentation by collision-induced dissociation (CID) at 35% normalized collision energy in the ion trap. Previously selected ions were dynamically excluded from re-selection for 60 seconds.

Proteins were identified from the MS raw files using Mascot search engine (Matrix Science, London, UK; version 2.5.1). MS/MS spectra were searched against the UniProt Human database (SwissProt 2019, 20303 entries). All searches included carbamidomethyl cysteine as a fixed modification and oxidized Met, deamidated Asn and Gln, acetylated N-term as variable modifications. Three missed tryptic cleavages were allowed. The MS^1 precursor mass tolerance was set to 10 ppm and the MS^2 tolerance was set to 0.6 Da. The search result was visualized by Scaffold (version 4.9.0. Proteome Software, INC., Portland, OR). Peptide identifications were accepted if they could be established at greater than 90.0% probability by the Peptide Prophet algorithm (Keller et

al., 2002) with Scaffold delta-mass correction. Protein identifications were accepted if they could be established at greater than 99.0% probability and contained at least 1 identified peptide. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii et al., 2003). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

RNA-sequencing

RNA was extracted using a Qiagen RNeasy kit; briefly, cells were lysed in buffer RLT, nucleic acids were precipitated with ethanol and applied to columns, columns were treated with DNase, and RNA was eluted after washing/cleaning. Libraries were prepped using a QuantSeq 3′ mRNA-Seq Library Prep Kit FWD for Illumina (Lexogen) using 100ng of input RNA in an automated protocol adapted for the SciClone. Libraries were then analyzed for quality using the Agilent High Sensitivity DNA kit and for quantity using Qubit dsDNA HS assay, in 384-well format, using 20µL reactions in triplicate (19µL working reagent + 1µL sample or standard). For the Qubit assay, 11 standards were prepared from either 0-3ng/µL or 0-10ng/µL depending on BioAnalyzer concentrations. On plate reader, shake for 5 seconds, then read fluorescence: excitation: 480nm, emission: 530nm. Excitation/emission bandwidth: 5nm, settle time: 100ms. Sample concentrations were determined using the standard curve. Libraries were then pooled and sequenced using a NovaSeq 6000 SP Reagent Kit (100 cycles). Libraries were pooled at 25nM each, denatured with 1M NaOH added to a 0.2M final concentration (5 min at room temperature), and quenched with 200mM Tris HCl (pH 7). 1% PhyX spike-in (Illumina)

was included. Pooled, denatured libraries were run on an Illumina NovaSeq using 51bp reads, 6bp index reads, and paired-end single read parameters.

Data and Code availability

All code and data for these analyses will be deposited in public repositories. Our approach is implemented in in an interactive web application (http://fireworks.mendillolab.org/).

Acknowledgements: We thank members of the Mendillo laboratory and Michael Drumm for comments on the manuscript. We thank Clara B. Peek for helpful discussions. We thank Elizabeth Bartom for developing the Ceto pipeline used for RNA-Seq analysis. D.R.A. was supported by the NIH (5T32GM008152-33). M.L.M is supported by the National Cancer Institute of the NIH (R00CA175293) and the Susan G. Komen Foundation (CCR17488145). M.L.M was also supported by Kimmel Scholar (SKF-16-135) and Lynn Sage Scholar awards. Proteomics services were performed by the Northwestern Proteomics Core Facility, generously supported by NCI CCSG P30 CA060553 awarded to the Robert H Lurie Comprehensive Cancer Center, instrumentation award (S10OD025194) from NIH Office of Director, and the National Resource for Translational and Developmental Proteomics supported by P41 GM108569. D.R.F. was supported by the NIH (R01 GM111907).

Author Contributions: Conceptualization: DRA and MLM. Data curation: DRA, JJ, BKC, YAG, NLK, MLM. Formal analysis: DRA, BKC. Funding acquisition: MLM. Investigation, DRA, KAM, DJA, RSS, SB, ST, BPO, BC, MLM. Methodology: DRA, JJ, KAM, SS, BKC, YAG, NLK, DRF, MLM. Project administration: MLM. Resources: MLM, NLK, SS, DRF. Software: DRA, JJ. Supervision: MLM. Validation: DRA, JJ, KAM, DJA. Visualization: DRA. Writing – original draft: DRA and MLM. Writing – review and editing: all authors.

FIGURE LEGENDS

Figure 1: Developing a coessentiality network for the master regulator of cytosolic proteostasis. A. Schematic of overall approach. B. HSF1 loss is deleterious to cell fitness across lineages and has an essentiality spectrum within lineages. Essentiality scores are defined as the inverse of the CERES score for sgRNAs targeting that gene; one is typical for a core fitness-essential gene and zero represents no fitness effect. C. HSF1 copy number is associated with HSF1 transcript level. D. The magnitude of HSF1 essentiality is not predicted by HSF1 copy number. E. The genes with the most similar fitness profile to HSF1 are highly enriched for genes in the same locus (chr8q). This bias is corrected by subtracting the median essentiality of 10 neighbor genes prior to genome-scale correlations. Subtracting more or less genes resulted in diminished specificity of this correction. F. Visualization of the HSF1-coessentiality network, where columns are cell lines, rows are genes, and hue corresponds to row-normalized gene essentiality. Also shown are enriched gene sets (FDR < 0.0001). G. Network visualization of the top 100 HSF1-correlated genes (primary nodes) plus the top 5 coessential genes for each primary node (i.e. secondary edges/nodes). Nodes are visualized here if they have more than one connection in the network. H. Cancer subtype-specific coessentiality patterns for modules enriched in the pan-cancer HSF1 network, including a conserved heat shock correlation in all subtypes but acute myeloid leukemia (AML). I. Multiomic data integration to characterize the high-HSF1-dependence state in AML lines. J. HSF1 essentiality is highly correlated with the transcriptional signature of translation/protein synthesis genes in AML but not most other cancers. K. The translation-dominant multiomic signature associated

with HSF1-dependence in AML stratifies AML patient prognosis, as depicted by the top statistically-associated gene, EIF3L.

Figure 2: A chaperone coessentiality analysis reveals functionally and spatially differentiated networks. A. Essentiality scores and transcript expression of heat shock factors and major chaperone family members across 625 cancer cell lines. B. Rank 15 positive correlation coessentiality network using all chaperone and heat shock factor genes in (A). Nodes are visualized if they have connections to more than one gene in the network (internal nodes). C. Selected Markov clusters from the network which represent biological functional groups corresponding to gene essentiality, localization, and/or divergent proteotoxic stress responses. RNAseq datasets derive from MDA-MB-231 cells treated with either 2-deoxyglusose (2DG) to induce ER stress or heat shock (HS; 42C x 1hr) to induce cytosolic proteotoxicity. Enrichment **$p < 0.01$, ***$p<0.001$, Kolmogorov–Smirnov test.

Figure 3: A global stress response coessentiality network. A. Essentiality scores and transcript expression of stress response master regulators across 625 cancer cell lines. B. Visualization of coessential gene relationships within stress response modules and between stress response modules at rank 30 for both positive and negative correlations. Stress networks thresholded at other ranks (10, 50, 100) are detailed in Figure S5. C. The stress network has greater connectivity (edge density ratio) than expected from 10,000 random network permutations using the same parameters. D. PANTHER functional classes of the 125 stress network genes pulled in as a function of coessentiality with source nodes; * indicates $P<0.005$, Fisher's exact test. E. Representation of the degree to which stress network genes have been previously studied, using NCBI Gene

References Into Function (GeneRIFs) as a surrogate for existing knowledge on a given gene. F. Transcriptome profiling by RNAseq of the stress network genes in MDA-MB-231 cells after treatment with Cyclophosphamide (CPA; genotoxic stressor), CoCl2 (hypoxia mimetic and oxidative stressor), 2-deoxyglucose (2DG; ER and nutrient stressor), heat shock (HS; 42C x 1hr; cytosolic proteotoxic stressor), or serum/amino acid starvation (SS; nutrient stressor). Selected genes elaborated in the text are labeled.

Figure 4: Systematic profiling of lineage-specific coessentiality patterns in the stress response. A. Lineage-specific subnetworks were created and profiled for relative enrichment of each possible source node pair subnetwork, with enrichment defined as membership relative to subnetwork membership across all other lineages and visualized as shares of a pie chart. B. The skin coessentiality network features the largest TP53-ATM genotoxicity network across cancer subsets. C. The autonomic ganglia (neuroblastoma) network contains a highly anticorrelated network between ATF3 and EIF2AK3 which was not present in other lineages. D. The hematopoietic lineage features a striking ERN1-XBP1 subnetwork, but no relationships between other branches of the UPR (ATF6, EIF2AK3/PERK). E. Multiple myeloma drives the correlation of ERN1 and XBP1 in hematopoietic tumors. F. Among all possible combinations, the ERN1-XBP1 network in myeloma is the most selectively essential as well as tightly correlated. Other hits for enriched networks include ARNT-EPAS1 (hypoxia) in renal cancers. G. The myeloma lines most dependent on XBP1 and ERN1 have a low OXPHOS multiomic signature. I. The high-ERN1-XBP1-dependent myeloma lines are more susceptible to the proteasome inhibitor Bortezomib and substantially less sensitive to Elesclomol, which targets mitochondrial metabolism in myeloma lines resistant to proteasome inhibition

(Tsvetkov et al., 2019). J. Patients enrolled in a clinical trial of Bortezomib (Mulligan et al., 2007) were more likely to respond if their tumors had >1 SD underexpression of the OXPHOS transcriptomic signature from (G). P values determined by Mann-Whitney U test.

Figure 5: C16orf72 is a conserved and stress-induced modulator of diverse stress responses. A. C16orf72 expression is prognostic in many tumors. Subtypes shown had log-rank P value less than 0.05. B. C16orf72 loss is deleterious to the fitness of cancer cell lines across diverse backgrounds, particularly kidney cancer lines. C. C16orf72 is a 275aa protein conserved through C. elegans. Region shown represents a conserved nuclear localization signal elaborated in Figure S6. D. A CRISPR screen in Drosophila cells (Viswanatha et al., 2018) indicates conservation of C16orf72 importance to cellular proliferation. E. Validation of the Project Achilles fitness screen data suggesting that SKBR3 cell proliferation is highly dependent on C16orf72, ZR-75-1 cells are moderately dependent on C16orf72, and MDA-MB-231 cells are minimally affected by C16orf72 loss. F. Transcriptomic profiling of C16orf72 knockdown in these three cell lines identifies conserved regulation of pathways including multiple stress response networks. G. The C16orf72 coessentiality network demonstrates enrichment for ubiquitination pathways and multiple stress response networks. H. C16orf72 knockdown greatly hampers the ability of SKBR3, ZR75, and MDA-MB-231 cells to adapt to diverse stressors. I, J, K: Live cell imaging and RNA-seq of the three conditions with had the greatest combined effect on fitness in the background of C16orf72 depletion in MDA-MB-231 cells.

Figure 6: C16orf72 functions as a cooperative binding partner with the E3 ligase HUWE1. A. Expression of C16orf72-FLAG in HEK293T cells revealed a predominantly

nuclear localization. B. C16orf72-FLAG affinity purification coupled with tandem mass spectrometry analysis identifies many binding partners for C16orf72, most prominently, the Hect E3 ligase HUWE1. Spectral counts were first normalized to the sum of spectral counts for the sample, then protein length, and finally compared vs. the maximum value for that protein from three separate control (GFP) IP-MS experiments. C. Immunoblot validation of HUWE1 IP using C16orf72-FLAG as bait. D. Co-IP of endogenous HUWE1 (bait) identifies C16orf72-FLAG. E. HUWE1 is the top-ranked gene in the C16orf72 coessentiality network. F. Many other proteins which interacted with C16orf72 were also coessential with C16orf72. G. C16orf72 protein levels and relative ubiquitination levels do not change with HUWE1 depletion, indicating that C16orf72 is not a ubiquitination substrate of HUWE1. H. Transcription profiling reveals that depletion of C16orf72, HUWE1, or both genes in MDA-MB-231 cells has a striking phenotypic overlap. I. The ubiquitinated proteome, as measured by FLAG-ubiquitin affinity purification coupled with mass spectrometry, is altered highly similarly by loss of either C16orf72 or HUWE1. J. C16orf72 and HUWE1 depletion converge on the under-ubiquitination of many proteins, such as GNL3/nucleostemin, and the over-ubiquitination of one protein, UBE2L3, which serves as an E2 for HUWE1. K. Representative immunoblots of established HUWE1 target protein levels after C16orf72 or HUWE1 siRNA treatment for 3 days. L. C16orf72 and HUWE1 depletion increase baseline levels of GNL3/nucleostemin and reduce the capacity to eliminate the protein under stress (serum starvation).

Figure S1: A top-down, genome-scale coessentiality network analysis does not resolve signal for major stress response genes. A. Lineage/subtype representation of the 625 cancer cell lines screened by the AVANA sgRNA library and included in these

analyses. B. Visualization of the most coessential genes reveals dominant clusters corresponding to major molecular machines, as well as clusters corresponding to a known bias of assigning gene essentiality in CRISPR screens: copy number altered regions. C. Representation of proteins which form physical interaction complexes is substantially higher than would be expected in the genes which are most coessential across the genome D. The top fitness correlation for a generic transcription factor of chaperone is less than that of a protein complex member, suggesting reduced representation of these genes in any non-rank-based, top-down coessentiality approach. E. Of the 21 stress response genes studied using coessential gene networks in this paper, only ATM and P53 have a functional module identified in a top-down genome scale clustering approach.

Figure S2: Cytosolic HSP90 isoform essentiality is predicted by HSP90 isoform expression in a manner related to cell lineage. A. HSP90AA1 and HSP90AB1 are not coessential, despite physically interacting in a major protein complex. B. Cytosolic HSP90 isoform expression predicts relative essentiality of that isoform. C. Certain tumor subtypes have higher expression and dependence on a particular cytosolic HSP90 isoform.

Figure S3: Alternative rank threshold stress networks. Rank 10, 30 (as shown in Figure 3, but with hierarchical layout), 50, and 100 stress network organization using a force-directed layout.

Figure S4: ATF3 binds to an EIF2AK3/PERK enhancer. ENCODE ChIP-seq data were queried for EIF2AK3, revealing that ATF3 binds strongly to a regulatory element of EIF2AK3 in K562 cells.

Figure S5: C16orf72 is expressed across broad cell types in adult tissues and tumors, as well as throughout human development. A. GEPIA output detailing tumor vs. normal transcriptional profiling of C16orf72; significant differences represented by red text. B. Expression of C16orf72 across human development; data from (Cardoso-Moreira et al., 2019).

Figure S6: C16orf72's nuclear localization is driven by an N-terminal bipartite NLS. A. Querying C16orf72's amino acid sequence with cNLS mapper (http://nls-mapper.iab.keio.ac.jp/) revealed a predicted bipartite NLS at position 252-275. Score filter was set at the most stringest threshold, 7. B. Deletion of the putative NLS is sufficient to establish a cytosolic/perinuclear localization of C16orf72 in 293T cells as compared with its baseline nuclear localization (Figure 6A)

REFERENCES

Adamson, B., Norman, T.M., Jost, M., Cho, M.Y., Nunez, J.K., Chen, Y., Villalta, J.E., Gilbert, L.A., Horlbeck, M.A., Hein, M.Y.*, et al.* (2016). A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. Cell *167*, 1867-1882 e1821.

Aguirre, A.J., Meyers, R.M., Weir, B.A., Vazquez, F., Zhang, C.Z., Ben-David, U., Cook, A., Ha, G., Harrington, W.F., Doshi, M.B.*, et al.* (2016). Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting. Cancer Discov *6*, 914-929.

Bagger, F.O., Kinalis, S., and Rapin, N. (2019). BloodSpot: a database of healthy and malignant haematopoiesis updated with purified and single cell mRNA sequencing profiles. Nucleic Acids Res *47*, D881-D885.

Bandyopadhyay, S., Mehta, M., Kuo, D., Sung, M.K., Chuang, R., Jaehnig, E.J., Bodenmiller, B., Licon, K., Copeland, W., Shales, M.*, et al.* (2010). Rewiring of genetic networks in response to DNA damage. Science *330*, 1385-1389.

Ben-Sahra, I., Hoxhaj, G., Ricoult, S.J.H., Asara, J.M., and Manning, B.D. (2016). mTORC1 induces purine synthesis through control of the mitochondrial tetrahydrofolate cycle. Science *351*, 728-733.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Statists Soc *57*, 289-300.

Boos, F., Kramer, L., Groh, C., Jung, F., Haberkant, P., Stein, F., Wollweber, F., Gackstatter, A., Zoller, E., van der Laan, M.*, et al.* (2019). Mitochondrial protein-induced stress triggers a global adaptive transcriptional programme. Nat Cell Biol *21*, 442-451.

Bosshard, M., Aprigliano, R., Gattiker, C., Palibrk, V., Markkanen, E., Backe, P.H., Pellegrino, S., Raymond, F.L., Froyen, G., Altmeyer, M.*, et al.* (2017). Impaired oxidative stress response characterizes HUWE1-promoted X-linked intellectual disability. Sci Rep *7*, 15050.

Boyle, E.A., Pritchard, J.K., and Greenleaf, W.J. (2018). High-resolution mapping of cancer cell networks using co-functional interactions. Mol Syst Biol *14*, e8594.

Calame, K.L., Lin, K.I., and Tunyaplin, C. (2003). Regulatory mechanisms that determine the development and function of plasma cells. Annu Rev Immunol *21*, 205-230.

Cardoso-Moreira, M., Halbert, J., Valloton, D., Velten, B., Chen, C., Shao, Y., Liechti, A., Ascencao, K., Rummel, C., Ovchinnikova, S.*, et al.* (2019). Gene expression across mammalian organ development. Nature *571*, 505-509.

Chen, K.C., Qu, S., Chowdhury, S., Noxon, I.C., Schonhoft, J.D., Plate, L., Powers, E.T., Kelly, J.W., Lander, G.C., and Wiseman, R.L. (2017). The endoplasmic reticulum HSP40 co-chaperone ERdj3/DNAJB11 assembles and functions as a tetramer. EMBO J *36*, 2296-2309.

Choe, K.N., Nicolae, C.M., Constantin, D., Imamura Kawasawa, Y., Delgado-Diaz, M.R., De, S., Freire, R., Smits, V.A., and Moldovan, G.L. (2016). HUWE1 interacts with PCNA to alleviate replication stress. EMBO Rep *17*, 874-886.

Christofk, H.R., Vander Heiden, M.G., Harris, M.H., Ramanathan, A., Gerszten, R.E., Wei, R., Fleming, M.D., Schreiber, S.L., and Cantley, L.C. (2008). The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. Nature *452*, 230-233.

Costanzo, M., VanderSluis, B., Koch, E.N., Baryshnikova, A., Pons, C., Tan, G., Wang, W., Usaj, M., Hanchard, J., Lee, S.D.*, et al.* (2016). A global genetic interaction network maps a wiring diagram of cellular function. Science *353*.

Cunningham, T.A., Chapman, E., and Schatz, J.H. (2018). eIF4A inhibition: ready for primetime? Oncotarget *9*, 35515-35516.
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.
Daniely, Y., Dimitrova, D.D., and Borowiec, J.A. (2002). Stress-dependent nucleolin mobilization mediated by p53-nucleolin complex formation. Mol Cell Biol *22*, 6014-6022.
Dobzhansky, T. (1946). Genetics of natural populations; recombination and variability in populations of Drosophila pseudoobscura. Genetics *31*, 269-290.
Fidalgo, M., Guerrero, A., Fraile, M., Iglesias, C., Pombo, C.M., and Zalvide, J. (2012). Adaptor protein cerebral cavernous malformation 3 (CCM3) mediates phosphorylation of the cytoskeletal proteins ezrin/radixin/moesin by mammalian Ste20-4 to protect cells from oxidative stress. J Biol Chem *287*, 11556-11565.
Filone, C.M., Caballero, I.S., Dower, K., Mendillo, M.L., Cowley, G.S., Santagata, S., Rozelle, D.K., Yen, J., Rubins, K.H., Hacohen, N.*, et al.* (2014). The master regulator of the cellular stress response (HSF1) is critical for orthopoxvirus infection. PLoS Pathog *10*, e1003904.
Ghandi, M., Huang, F.W., Jane-Valbuena, J., Kryukov, G.V., Lo, C.C., McDonald, E.R., 3rd, Barretina, J., Gelfand, E.T., Bielski, C.M., Li, H.*, et al.* (2019). Next-generation characterization of the Cancer Cell Line Encyclopedia. Nature *569*, 503-508.
Ginisty, H., Amalric, F., and Bouvet, P. (1998). Nucleolin functions in the first step of ribosomal RNA processing. EMBO J *17*, 1476-1486.
Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, K.R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S.*, et al.* (2015). High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. Cell *163*, 1515-1526.
Hillenmeyer, M.E., Fung, E., Wildenhain, J., Pierce, S.E., Hoon, S., Lee, W., Proctor, M., St Onge, R.P., Tyers, M., Koller, D.*, et al.* (2008). The chemical genomic portrait of yeast: uncovering a phenotype for all genes. Science *320*, 362-365.
Ioakeimidis, F., Ott, C., Kozjak-Pavlovic, V., Violitzi, F., Rinotas, V., Makrinou, E., Eliopoulos, E., Fasseas, C., Kollias, G., and Douni, E. (2014). A splicing mutation in the novel mitochondrial protein DNAJC11 causes motor neuron pathology associated with cristae disorganization, and lymphoid abnormalities in mice. PLoS One *9*, e104237.
Jeong, W., Yoon, H.W., Lee, S.R., and Rhee, S.G. (2004). Identification and characterization of TRP14, a thioredoxin-related protein of 14 kDa. New insights into the specificity of thioredoxin function. J Biol Chem *279*, 3142-3150.
Jiang, H.Y., Wek, S.A., McGrath, B.C., Lu, D., Hai, T., Harding, H.P., Wang, X., Ron, D., Cavener, D.R., and Wek, R.C. (2004). Activating transcription factor 3 is integral to the eukaryotic initiation factor 2 kinase stress response. Mol Cell Biol *24*, 1365-1377.
Kampinga, H.H., Andreasson, C., Barducci, A., Cheetham, M.E., Cyr, D., Emanuelsson, C., Genevaux, P., Gestwicki, J.E., Goloubinoff, P., Huerta-Cepas, J.*, et al.* (2019). Function, evolution, and structure of J-domain proteins. Cell Stress Chaperones *24*, 7-15.
Kao, S.H., Wu, H.T., and Wu, K.J. (2018). Ubiquitination by HUWE1 in tumorigenesis and beyond. J Biomed Sci *25*, 67.
Keller, A., Nesvizhskii, A.I., Kolker, E., and Aebersold, R. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem *74*, 5383-5392.
Kim, E., Dede, M., Lenoir, W.F., Wang, G., Srinivasan, S., Colic, M., and Hart, T. (2019). A network of human functional gene interactions from knockout fitness screens in cancer cells. Life Sci Alliance *2*.
Levinson, D.F., Duan, J., Oh, S., Wang, K., Sanders, A.R., Shi, J., Zhang, N., Mowry, B.J., Olincy, A., Amin, F.*, et al.* (2011). Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. Am J Psychiatry *168*, 302-316.

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 Syst *1*, 417-425.
Lindstrom, R., Lindholm, P., Kallijarvi, J., Palgi, M., Saarma, M., and Heino, T.I. (2016). Exploring the Conserved Role of MANF in the Unfolded Protein Response in Drosophila melanogaster. PLoS One *11*, e0151550.
Luo, J., Solimini, N.L., and Elledge, S.J. (2009). Principles of cancer therapy: oncogene and non-oncogene addiction. Cell *136*, 823-837.
Lynch, J.M., Maillet, M., Vanhoutte, D., Schloemer, A., Sargent, M.A., Blair, N.S., Lynch, K.A., Okada, T., Aronow, B.J., Osinska, H.*, et al.* (2012). A thrombospondin-dependent pathway for a protective ER stress response. Cell *149*, 1257-1268.
Maghames, C.M., Lobato-Gil, S., Perrin, A., Trauchessec, H., Rodriguez, M.S., Urbach, S., Marin, P., and Xirodimas, D.P. (2018). NEDDylation promotes nuclear protein aggregation and protects the Ubiquitin Proteasome System upon proteotoxic stress. Nat Commun *9*, 4376.
Maxwell, P.H., Wiesener, M.S., Chang, G.W., Clifford, S.C., Vaux, E.C., Cockman, M.E., Wykoff, C.C., Pugh, C.W., Maher, E.R., and Ratcliffe, P.J. (1999). The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. Nature *399*, 271-275.
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.
Meyers, R.M., Bryan, J.G., McFarland, J.M., Weir, B.A., Sizemore, A.E., Xu, H., Dharia, N.V., Montgomery, P.G., Cowley, G.S., Pantel, S.*, et al.* (2017). Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. Nat Genet *49*, 1779-1784.
Mi, H., Muruganujan, A., Ebert, D., Huang, X., and Thomas, P.D. (2019). PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res *47*, D419-D426.
Mulligan, G., Mitsiades, C., Bryant, B., Zhan, F., Chng, W.J., Roels, S., Koenig, E., Fergus, A., Huang, Y., Richardson, P.*, et al.* (2007). Gene expression profiling and correlation with outcome in clinical trials of the proteasome inhibitor bortezomib. Blood *109*, 3177-3188.
Nagy, A., Lanczky, A., Menyhart, O., and Gyorffy, B. (2018). Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. Sci Rep *8*, 9227.
Nesvizhskii, A.I., Keller, A., Kolker, E., and Aebersold, R. (2003). A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem *75*, 4646-4658.
Obeng, E.A., Carlson, L.M., Gutman, D.M., Harrington, W.J., Jr., Lee, K.P., and Boise, L.H. (2006). Proteasome inhibitors induce a terminal unfolded protein response in multiple myeloma cells. Blood *107*, 4907-4916.
Pan, J., Meyers, R.M., Michel, B.C., Mashtalir, N., Sizemore, A.E., Wells, J.N., Cassel, S.H., Vazquez, F., Weir, B.A., Hahn, W.C.*, et al.* (2018). Interrogation of Mammalian Protein Complex Structure, Function, and Membership Using Genome-Scale Fitness Screens. Cell Syst *6*, 555-568 e557.
Sanders, S.J., Ercan-Sencicek, A.G., Hus, V., Luo, R., Murtha, M.T., Moreno-De-Luca, D., Chu, S.H., Moreau, M.P., Gupta, A.R., Thomson, S.A.*, et al.* (2011). Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron *70*, 863-885.
Santagata, S., Mendillo, M.L., Tang, Y.C., Subramanian, A., Perley, C.C., Roche, S.P., Wong, B., Narayan, R., Kwon, H., Koeva, M.*, et al.* (2013). Tight coordination of protein translation and HSF1 activation supports the anabolic malignant state. Science *341*, 1238303.
Scherz-Shouval, R., Santagata, S., Mendillo, M.L., Sholl, L.M., Ben-Aharon, I., Beck, A.H., Dias-Santagata, D., Koeva, M., Stemmer, S.M., Whitesell, L.*, et al.* (2014). The reprogramming of tumor stroma by HSF1 is a potent enabler of malignancy. Cell *158*, 564-578.

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.*, et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545-15550.
Tan, W.X., Xu, T.M., Zhou, Z.L., Lv, X.J., Liu, J., Zhang, W.J., and Cui, M.H. (2019). TRP14 promotes resistance to cisplatin by inducing autophagy in ovarian cancer. Oncol Rep.
Tsai, R.Y. (2014). Turning a new page on nucleostemin and self-renewal. J Cell Sci *127*, 3885-3891.
Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M.*, et al.* (2017). Defining a Cancer Dependency Map. Cell *170*, 564-576 e516.
Tsvetkov, P., Detappe, A., Cai, K., Keys, H.R., Brune, Z., Ying, W., Thiru, P., Reidy, M., Kugener, G., Rossen, J.*, et al.* (2019). Mitochondrial metabolism promotes adaptation to proteotoxic stress. Nat Chem Biol *15*, 681-689.
Viswanatha, R., Li, Z., Hu, Y., and Perrimon, N. (2018). Pooled genome-wide CRISPR screening for basal and context-specific fitness gene essentiality in Drosophila cells. Elife *7*.
Wang, T., Yu, H., Hughes, N.W., Liu, B., Kendirli, A., Klein, K., Chen, W.W., Lander, E.S., and Sabatini, D.M. (2017). Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. Cell *168*, 890-903 e815.
Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J.D., Bussey, H.*, et al.* (1999). Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science *285*, 901-906.
Yan, Y., Rato, C., Rohland, L., Preissler, S., and Ron, D. (2019). MANF antagonizes nucleotide exchange by the endoplasmic reticulum chaperone BiP. Nat Commun *10*, 541.
Zhang, S.F., Wang, X.Y., Fu, Z.Q., Peng, Q.H., Zhang, J.Y., Ye, F., Fu, Y.F., Zhou, C.Y., Lu, W.G., Cheng, X.D.*, et al.* (2015). TXNDC17 promotes paclitaxel resistance via inducing autophagy in ovarian cancer. Autophagy *11*, 225-238.
Zhang, W., Bojorquez-Gomez, A., Velez, D.O., Xu, G., Sanchez, K.S., Shen, J.P., Chen, K., Licon, K., Melton, C., Olson, K.M.*, et al.* (2018). A global transcriptional network connecting noncoding mutations to changes in tumor gene expression. Nat Genet *50*, 613-620.
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.

Figure 1

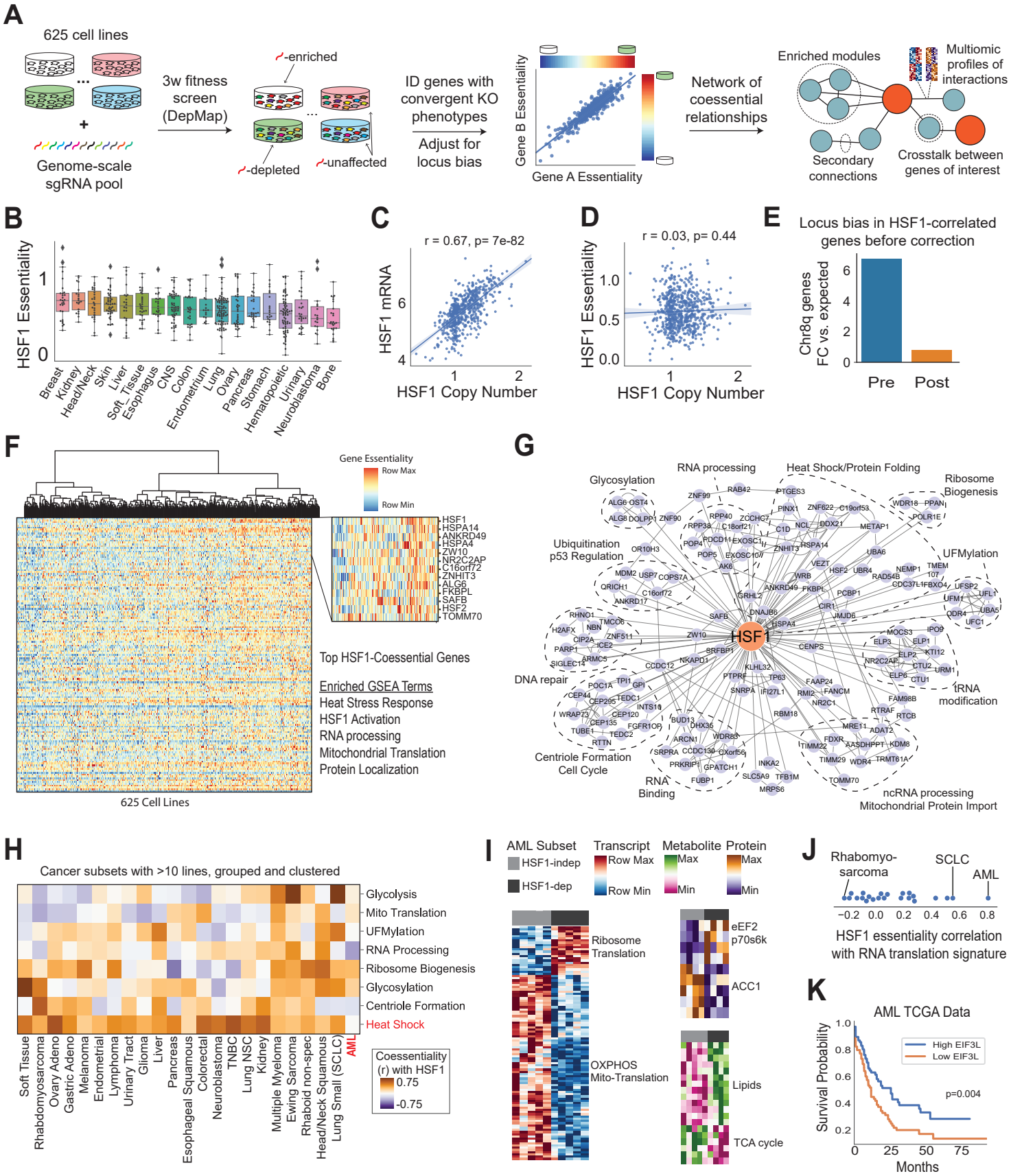

A
625 cell lines
Genome-scale sgRNA pool
3w fitness screen (DepMap)
-enriched
-depleted
-unaffected
ID genes with convergent KO phenotypes
Adjust for locus bias
Gene B Essentiality
Gene A Essentiality
Network of coessential relationships
Enriched modules
Multiomic profiles of interactions
Secondary connections
Crosstalk between genes of interest
B
HSF1 Essentiality
Breast
Kidney
Head/Neck
Skin
Liver
Soft_Tissue
Esophagus
CNS
Colon
Endometrium
Lung
Ovary
Pancreas
Stomach
Hematopoietic
Urinary
Neuroblastoma
Bone
C
r = 0.67, p= 7e-82
HSF1 mRNA
HSF1 Copy Number
D
r = 0.03, p= 0.44
HSF1 Essentiality
HSF1 Copy Number
E
Locus bias in HSF1-correlated genes before correction
Chr8q genes FC vs. expected
Pre
Post
F
Gene Essentiality
Row Max
Row Min
HSF1
HSPA14
ANKRD49
HSPA4
ZW10
NR2C2AP
C16orf72
ZNHIT3
ALG6
FKBPL
SAFB
HSF2
TOMM70
Top HSF1-Coessential Genes
Enriched GSEA Terms
Heat Stress Response
HSF1 Activation
RNA processing
Mitochondrial Translation
Protein Localization
625 Cell Lines
G
Glycosylation
RNA processing
Heat Shock/Protein Folding
Ribosome Biogenesis
Ubiquitination p53 Regulation
UFMylation
DNA repair
tRNA modification
Centriole Formation Cell Cycle
RNA Binding
ncRNA processing Mitochondrial Protein Import
HSF1
H
Cancer subsets with >10 lines, grouped and clustered
Glycolysis
Mito Translation
UFMylation
RNA Processing
Ribosome Biogenesis
Glycosylation
Centriole Formation
Heat Shock
Soft Tissue
Rhabdomyosarcoma
Ovary Adeno
Gastric Adeno
Melanoma
Endometrial
Lymphoma
Urinary Tract
Glioma
Liver
Pancreas
Esophageal Squamous
Colorectal
Neuroblastoma
TNBC
Lung NSC
Kidney
Multiple Myeloma
Ewing Sarcoma
Rhaboid non-spec
Head/Neck Squamous
Lung Small (SCLC)
AML
Coessentiality (r) with HSF1
0.75
-0.75
I
AML Subset
HSF1-indep
HSF1-dep
Transcript
Row Max
Row Min
Metabolite
Max
Min
Protein
Max
Min
Ribosome Translation
OXPHOS Mito-Translation
eEF2
p70s6k
ACC1
Lipids
TCA cycle
J
Rhabomyo-sarcoma
SCLC
AML
HSF1 essentiality correlation with RNA translation signature
K
AML TCGA Data
High EIF3L
Low EIF3L
p=0.004
Survival Probability
Months

Figure 2

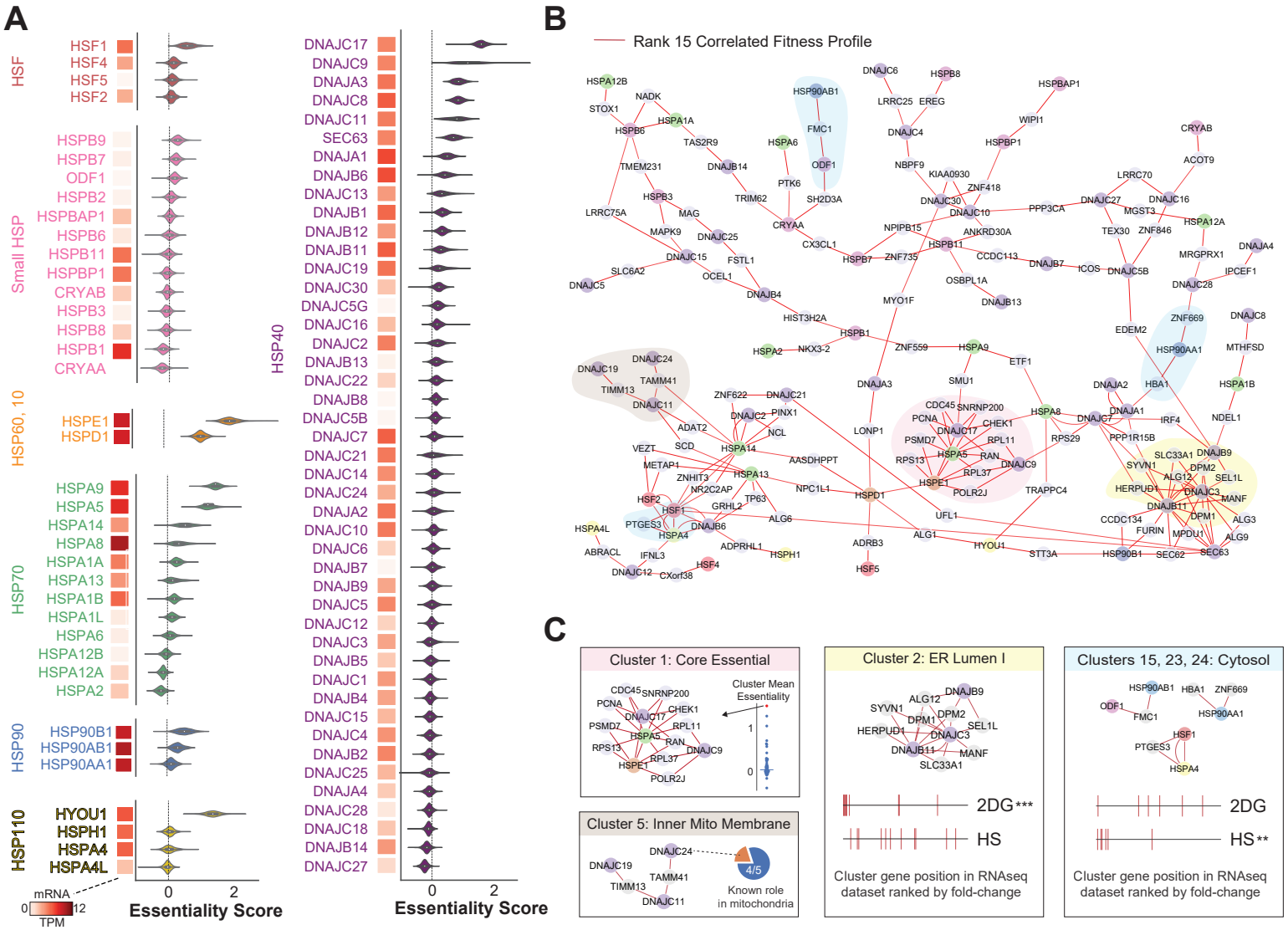

A
HSF
HSF1
HSF4
HSF5
HSF2
Small HSP
HSPB9
HSPB7
ODF1
HSPB2
HSPBAP1
HSPB6
HSPB11
HSPBP1
CRYAB
HSPB3
HSPB8
HSPB1
CRYAA
HSP60, 10
HSPE1
HSPD1
HSP70
HSPA9
HSPA5
HSPA14
HSPA8
HSPA1A
HSPA13
HSPA1B
HSPA1L
HSPA6
HSPA12B
HSPA12A
HSPA2
HSP90
HSP90B1
HSP90AB1
HSP90AA1
HSP110
HYOU1
HSPH1
HSPA4
HSPA4L
mRNA
0
12
TPM
0
2
Essentiality Score
HSP40
DNAJC17
DNAJC9
DNAJA3
DNAJC8
DNAJC11
SEC63
DNAJA1
DNAJB6
DNAJC13
DNAJB1
DNAJB12
DNAJB11
DNAJC19
DNAJC30
DNAJC5G
DNAJC16
DNAJC2
DNAJB13
DNAJC22
DNAJB8
DNAJC5B
DNAJC7
DNAJC21
DNAJC14
DNAJC24
DNAJA2
DNAJC10
DNAJC6
DNAJB7
DNAJB9
DNAJC5
DNAJC12
DNAJC3
DNAJB5
DNAJC1
DNAJB4
DNAJC15
DNAJC4
DNAJB2
DNAJC25
DNAJA4
DNAJC28
DNAJC18
DNAJB14
DNAJC27
0
2
Essentiality Score
B
Rank 15 Correlated Fitness Profile
C
Cluster 1: Core Essential
Cluster Mean Essentiality
1
0
Cluster 5: Inner Mito Membrane
4/5
Known role in mitochondria
Cluster 2: ER Lumen I
2DG***
HS
Cluster gene position in RNAseq dataset ranked by fold-change
Clusters 15, 23, 24: Cytosol
2DG
HS**
Cluster gene position in RNAseq dataset ranked by fold-change

Figure 3

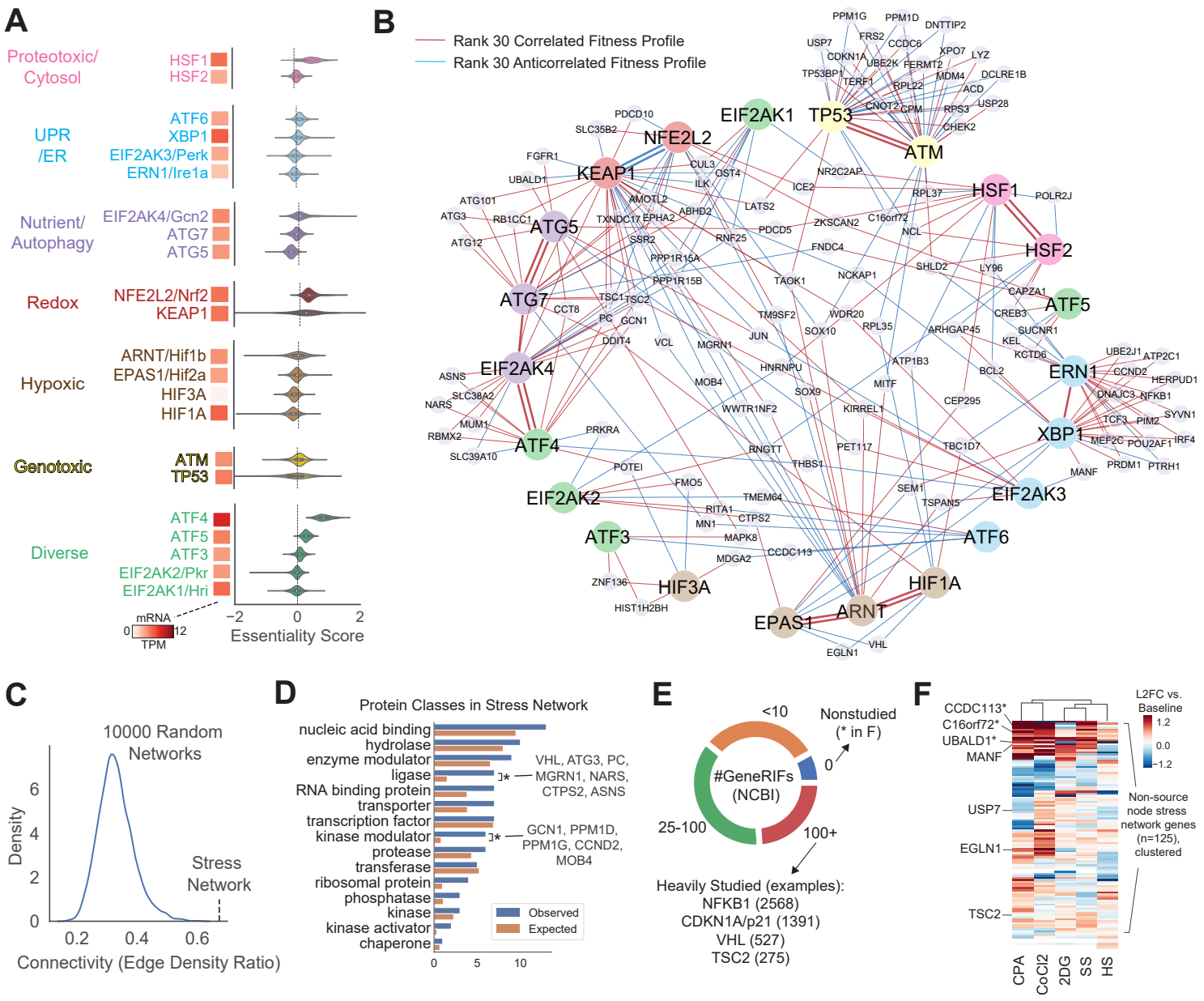
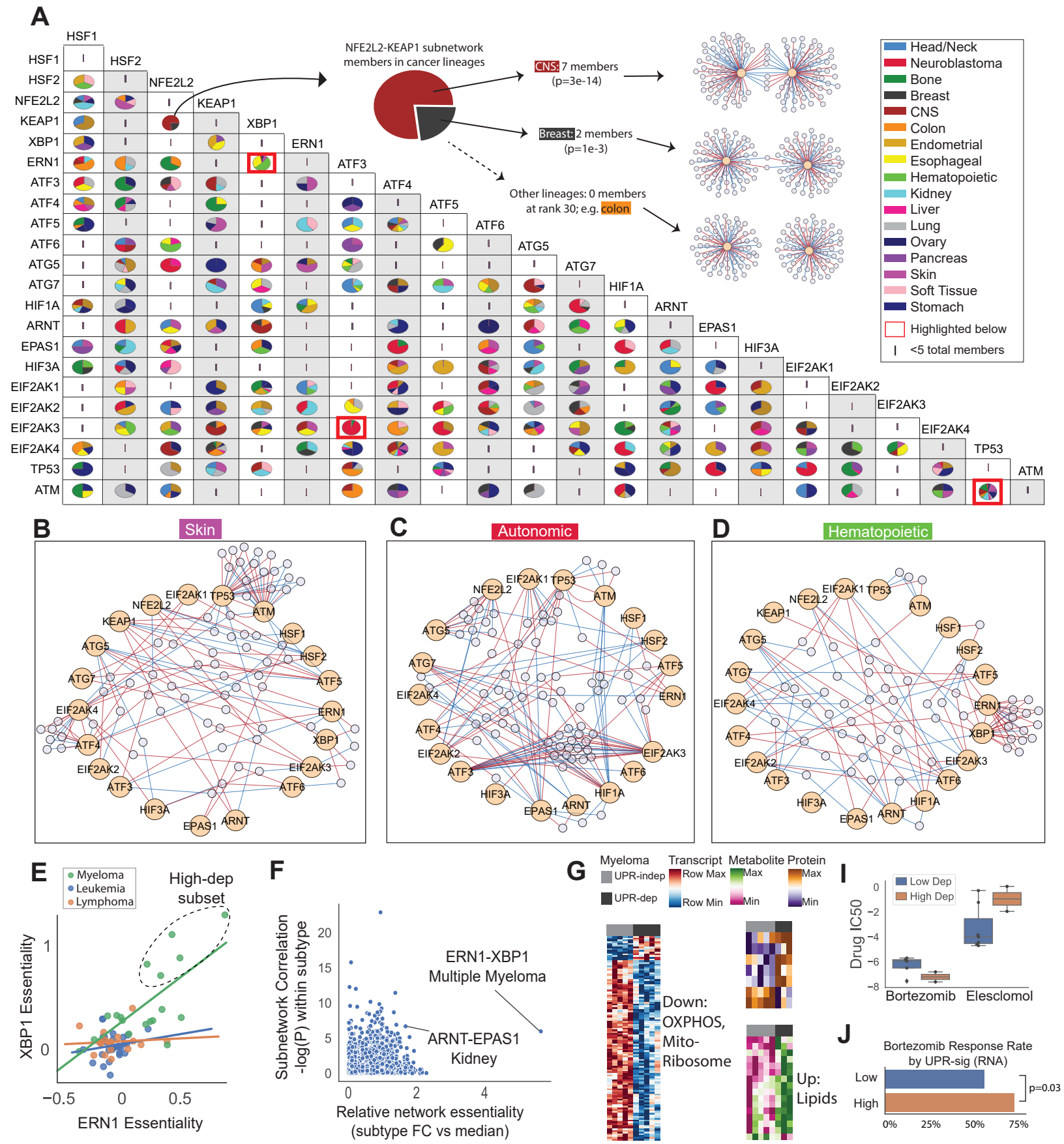

A
Proteotoxic/ Cytosol
HSF1
HSF2
UPR /ER
ATF6
XBP1
EIF2AK3/Perk
ERN1/Ire1a
Nutrient/ Autophagy
EIF2AK4/Gcn2
ATG7
ATG5
Redox
NFE2L2/Nrf2
KEAP1
Hypoxic
ARNT/Hif1b
EPAS1/Hif2a
HIF3A
HIF1A
Genotoxic
ATM
TP53
Diverse
ATF4
ATF5
ATF3
EIF2AK2/Pkr
EIF2AK1/Hri
mRNA
TPM
0
12
−2
0
2
Essentiality Score
B
Rank 30 Correlated Fitness Profile
Rank 30 Anticorrelated Fitness Profile
C
10000 Random Networks
Density
Stress Network
Connectivity (Edge Density Ratio)
D
Protein Classes in Stress Network
nucleic acid binding
hydrolase
enzyme modulator
ligase
RNA binding protein
transporter
transcription factor
kinase modulator
protease
transferase
ribosomal protein
phosphatase
kinase
kinase activator
chaperone
VHL, ATG3, PC, MGRN1, NARS, CTPS2, ASNS
GCN1, PPM1D, PPM1G, CCND2, MOB4
Observed
Expected
E
<10
Nonstudied (* in F)
#GeneRIFs (NCBI)
25-100
100+
Heavily Studied (examples):
NFKB1 (2568)
CDKN1A/p21 (1391)
VHL (527)
TSC2 (275)
F
L2FC vs. Baseline
CCDC113*
C16orf72*
UBALD1*
MANF
USP7
EGLN1
TSC2
Non-source node stress network genes (n=125), clustered
CPA
CoCl2
2DG
SS
HS

Figure 4


A
NFE2L2-KEAP1 subnetwork members in cancer lineages
CNS: 7 members (p=3e-14)
Breast: 2 members (p=1e-3)
Other lineages: 0 members at rank 30; e.g. colon
Head/Neck
Neuroblastoma
Bone
Breast
CNS
Colon
Endometrial
Esophageal
Hematopoietic
Kidney
Liver
Lung
Ovary
Pancreas
Skin
Soft Tissue
Stomach
Highlighted below
<5 total members
B Skin
C Autonomic
D Hematopoietic

E
Myeloma
Leukemia
Lymphoma
High-dep subset
XBP1 Essentiality
ERN1 Essentiality

F
ERN1-XBP1 Multiple Myeloma
ARNT-EPAS1 Kidney
Subnetwork Correlation -log(P) within subtype
Relative network essentiality (subtype FC vs median)

G
Myeloma
UPR-indep
UPR-dep
Transcript
Row Max
Row Min
Metabolite
Max
Min
Protein
Max
Min
Down: OXPHOS, Mito-Ribosome
Up: Lipids

I
Low Dep
High Dep
Drug IC50
Bortezomib
Elesclomol

J
Bortezomib Response Rate by UPR-sig (RNA)
Low
High
p=0.03
0%
25%
50%
75%

Figure 5

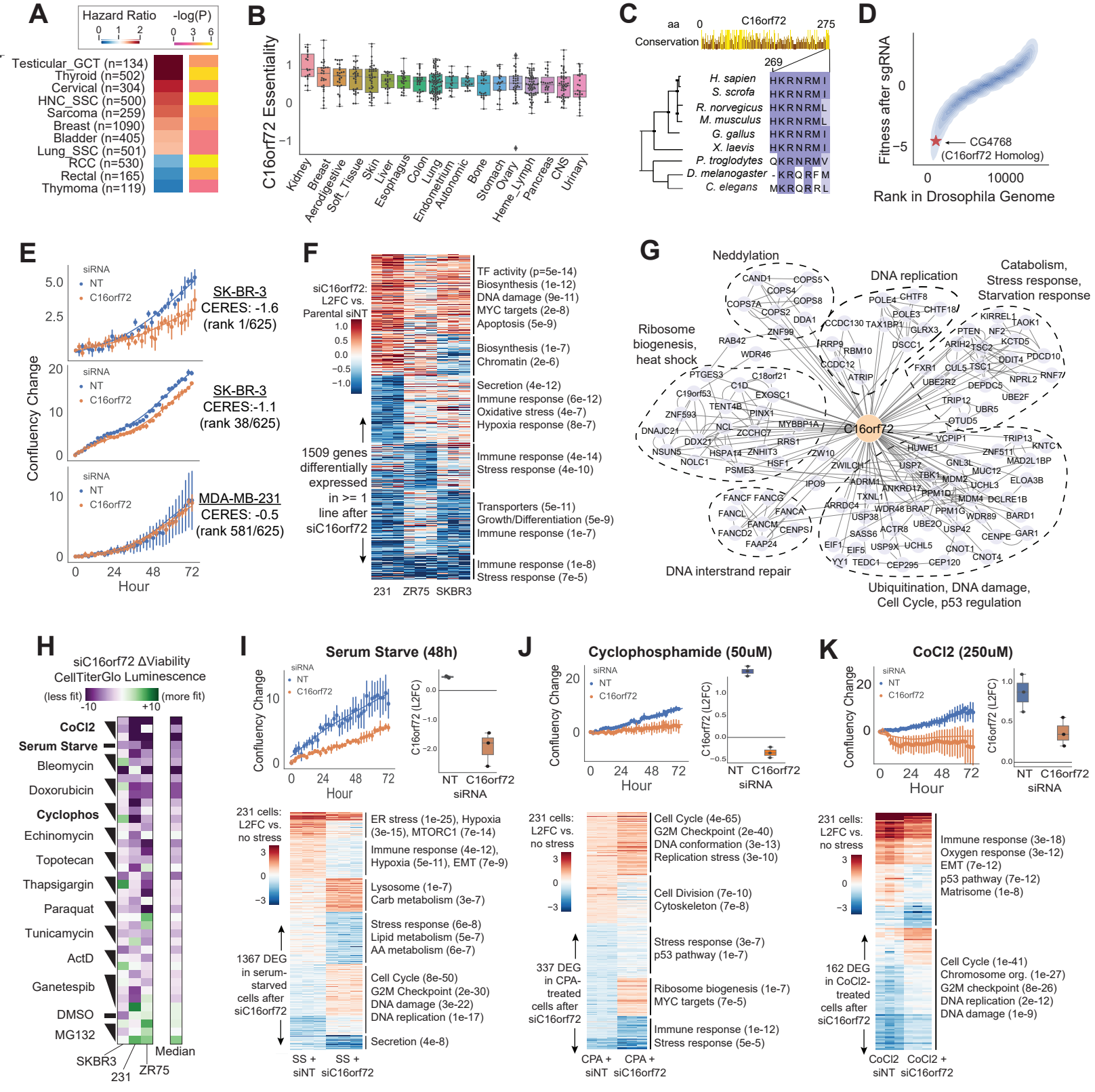
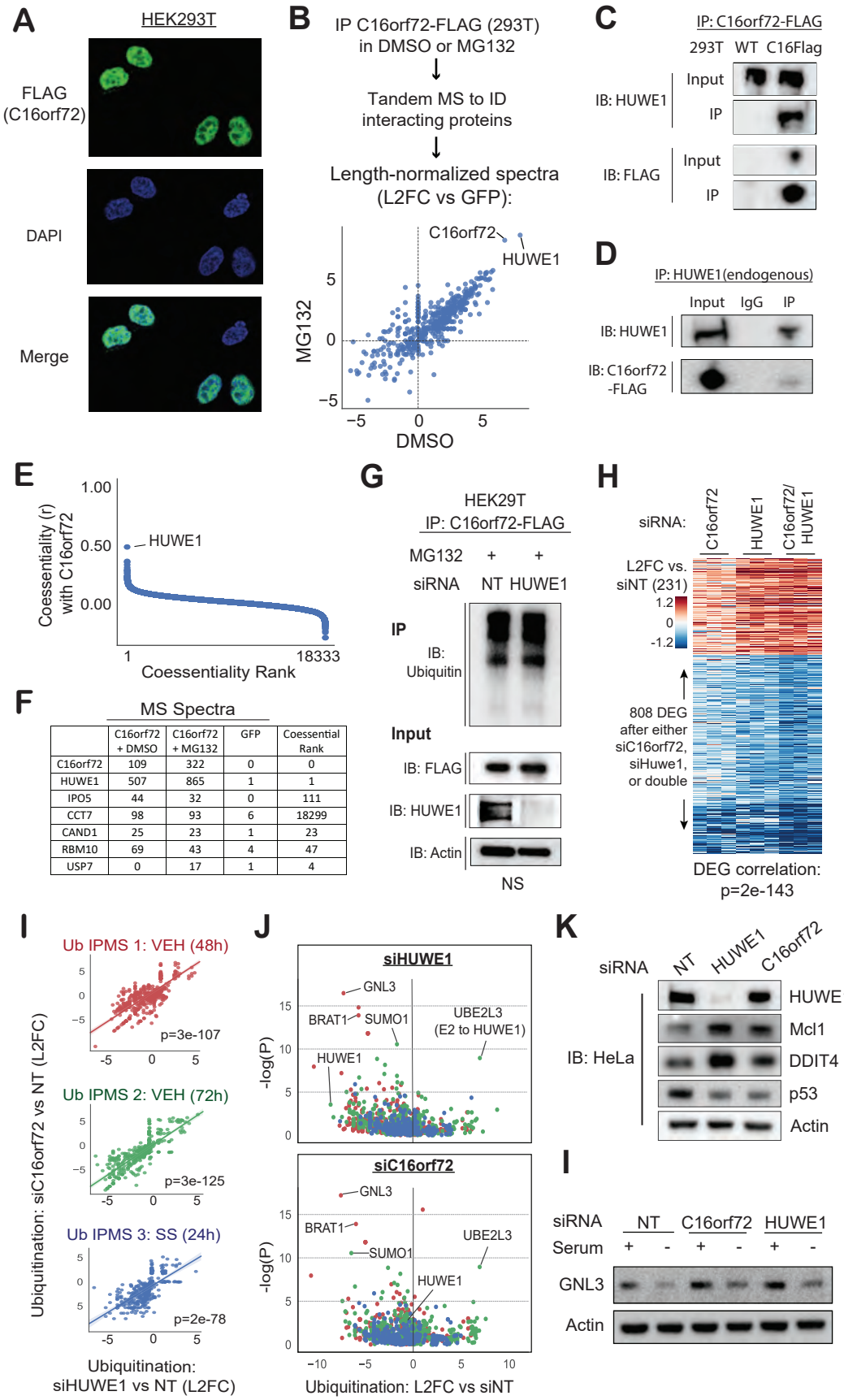
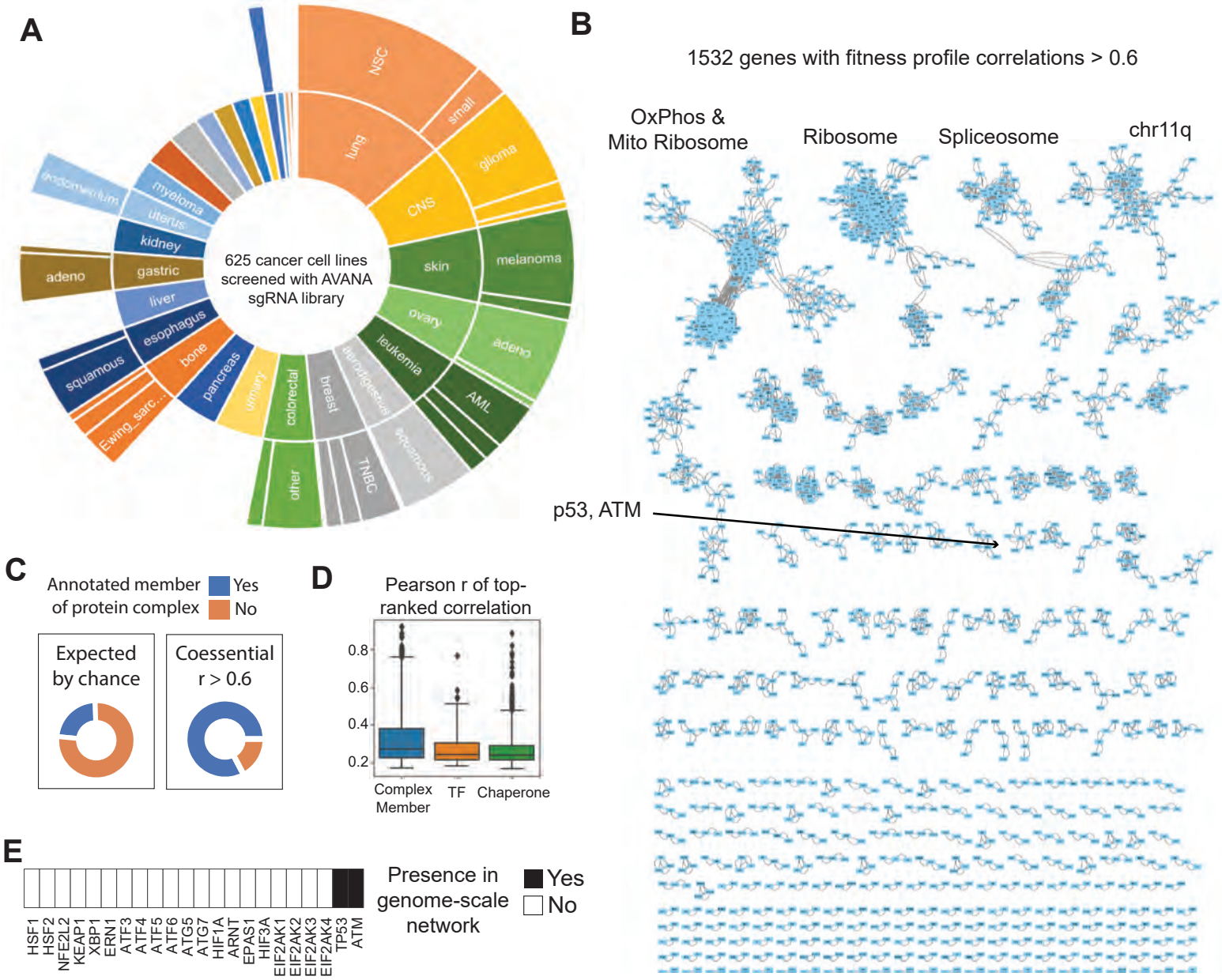

A
Hazard Ratio
-log(P)
Testicular_GCT (n=134)
Thyroid (n=502)
Cervical (n=304)
HNC_SSC (n=500)
Sarcoma (n=259)
Breast (n=1090)
Bladder (n=405)
Lung_SSC (n=501)
RCC (n=530)
Rectal (n=165)
Thymoma (n=119)
B
C16orf72 Essentiality
Kidney
Breast
Aerodigestive
Soft_Tissue
Skin
Liver
Esophagus
Colon
Lung
Endometrium
Autonomic
Bone
Stomach
Ovary
Heme_Lymph
Pancreas
CNS
Urinary
C
aa
C16orf72
Conservation
H. sapien
S. scrofa
R. norvegicus
M. musculus
G. gallus
X. laevis
P. troglodytes
D. melanogaster
C. elegans
D
Fitness after sgRNA
CG4768
(C16orf72 Homolog)
Rank in Drosophila Genome
E
siRNA
NT
C16orf72
SK-BR-3
CERES: -1.6
(rank 1/625)
SK-BR-3
CERES:-1.1
(rank 38/625)
MDA-MB-231
CERES: -0.5
(rank 581/625)
Confluency Change
Hour
F
siC16orf72:
L2FC vs.
Parental siNT
1509 genes
differentially
expressed
in >= 1
line after
siC16orf72
TF activity (p=5e-14)
Biosynthesis (1e-12)
DNA damage (9e-11)
MYC targets (2e-8)
Apoptosis (5e-9)
Biosynthesis (1e-7)
Chromatin (2e-6)
Secretion (4e-12)
Immune response (6e-12)
Oxidative stress (4e-7)
Hypoxia response (8e-7)
Immune response (4e-14)
Stress response (4e-10)
Transporters (5e-11)
Growth/Differentiation (5e-9)
Immune response (1e-7)
Immune response (1e-8)
Stress response (7e-5)
231
ZR75
SKBR3
G
Neddylation
DNA replication
Catabolism,
Stress response,
Starvation response
Ribosome
biogenesis,
heat shock
C16orf72
DNA interstrand repair
Ubiquitination, DNA damage,
Cell Cycle, p53 regulation
H
siC16orf72 ΔViability
CellTiterGlo Luminescence
(less fit)
(more fit)
-10
+10
CoCl2
Serum Starve
Bleomycin
Doxorubicin
Cyclophos
Echinomycin
Topotecan
Thapsigargin
Paraquat
Tunicamycin
ActD
Ganetespib
DMSO
MG132
SKBR3
231
ZR75
Median
I
Serum Starve (48h)
siRNA
NT
C16orf72
Confluency Change
Hour
C16orf72 (L2FC)
NT
C16orf72
siRNA
231 cells:
L2FC vs.
no stress
ER stress (1e-25), Hypoxia
(3e-15), MTORC1 (7e-14)
Immune response (4e-12),
Hypoxia (5e-11), EMT (7e-9)
Lysosome (1e-7)
Carb metabolism (3e-7)
Stress response (6e-8)
Lipid metabolism (5e-7)
AA metabolism (6e-7)
Cell Cycle (8e-50)
G2M Checkpoint (2e-30)
DNA damage (3e-22)
DNA replication (1e-17)
Secretion (4e-8)
1367 DEG
in serum-
starved
cells after
siC16orf72
SS +
siNT
SS +
siC16orf72
J
Cyclophosphamide (50uM)
siRNA
NT
C16orf72
Confluency Change
Hour
C16orf72 (L2FC)
NT
C16orf72
siRNA
231 cells:
L2FC vs.
no stress
Cell Cycle (4e-65)
G2M Checkpoint (2e-40)
DNA conformation (3e-13)
Replication stress (3e-10)
Cell Division (7e-10)
Cytoskeleton (7e-8)
Stress response (3e-7)
p53 pathway (1e-7)
Ribosome biogenesis (1e-7)
MYC targets (7e-5)
Immune response (1e-12)
Stress response (5e-5)
337 DEG
in CPA-
treated
cells after
siC16orf72
CPA +
siNT
CPA +
siC16orf72
K
CoCl2 (250uM)
siRNA
NT
C16orf72
Confluency Change
Hour
C16orf72 (L2FC)
NT
C16orf72
siRNA
231 cells:
L2FC vs.
no stress
Immune response (3e-18)
Oxygen response (3e-12)
EMT (7e-12)
p53 pathway (7e-12)
Matrisome (1e-8)
Cell Cycle (1e-41)
Chromosome org. (1e-27)
G2M checkpoint (8e-26)
DNA replication (2e-12)
DNA damage (1e-9)
162 DEG
in CoCl2-
treated
cells after
siC16orf72
CoCl2
siNT
CoCl2 +
siC16orf72

Figure 6



	C16orf72 + DMSO	C16orf72 + MG132	GFP	Coessential Rank
C16orf72	109	322	0	0
HUWE1	507	865	1	1
IPO5	44	32	0	111
CCT7	98	93	6	18299
CAND1	25	23	1	23
RBM10	69	43	4	47
USP7	0	17	1	4

Figure S1



A
625 cancer cell lines screened with AVANA sgRNA library
NSC
lung
small
glioma
CNS
skin
melanoma
ovary
adeno
leukemia
AML
aerodigestus
breast
TNBC
squamous
other
colorectal
urinary
pancreas
bone
esophagus
liver
gastric
kidney
uterus
myeloma
endometrium
adeno
squamous
Ewing_sarc...
B
1532 genes with fitness profile correlations > 0.6
OxPhos & Mito Ribosome
Ribosome
Spliceosome
chr11q
p53, ATM
C
Annotated member of protein complex
Yes
No
Expected by chance
Coessential r > 0.6
D
Pearson r of top-ranked correlation
0.8
0.6
0.4
0.2
Complex Member
TF
Chaperone
E
HSF1
HSF2
NFE2L2
KEAP1
XBP1
ERN1
ATF3
ATF4
ATF5
ATF6
ATG5
ATG7
HIF1A
ARNT
EPAS1
HIF3A
EIF2AK1
EIF2AK2
EIF2AK3
EIF2AK4
TP53
ATM
Presence in genome-scale network
Yes
No

Figure S2

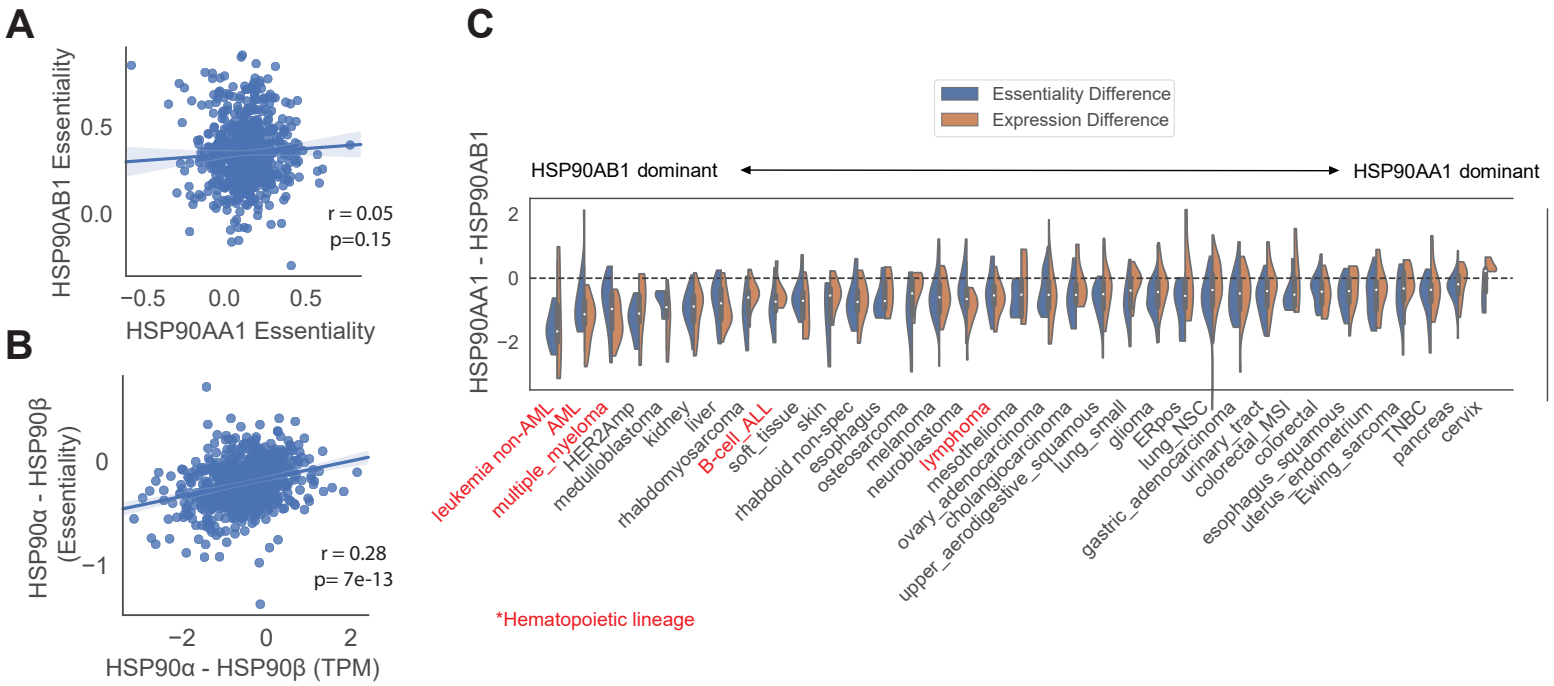

A
HSP90AB1 Essentiality
0.5
0.0
r = 0.05
p=0.15
−0.5
0.0
0.5
HSP90AA1 Essentiality
B
HSP90α - HSP90β (Essentiality)
0
−1
r = 0.28
p= 7e-13
−2
0
2
HSP90α - HSP90β (TPM)
C
Essentiality Difference
Expression Difference
HSP90AB1 dominant
HSP90AA1 dominant
HSP90AA1 - HSP90AB1
2
0
−2
leukemia non-AML
AML
multiple_myeloma
HER2Amp
medulloblastoma
kidney
liver
rhabdomyosarcoma
B-cell_ALL
soft_tissue
skin
rhabdoid non-spec
esophagus
osteosarcoma
melanoma
neuroblastoma
lymphoma
mesothelioma
ovary_adenocarcinoma
cholangiocarcinoma
upper_aerodigestive_squamous
lung_small
glioma
ERpos
lung_NSC
gastric_adenocarcinoma
urinary_tract
colorectal_MSI
colorectal
esophagus_squamous
uterus_endometrium
Ewing_sarcoma
TNBC
pancreas
cervix
*Hematopoietic lineage

Figure S3

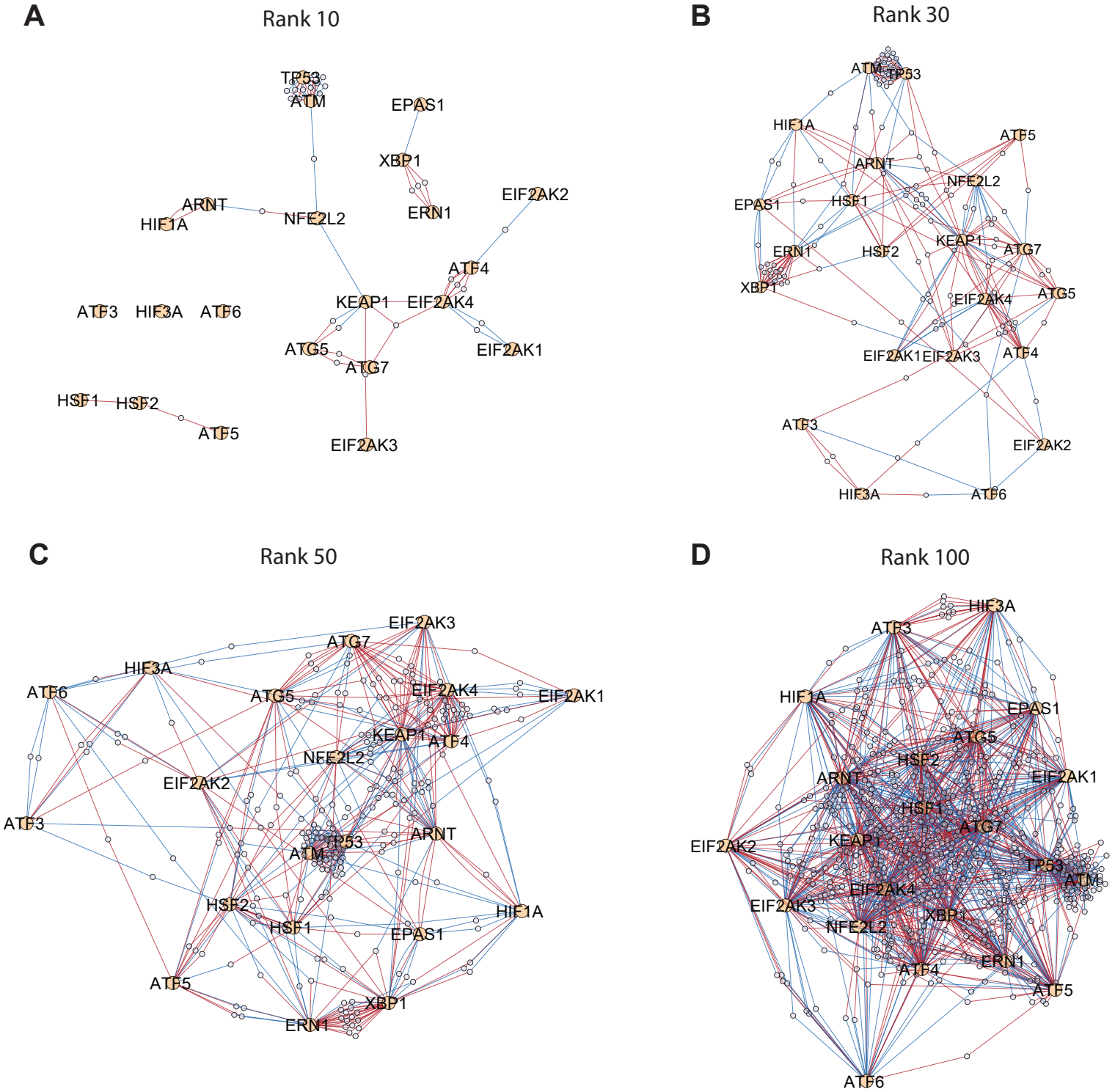

A
Rank 10
TP53
ATM
EPAS1
XBP1
ERN1
EIF2AK2
ARNT
HIF1A
NFE2L2
ATF4
ATF3
HIF3A
ATF6
KEAP1
EIF2AK4
ATG5
ATG7
EIF2AK1
HSF1
HSF2
ATF5
EIF2AK3
B
Rank 30
ATM
TP53
HIF1A
ATF5
ARNT
NFE2L2
EPAS1
HSF1
KEAP1
ATG7
ERN1
HSF2
XBP1
EIF2AK4
ATG5
EIF2AK1
EIF2AK3
ATF4
ATF3
EIF2AK2
HIF3A
ATF6
C
Rank 50
EIF2AK3
ATG7
HIF3A
ATF6
ATG5
EIF2AK4
EIF2AK1
KEAP1
ATF4
NFE2L2
EIF2AK2
ATF3
TP53
ARNT
ATM
HSF2
HIF1A
HSF1
EPAS1
ATF5
XBP1
ERN1
D
Rank 100
HIF3A
ATF3
HIF1A
EPAS1
ATG5
HSF2
ARNT
EIF2AK1
HSF1
ATG7
EIF2AK2
KEAP1
TP53
ATM
EIF2AK4
EIF2AK3
XBP1
NFE2L2
ERN1
ATF4
ATF5
ATF6

Figure S4

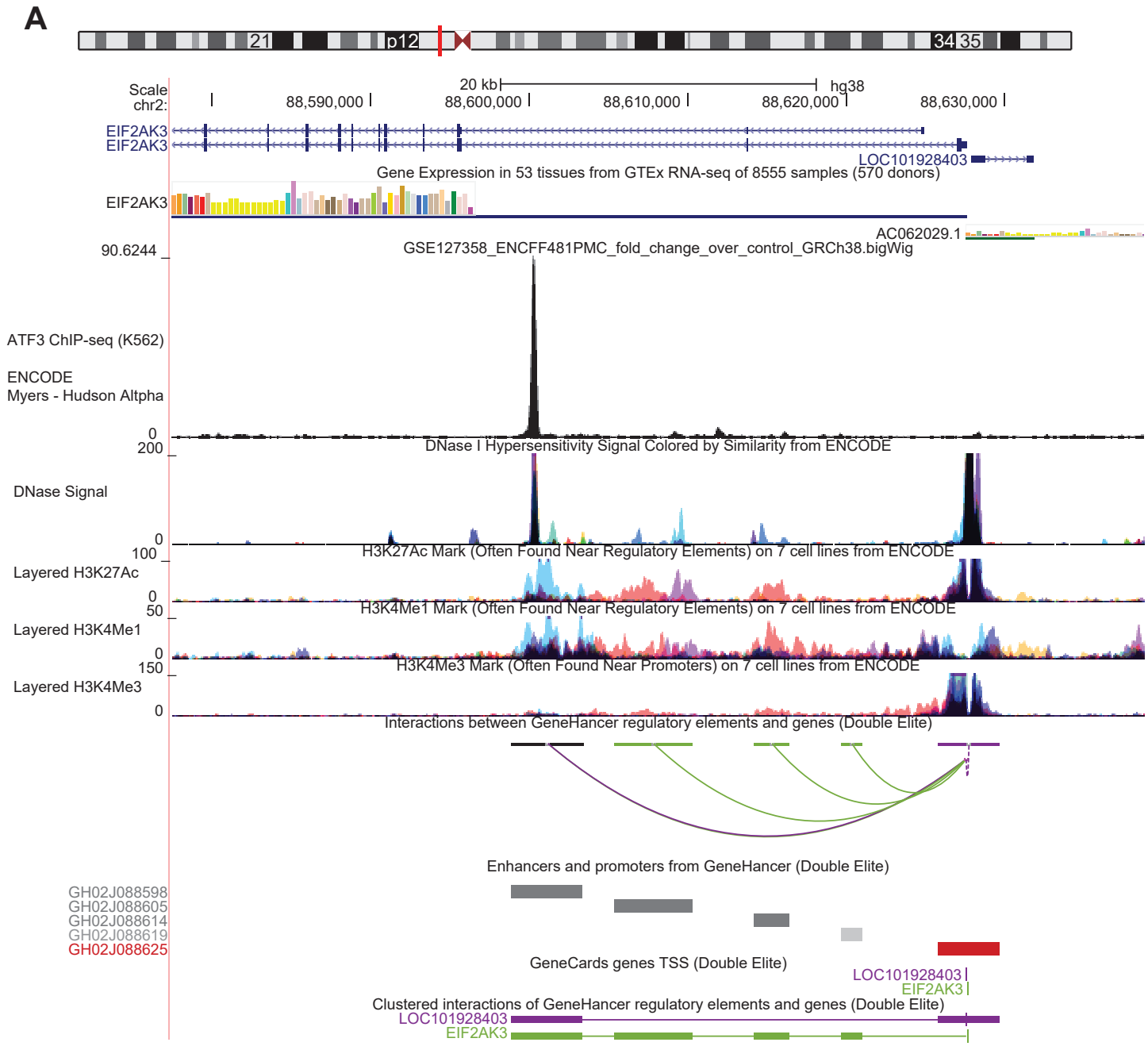
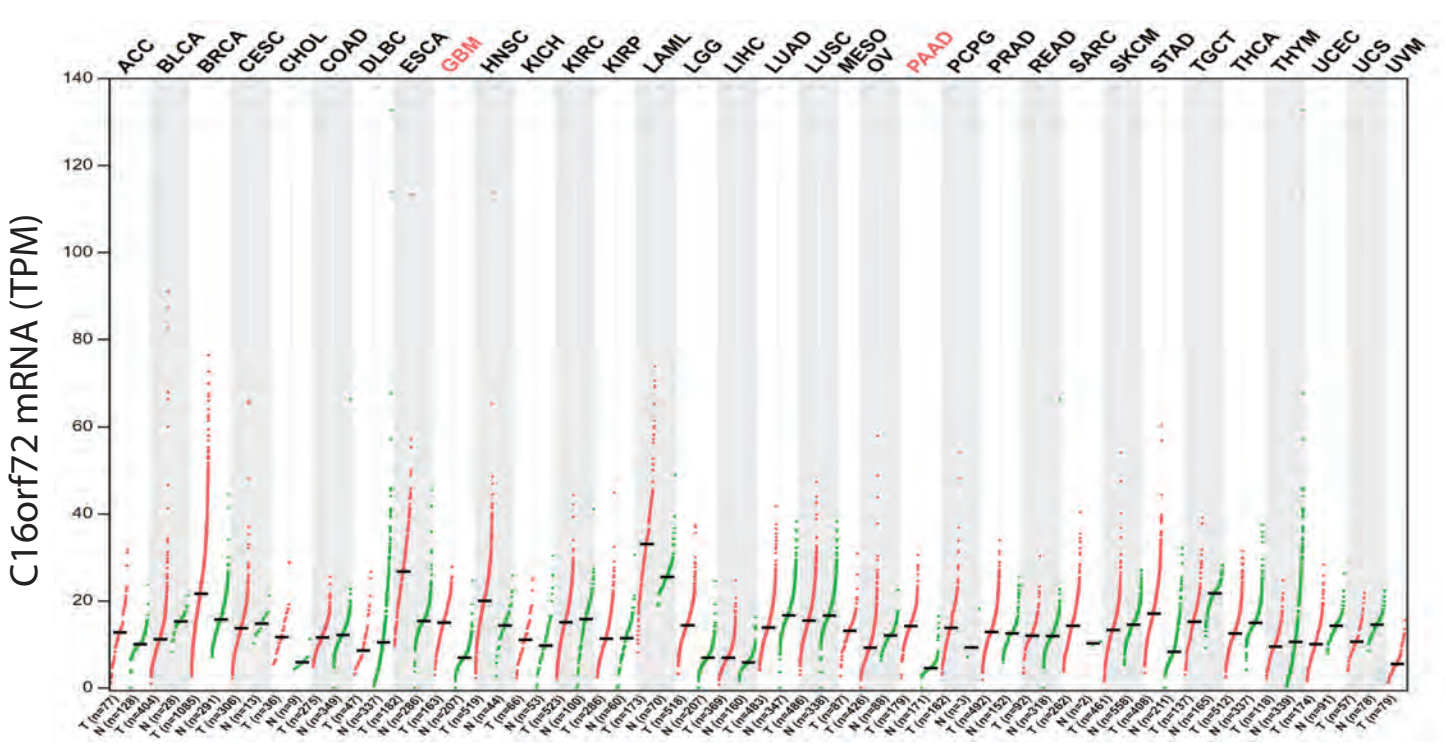

A
21
p12
34
35
Scale
chr2:
20 kb
hg38
88,590,000
88,600,000
88,610,000
88,620,000
88,630,000
EIF2AK3
EIF2AK3
LOC101928403
Gene Expression in 53 tissues from GTEx RNA-seq of 8555 samples (570 donors)
EIF2AK3
AC062029.1
90.6244
GSE127358_ENCFF481PMC_fold_change_over_control_GRCh38.bigWig
ATF3 ChIP-seq (K562)
ENCODE
Myers - Hudson Altpha
0
200
DNase I Hypersensitivity Signal Colored by Similarity from ENCODE
DNase Signal
0
100
H3K27Ac Mark (Often Found Near Regulatory Elements) on 7 cell lines from ENCODE
Layered H3K27Ac
0
50
H3K4Me1 Mark (Often Found Near Regulatory Elements) on 7 cell lines from ENCODE
Layered H3K4Me1
0
150
H3K4Me3 Mark (Often Found Near Promoters) on 7 cell lines from ENCODE
Layered H3K4Me3
0
Interactions between GeneHancer regulatory elements and genes (Double Elite)
Enhancers and promoters from GeneHancer (Double Elite)
GH02J088598
GH02J088605
GH02J088614
GH02J088619
GH02J088625
GeneCards genes TSS (Double Elite)
LOC101928403
EIF2AK3
Clustered interactions of GeneHancer regulatory elements and genes (Double Elite)
LOC101928403
EIF2AK3

Figure S5

A


ACC
BLCA
BRCA
CESC
CHOL
COAD
DLBC
ESCA
GBM
HNSC
KICH
KIRC
KIRP
LAML
LGG
LIHC
LUAD
LUSC
MESO
OV
PAAD
PCPG
PRAD
READ
SARC
SKCM
STAD
TGCT
THCA
THYM
UCEC
UCS
UVM
C16orf72 mRNA (TPM)
0
20
40
60
80
100
120
140

B

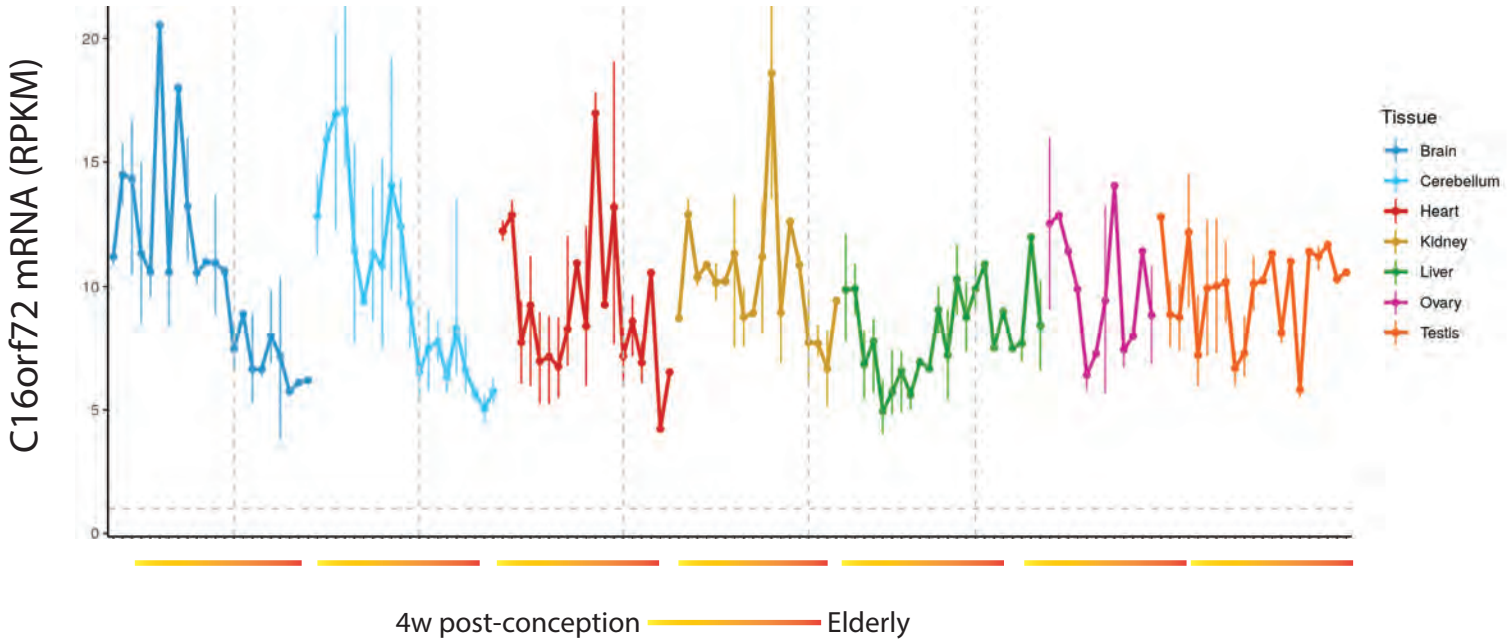
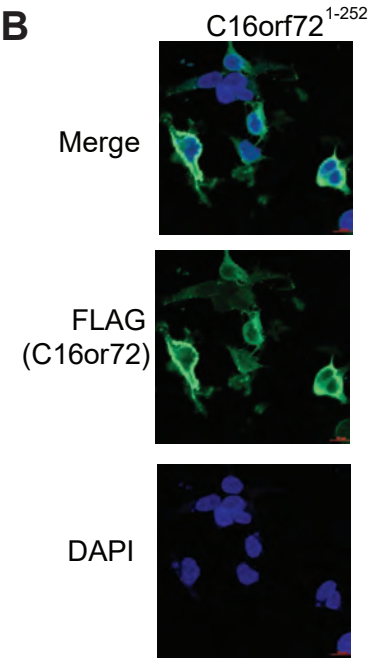

C16orf72 mRNA (RPKM)
0
5
10
15
20
Tissue
Brain
Cerebellum
Heart
Kidney
Liver
Ovary
Testis
4w post-conception
Elderly

Figure S6

A

Predicted NLSs in query sequence	
MEERKEEGEAEIQEHGPEHWFSKWERQCLAEAEQDEQLPPELQEEAAAAA	50
QPEHKQQKLWHLFQNSATAVAQLYKDRVCQQPGLSLWVPFQNAATAVTNL	100
YKESVDTHQRSFDIGIQIGYQRRNKDVLAWVKKRRRTIRREDLISFLCGK	150
VPPPRNSRAPPRLTVVSPNRATSTETSSSVETDLQPFREAIALHGLSGAM	200
ASISVRSSTPGSPTHVSSGSNASRRRNGLHDVDLNTFISEEMALHLDNGG	250
TRKRTSAQCGDVITDSPTHKRNRMI	275

Predicted bipartite NLS		
Pos.	Sequence	Score
252	RKRTSAQCGDVITDSPTHKRNRMI	11.5

Predicted monopartite NLS		
Pos.	Sequence	Score
128	LAWVKKRRRTI	8
130	WVKKRRRTIR	8

B