Cancers adapt to their mutational load by buffering protein misfolding stress

Susanne Tilk[1*], Judith Frydman[1], Christina Curtis[2,3,4*], Dmitri Petrov[1,4*]

[1]Department of Biology, Stanford University, Stanford, CA, USA.
[2]Department of Medicine, Division of Oncology, Stanford University School of Medicine, Stanford, CA, USA
[3]Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA.
[4]Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA.
*Correspondence to:
tilk@stanford.edu, cncurtis@stanford.edu, petrov@stanford.edu

Statement of Significance

Cancers can successfully survive an accumulation of thousands of protein-damaging mutations. Here, we show that high mutational load tumors mitigate these damaging consequences by up-regulating complexes that buffer against protein misfolding stress – providing novel therapeutic vulnerabilities and suggesting that disruption of proteostasis is a hallmark of somatic evolution.

Abstract

In asexual populations that don't undergo recombination, such as cancer, deleterious mutations are expected to accrue readily due to genome-wide linkage between mutations. Despite this mutational load of often thousands of deleterious mutations, many tumors thrive. How tumors survive the damaging consequences of this mutational load is not well understood. Here, we investigate the functional consequences of mutational load in 10,295 human tumors by quantifying their phenotypic response through changes in gene expression. Using a generalized linear mixed model (GLMM), we find that high mutational load tumors up-regulate proteostasis machinery related to the mitigation and prevention of protein misfolding. We replicate these expression responses in cancer cell lines and show that the viability in high mutational load cancer cells is strongly dependent on complexes that degrade and refold proteins. This indicates that upregulation of proteostasis machinery is causally important for high mutational burden tumors and uncovers new therapeutic vulnerabilities.

Introduction

Cancer develops from an accumulation of somatic mutations over time. While a small subset of these mutations drive tumor progression, the vast majority of remaining mutations, known as passengers, don't help and might hinder cancer growth. The role that passengers play in tumor progression has traditionally received little attention despite their abundance and variation across cancer types. The number of passengers in a tumor can vary by over four orders of magnitude, even within the same cancer type, from just a few to tens of thousands of point mutations[1].

Whether these passengers are neutral or damaging to tumors has long been a matter of debate[2–10]. Some have argued that passengers are functionally unimportant to tumors given that most non-synonymous mutations are not removed by negative selection in somatic tissues[2,3]. This is in direct contrast to the human germ-line, where non-synonymous mutations are functionally damaging to most genes[11] and signals of negative selection are pervasive[3]. The common explanation for why damaging protein-coding mutations are removed in the human-germline but maintained in somatic tissues is that most genes are only important for multi-cellular function at the organismal level (e.g. during development), but not during somatic growth[2,12].

However, the notion that non-synonymous mutations are only selectively neutral in somatic tissues is surprising given their known functional consequences in the germ-line. Non-synonymous mutations are known to be damaging in the human germ-line due to their effects on protein folding and stability[13], which ought to be shared between somatic and germline evolution. An alternative explanation is that non-synonymous mutations are indeed damaging in somatic evolution, but negative selection is too inefficient at removing them due to linkage effects driven by the lack of recombination in somatic cells[10]. Without recombination to break apart combinations of mutations, selection must act on beneficial drivers and deleterious passengers that arise in the same genome together. This makes it less efficient for selection to individually favor beneficial drivers or remove deleterious passengers[14]. As a result, a substantial number of weakly damaging passengers can accrue in cancer due to inefficient negative selection over time. In support of this model, tumors with very small numbers of passengers – where linkage effects are expected to be negligible – have recently been shown to exhibit signatures of negative selection and weed out damaging non-synonymous mutations[10]. In contrast, the remaining majority (>95%) of tumors, which contain much larger numbers of linked mutations, display patterns of inefficient negative selection. This provides evidence in favor of the inefficient selection model and implies that most tumors carry a correspondingly large deleterious mutational load.

If individual passengers are in fact substantially damaging in cancer, successful tumors with thousands of linked mutations must find ways to maintain their viability by mitigating this large mutational load. While paths to mitigation are difficult to predict for non-coding mutations, tumors with mutations in protein-coding genes are expected to minimize the damaging phenotypic effects of protein mis-folding stress. Here, we investigate this hypothesis by analyzing tumor tissues with paired mutational and gene expression profiles to assess how the physiological state of cancer cells change as they accumulate protein coding mutations. Using a general linear mixed effects regression model (GLMM), we leverage variation across 10,295 tumors from 33 cancer types and find that complexes that re-fold proteins (chaperones), degrade proteins (proteasome)

and splice mRNA (spliceosome) are up-regulated in high mutation load tumors. We validate these results by showing that similar physiological responses occur in high mutational load cancer cell lines as well. Finally, we establish a causal connection by showing that high mutational load cell lines are particularly sensitive when proteasome and chaperone function is disrupted through downregulation of expression via short-hairpin RNA (shRNA) knock-down or targeted therapies. Collectively, these data indicate that the viability of high mutational load tumors is strongly dependent on the up-regulation of complexes that degrade and refold proteins, revealing a generic vulnerability of cancer that can potentially be therapeutically exploited.

Results

Quantifying transcriptional response to mutational load in human tumors.

We first performed a genome-wide screen to systematically identify which genes are transcriptionally upregulated in response to mutational load in human tumors. To do so, we utilized publicly available whole-exome and gene expression data from 10,295 human tumors across 33 cancer types from The Cancer Genome Atlas (TCGA)[15,16]. We considered multiple classes of mutations to define mutational load and investigated their degree of collinearity, focusing on protein-coding regions since the use of whole-exome data limits the ability to accurately assess mutations in non-coding regions. We find that there is a high degree of collinearity among synonymous, non-synonymous and nonsense point mutations in protein coding genes ($R > 0.9$) but weak collinearity between point mutations and copy number alterations ($R < 0.05$) (Supplemental Figure 1). Thus, we decided to focus on the aggregate effects of protein-coding mutations and for all analyses defined mutational load as $\log_{10}$ of the total number of point mutations in protein-coding genes. For simplicity, we used all mutations rather than focusing only on passenger mutations since identifying genuine drivers against a background of linked passenger events can be difficult, especially for tumors with many mutations.

Since gene expression can vary across tumors due to many factors, such as cancer type, tumor purity and other unknown factors, we utilized a generalized linear mixed model (GLMM) to measure the association of mutational load and gene expression while accounting for these potential confounders (Fig. 1A). Within the GLMM, tumor purity and mutational load were modeled as fixed effects whereas cancer type was modeled as a random effect since it varies across groups of patients and can be interpreted as repeated measurements across groups. The following GLMM was applied separately to each gene,

$$Y \sim \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \mathrm{v} + e$$

where Y is a vector of normalized expression values across all tumors, β_0 is the fixed intercept, β_1 is the fixed slope for the predictor variable X_1 which is a vector of mutational load values for each tumor, β_2 is the fixed slope for the predictor variable X_2 which is a vector of the purity of each tumor, v is the random intercept for each cancer type, and e is a Gaussian error term (Methods).

Using this approach, we applied the GLMM to all tumors in TCGA and identified 5,330 genes that are significantly up-regulated in response to mutational load ($\boldsymbol{\beta}_1 > 0$, FDR < 0.05). Next, we linked these genes to cellular function by performing gene set enrichment to known protein complexes (CORUM database[17], Fig. 1B) and pathways (KEGG database[18], Fig. 1C) using gprofiler2[19]. As expected for tumors with many mutations, pathways and protein complexes related to cell cycle, DNA replication and DNA repair were enriched in tumors with a high mutational load. However, some of the most significant enrichment terms were for protein complexes and pathways that regulate translation (mitochondrial ribosomes), protein degradation (proteasome complex), and protein folding (CCT complex/HSP60), consistent with the hypothesis that high mutational load tumors experience protein misfolding stress. Surprisingly, we also found that the spliceosome, a large protein complex that regulates alternative splicing in cells, is up-regulated in response to mutational load. This suggests that transcription itself could also be regulated in response to protein misfolding stress, as seen in other studies[20,21]. In addition, we confirmed that the same proteostasis complexes are identified when performing gene set enrichment analysis only genes with the largest effect sizes from the transcriptional screen (in the upper quartile of $\boldsymbol{\beta}_1$ regression coefficients), which contain half the number of significant genes as identified previously (N=2,152 vs 5,330; Supplemental Figure 2).

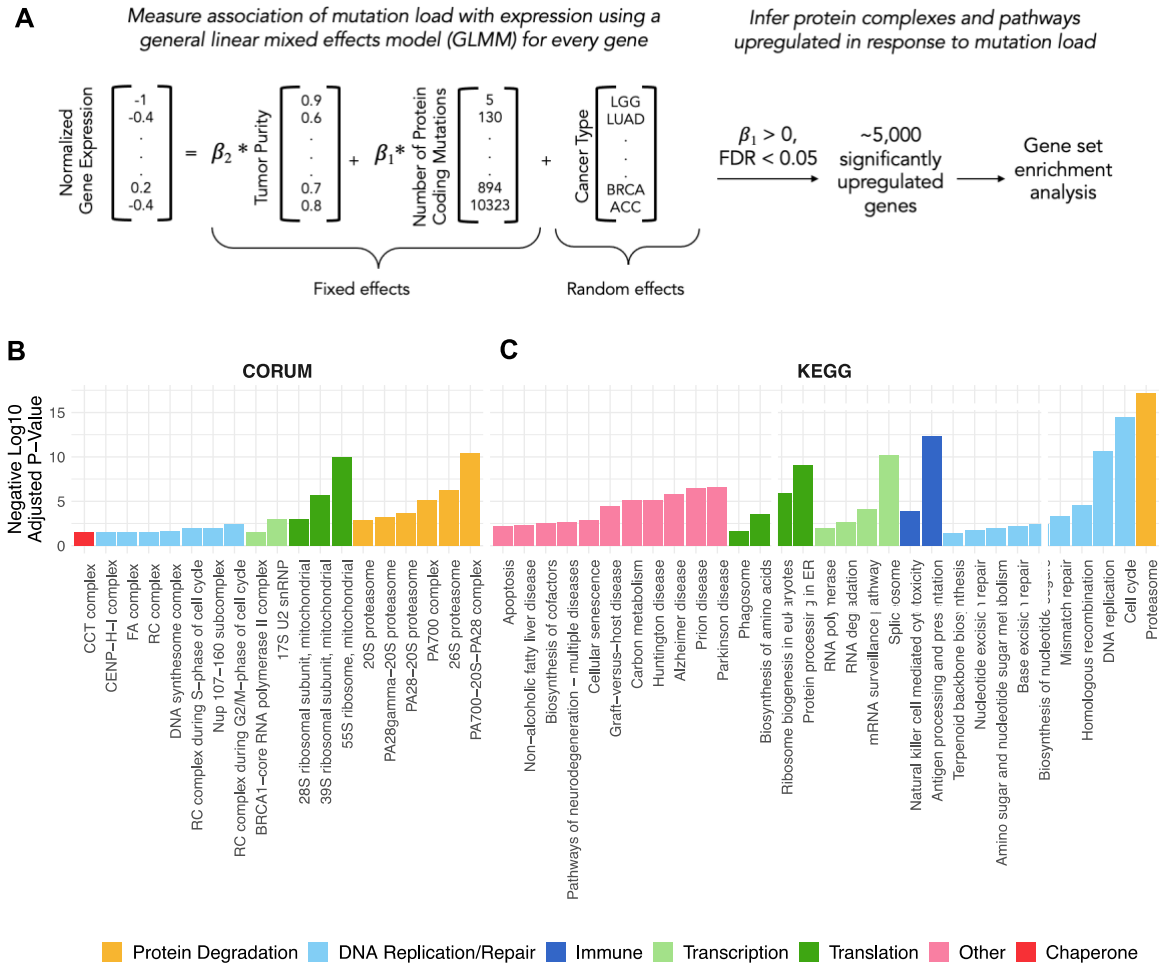


Figure 1. General linear mixed effects model (GLMM) identifies protein complexes and pathways up-regulated in response to mutational load in human tumors. (A) Overview of the GLMM used to measure the association of mutation load with gene expression while controlling for potential co-variates (purity and cancer type). Genes with a significant, positive $\boldsymbol{\beta}_1$ regression coefficient and false discovery rate (FDR) < 0.05 are used for gene set enrichment analysis. **(B-C)** Bar plots of protein complexes from the CORUM database (left) and pathways from the KEGG database (right) that are significantly enriched ($p < 0.05$) in response to mutational load. Length of bars denote negative $\log_{10}$ of adjusted p-value and colors denote broad functional groups enriched in both databases.

Gene silencing through alternative splicing in high mutational load tumors.

We next investigated in detail how these protein complexes could mitigate the damaging effects of protein misfolding in high mutational load tumors by examining the role of the spliceosome in gene silencing. We hypothesized that the up-regulation of the spliceosome in high mutational load tumors prevents further protein misfolding by regulating pre-mRNA transcripts to be degraded rather than translated. The down-regulation of gene expression via alternative splicing events, such as intron retention, is one known mechanism to silence genes by funneling transcripts to mRNA decay pathways.[22–24]

To test whether gene expression is down-regulated in high mutational load tumors through intron retention, we utilized previously called alternative splicing events in TCGA[25]. Alternative splicing events within this dataset were quantified through a metric called *percent spliced in* or PSI. PSI is calculated as the number of reads that overlap the alternative splicing event (e.g. for intron retention, either at intronic regions or those at the boundary of exon to intron junctions) divided by the total number of reads that support and don't support the alternative splicing event. Thus, PSI estimates the probability of alternative splicing events only at specific exonic boundaries in the entire transcript population without requiring information on the complete underlying composition of each full length-transcript.

Using these alternative splicing calls, we reasoned that if a transcript contains an intron retention event and is downregulated in expression, the transcript is more likely to have been degraded by mRNA decay pathways. For all genes, we first quantified whether intron retention events were present based on a threshold value >80% PSI. For each gene with an intron retention event, we quantified whether the expression of the same gene was under-expressed. Each gene was counted as under-expressed if it was one standard deviation below the mean expression within the same cancer type. To control for mutations that might affect patterns of expression, (i.e., expression quantitative trait loci or eQTL effects), alternative splicing events that contained a point mutation within the same gene were removed from the analysis (which only represent ~1% of intron retention events across all tumors; Methods). We find that relative to all transcripts with intron retention events, the number of transcripts that are under-expressed increases with tumor mutational load (Fig. 2A), suggesting that the degree of intron-retention driven mRNA decay is elevated in high mutational load tumors. This trend is robust to other PSI value thresholds (>50-90% PSI), even for other alternative splicing events (e.g., exon skipping, mutually exclusive exons, etc.) and when not filtering for potential eQTL effects (Supplemental Figure 3 and 4).

We next investigated which genes are more likely to be silenced through mRNA decay between low and high mutational load tumors. For each intron retention event,

we calculated whether PSI values were significantly different in low mutational load tumors (<10 total protein-coding mutations) compared to high mutational load tumors (>1000 total protein-coding mutations) using a t-test. This approach identified 606 and 201 genes that have more and less intron retention events in high mutational load tumors, respectively. Using gene set enrichment analysis, we find that cytoplasmic ribosomes contain more intron retention events in high mutational load tumors, potentially leading to their down-regulation through mRNA decay to prevent further protein mis-folding (Fig. 2B). Genes that contain fewer intron retention events in high mutational load tumors, which are less likely to undergo mRNA decay, are primarily related to mRNA splicing.

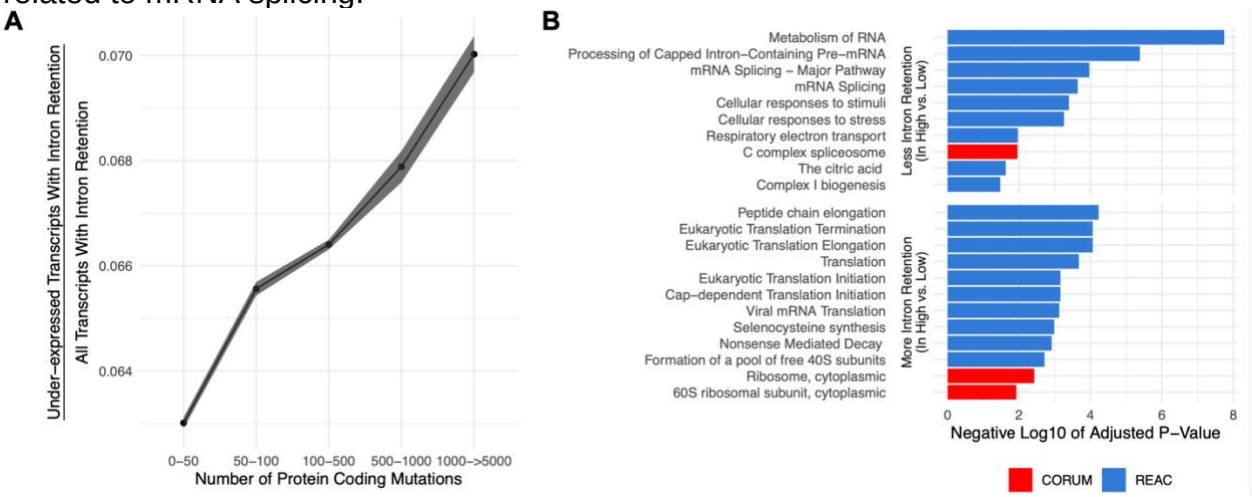


Figure 2. Gene silencing is elevated in high mutational load tumors likely through the coupling of intron retention with mRNA decay. (A) Counts of the number of under-expressed transcripts with intron retention events, relative to counts of all intron retention events in tumors binned by the total number of protein-coding mutations. Intron retention events with PSI > 80% are counted. Error bars are 95% confidence intervals determined by bootstrap sampling. **(B)** Barplot of significant protein complexes in the CORUM database (in red) and Reactome pathway database (in blue) with more (bottom) and less (top) intron retention events in high mutational load tumors compared to low mutational load tumors.

Regulation of translation, protein folding and protein degradation in high mutational load tumors.

Next, we investigated in detail how the remaining proteostasis complexes that were significant in our genome-wide screen, which regulate protein synthesis, degradation and folding, could mitigate protein misfolding in high mutational load tumors. To do so, we expanded our gene sets to include other chaperone families, all ribosomal complexes and proteasomal subunits (Fig. 3A). Using the GLMM framework detailed above, we find that the expression of nearly all individual genes in chaperone families that participate in protein folding (HSP60, HSP70 and HSP90), protein disaggregation (HSP100), and have organelle-specific roles (ER and mitochondrial) are significantly up-regulated in response to mutational load. Interestingly, however, small heat shock proteins, which don't participate in protein folding or disaggregation, are significantly down-regulated in response to increased protein coding mutations. The role of small heat shock proteins is primarily to hold unfolded proteins in a reversible state

for re-folding or degradation by other chaperones[26] and thus, could possibly be down-regulated due to their inefficiency in mitigating protein misfolding.

We further examined differences in expression of different structural components of the proteasome, a large protein complex responsible for degradation of intracellular proteins. Consistent with the over-expression of chaperone families that mitigate protein mis-folding, both the 19s regulatory particle (which recognizes and imports proteins for degradation) and the 20s core (which cleaves peptides) of the proteasome are up-regulated in response to mutational load in TCGA (Fig. 3A). In addition, we find that specifically mitochondrial — but not cytoplasmic — ribosome complexes are up-regulated in high mutational load tumors. As previously reported in yeast[27] and human cells[28], mitochondrial ribosome biogenesis has been shown to occur under conditions of chronic protein misfolding as a mechanism of compartmentalization and degradation of proteins. In contrast, translation of proteins through cytosolic ribosome biogenesis has been previously characterized to be attenuated and slowed to prevent further protein mis-folding[29]. This decrease in expression of cytoplasmic ribosomes is also consistent with observed patterns of alternative splicing coupled to mRNA decay pathways in high mutational load tumors (Fig. 2B).

Finally, we performed a jackknife re-sampling procedure to confirm that specific cancer types aren't driving patterns of association within the GLMM. This was achieved by removing each cancer type from the regression model one at a time, and re-calculating regression coefficients on the remaining set of samples. Overall, regression coefficients were stable across cancer types and trends were unchanged (Supplemental Figure 5). In addition, we also performed linear regression within cancer types and found similar expression responses to mutational load across proteostasis complexes (Supplemental Figure 6). Finally, we also confirmed that patient age was not driving patterns of association of mutational load and gene expression within the GLMM (Supplemental Figure 7). Taken together, this suggests that protein re-folding, protein disaggregation, protein degradation, and down-regulation of cytoplasmic translation are potential mechanisms to mitigate and prevent protein misfolding in high mutational load tumors.

Validating proteostasis expression responses in cancer cell lines and establishing a causal connection through perturbation experiments.

We next sought to validate these results by first examining whether the expression patterns observed in human tumors replicate within cancer cell lines from the Cancer Cell Line Encyclopedia (CCLE)[30]. Unlike TCGA, samples within each cancer type in CCLE can be small and are unbalanced (i.e., some cancer types have <10 samples and others have >100 samples). Since GLMMs may not be able to estimate among-population variance accurately in these cases[31], we utilized a simple generalized linear model (GLM) instead to measure the effect of mutational load on patterns of expression without over-constraining the model. Indeed, we find that expression patterns seen in human tumors broadly replicate in cancer cell lines (Fig. 3). Similar to the expression analysis in TCGA, we also confirmed through a jackknife re-sampling procedure that specific cancer types aren't driving patterns of association within the GLM (Supplemental Figure 8). Finally, we further validated these trends by

incorporating protein abundance estimates in CCLE, which contains the largest dataset available of RNA (n=1377) and protein (n=373) abundances that are harmonized across samples. We find similar patterns of expression and protein abundances in response to mutational load in CCLE within proteostasis complexes (Supplemental Figure 9).

Overall, this indicated that the expression patterns observed are cell autonomous (i.e., independent of organismal effects such as the immune system, age or microenvironment) and consistent across high mutational load cancer cells. Importantly, it also demonstrates that cancer cell lines are a reasonable model to causally interrogate these effects further through functional and pharmacological perturbation experiments.

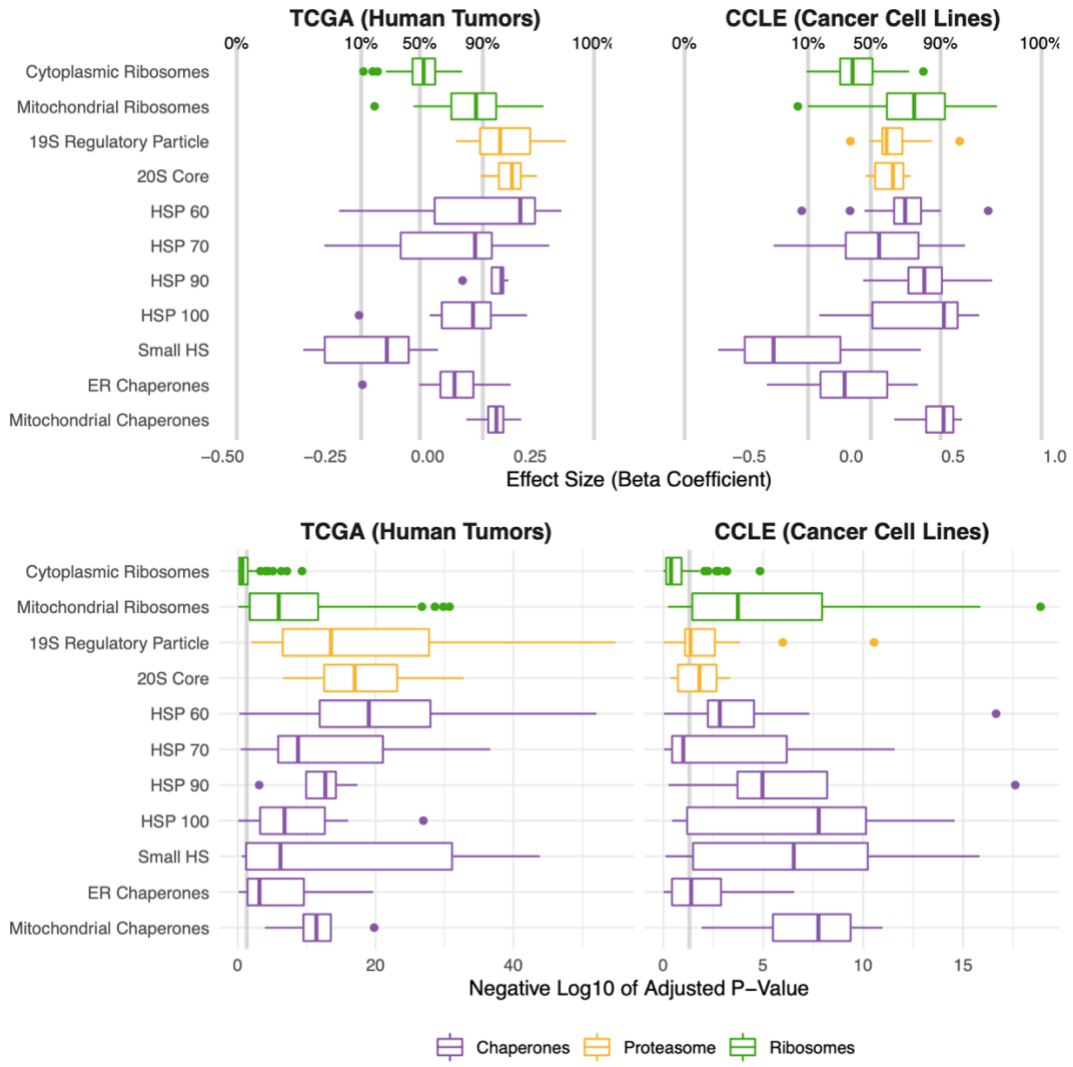


Figure 3. Protein folding, degradation, and synthesis are regulated in both high mutational load tumors (TCGA) and cell lines (CCLE). Box plots of β_1 regression coefficients (top panels) and negative $\log_{10}$ adjusted *p*-values (bottom panels) measuring the association of mutation load and the expression of individual genes in chaperone (purple), proteasome (yellow), and ribosome (green) complexes. Shown are regression coefficients from human tumors (TCGA) on the left and cell lines (CCLE) on the right. Percentages and grey lines on top panels show the quantile distribution of regression coefficients

measuring the association of mutational load and expression for all genes in the genome within each dataset. Vertical grey line on bottom panels shows threshold of significance (p = 0.05).

To establish a causal relationship between the over-expression of proteostasis machinery and maintenance of cell viability under high mutational load, we utilized expression knock-down (shRNA) estimates from project Achilles[32] for the same cancer cell lines as in CCLE. We sought to measure how mutational load impacts cell viability when protein complexes and gene families undergo a loss of function through expression knock-down. Since the shRNA screen was performed on an individual gene basis, we utilized a GLM framework that aggregates expression knock-down estimates of all genes within a given proteostasis gene family to jointly measure how mutational load impacts cell viability after loss of function. Specifically, we included an additional categorical variable of the gene name within each gene family to allow for a change in the intercept within each gene in the GLM when measuring the association of mutational load and cell viability after expression knock-down. In addition, we similarly evaluated whether specific cancer types were driving patterns of association within the GLM through jackknife re-sampling by cancer type (Fig. 4A).

Overall, we find that elevated mutational load is associated with decreased cell viability when the function of most chaperone gene families are disrupted through expression knock-down (Fig. 4A). However, only chaperones within the HSP100 family, which have the unique ability to rescue and reactivate existing protein aggregates in cooperation with other chaperone families[33], show a significant negative relationship between mutational load and cell viability across almost all cancer types. Similarly, we find specificity in the vulnerability that mutational load generates when the function of the proteasome and different ribosomal complexes are disrupted (Fig. 4A). Mutational load significantly decreases cell viability only when expression knock-down of the 19s regulatory particle of the proteasome is disrupted, suggesting that targeting the protein import machinery of the proteasome is more effective than targeting the protein cleaving machinery in the 20s core. Finally, mutational load significantly increases cell viability when cytoplasmic ribosomes – which are already down-regulated in response to mutational load (Fig. 2B) – undergo a loss of function through expression knock-down. Conversely, expression knock-down of mitochondrial ribosomes significantly decreases viability with increased mutational load in cell lines, which is also consistent with the patterns of expression observed.

Since functional redundancy in the human genome can make expression knock-down estimates within individual genes noisy, we also examined how drugs targeting the function of whole complexes impacts viability with mutational load across all cancer types and when removing individual cancer types through jackknife re-sampling. To do so, we utilized drug sensitivity screening data in project PRISM[34] within CCLE and used a simple GLM to measure the association of mutational load and cell viability after drug inhibition. We find that treatment with the majority of proteasome inhibitors (6/8) and ubiquitin-specific proteasome inhibitors (2/3), which target protein degradation complexes, are significantly associated with a decrease in cell viability in high mutational load cell lines. Similarly, most HSP90 inhibitors decrease cell viability with mutational load (8/10), although only a few drugs show a significant relationship. This variability in the efficacy of drugs with similar mechanisms of action likely reflects that

the efficacy to disrupt the function of proteostasis machinery is dependent on the specific molecular affinity of a compound to its target and downstream effectors. While these are the only relevant proteostasis drugs in the PRISM dataset that are currently available, we anticipate that drugs targeting other chaperone machinery or splicing complexes could also target other potential vulnerabilities in high mutational load cancers. Collectively, these results indicate that elevated expression of protein degradation and folding machinery is causally related to the maintenance of viability in in high mutational load cell lines, and likely in high mutational load tumors by extension.

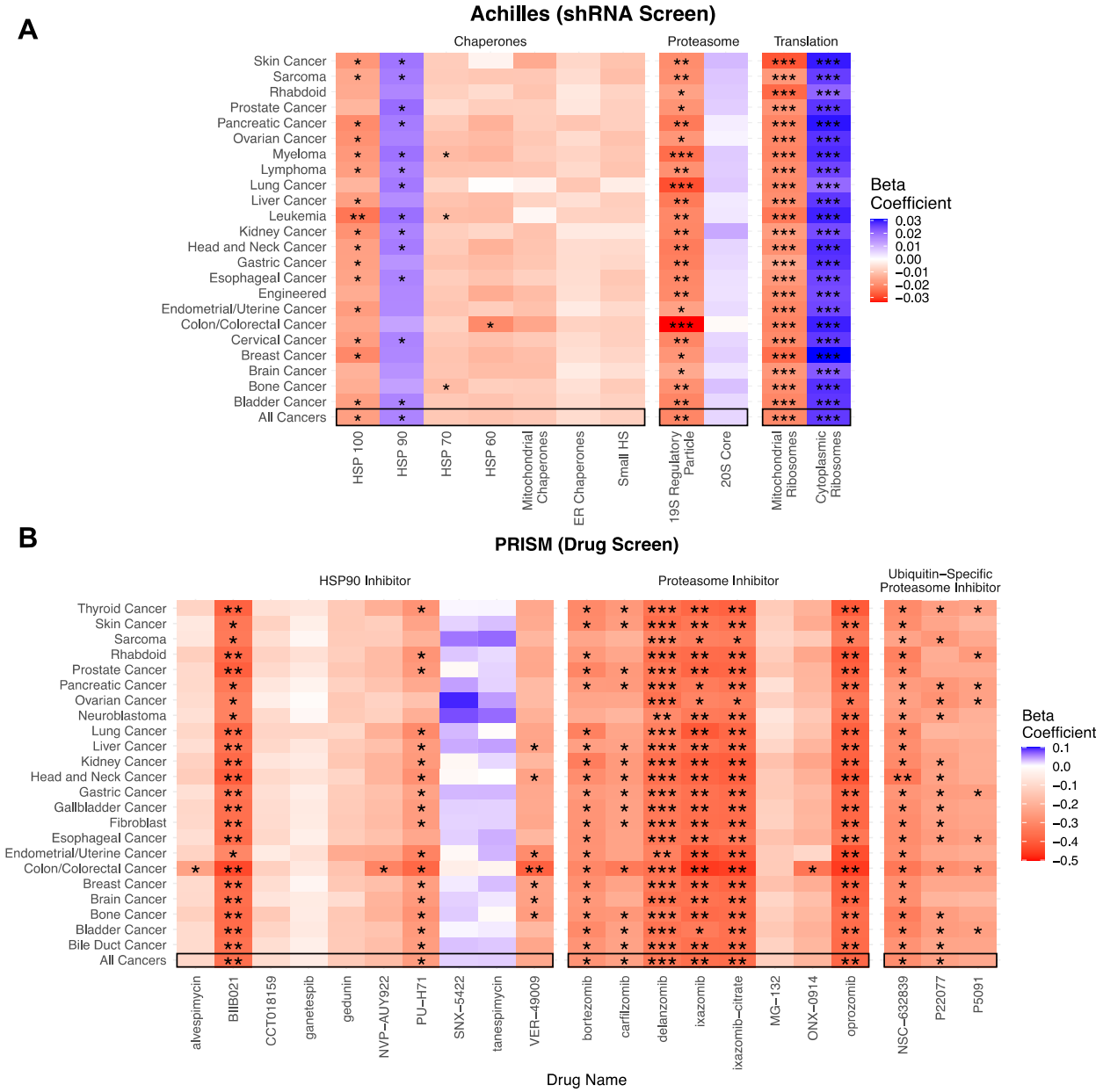


Figure 4. Viability in high mutational load cell lines decreases when proteostasis machinery is disrupted. (A) Heatmap of β_1 regression coefficients jointly measuring the association of mutational load and cell viability after expression knockdown of individual genes in proteostasis complexes. **(B)** Heatmap of β_1 regression coefficients measuring the association mutational load and cell viability after inhibition of proteostasis machinery via drugs. Both panels show how stable regression estimates are when including

all cancer types ('All Cancers') shown in black boxes and when removing each individual cancer type on the y-axis. Colors denote a positive (blue), zero (grey), or negative (red) relationship of mutational load and cell viability after expression knock-down or drug inhibition. Stars denote whether the relationship is significant (* = $p < 0.05$; ** = $p < 0.005$; *** = $p < 0.0005$).

Lastly, we find that most drugs in the PRISM database do not significantly decrease cell viability with mutational load (Fig. 5A), suggesting that high mutational load cancer cells are not generically vulnerable to all classes of drugs. Specifically, we find that drugs which inhibit transcription, cytoskeleton organization, protein degradation, chaperones, protein synthesis and promote apoptosis are most effective at targeting high mutational load cancer cells – delineating additional potential therapeutic vulnerabilities in high mutational burden tumors (Fig. 5B).

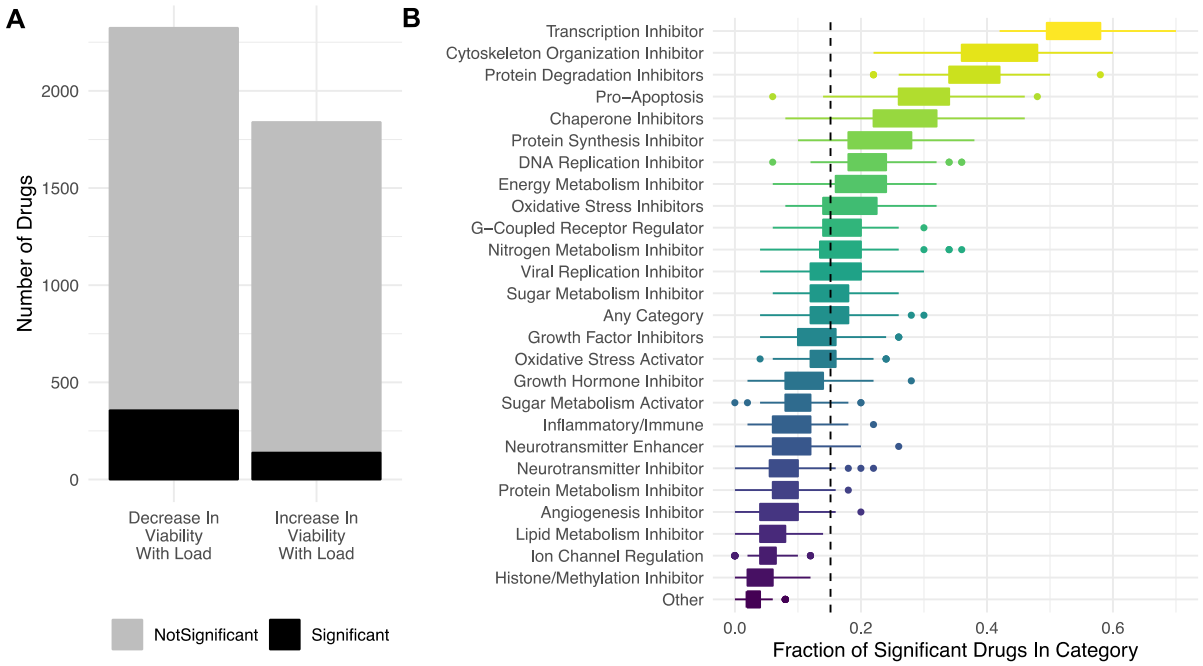


Figure 5. Targeting proteostasis machinery is a key vulnerability in high mutational load cell lines. (A) Bar plot of the number of drugs in the PRISM database significantly (black) and not significantly (grey) associated with mutational load and cell viability using a simple generalized linear model (GLM). **(B)** Fraction of drugs in broad functional categories significantly negatively associated with mutational load and cell viability from the GLM. Confidence intervals were determined by randomly sampling 50 drugs in each functional category 100 times. Dashed line is the median of randomly sampled drugs across all categories.

Discussion

Here, we test the hypothesis that cancer cells regulate their proteostasis machinery to mitigate the damaging effects of passenger mutations, which can destabilize and misfold proteins. Misfolded proteins can arise from non-synonymous or nonsense passengers which cause abnormal amino acid modifications or pre-mature truncations in proteins. Even synonymous passengers, which are traditionally thought to be functionally silent, can lead to misfolding of proteins through changes in mRNA stability[35], translational pausing[36,37], and non-optimal codon usage.[38,39] As a result, protein misfolding can be damaging in cells not only due to a loss of function of the original protein, but also due to a gain in toxicity caused by the aggregation of aberrant peptides. It is intriguing to consider the possibility that the need to manage protein misfolding stress is a hallmark of somatic evolution in cancer.

To maintain viability by minimizing these cytotoxic effects, we find that high mutational load tumors – similar to yeast[40], bacteria[41,42], and viruses[43] – up-regulate the expression of chaperones, which allow mutated proteins that would otherwise be misfolded to retain function. We find evidence suggesting that specific chaperone families that actively participate in protein re-folding (HSP60, HSP90 and HSP70) or disaggregation (HSP100) are up-regulated in response to mutational load, while other chaperone machinery that salvage proteins (Small HS) are downregulated. In addition, we find degradation of mutated proteins through up-regulation of the proteasome to be another possible strategy high mutational load tumors use to mitigate protein misfolding stress.

Finally, we find additional mechanisms that high mutational load tumors use to not just mitigate but also prevent protein misfolding. By utilizing post-transcriptional processes that couple alternative splicing with mRNA decay pathways known to occur in normal human tissues[22,44,45], high mutational load tumors appear to selectively prevent protein production by regulating certain pre-mRNA transcripts to be degraded rather than translated. We find evidence suggesting that the targets of this coordinated un-productive splicing are primarily related to cytoplasmic ribosomal gene expression that controls the translation of proteins, consistent with observations in other organisms[46–48]. Intriguingly, we find that while cytoplasmic ribosome expression is attenuated, mitochondrial ribosome biogenesis in human tumors is up-regulated in response to mutational load. This could both be another mechanism that high mutational load tumors use to compartmentalize and degrade proteins[27] and reflect the increased energetic demands of proteostasis maintenance[49].

The expression responses observed here are not only consistent with protein misfolding stress in other organisms, but also cross-validate in cancer cell lines, where we find similar expression responses to mutational load. This provides further evidence of a generic, cell intrinsic phenomenon occurring that cannot be explained by extrinsic organismal effects, such as aging, changes in the immune system or microenvironment. Furthermore, we move beyond correlations of gene expression responses to mutational load and establish a causal connection by demonstrating that mitigation of protein misfolding through protein degradation and re-folding is necessary for high mutational

load cancer cells to maintain viability through perturbation experiments via knockdown experiments with shRNA and drug profiling.

The results presented here have many implications. First, they suggest that while there is direct selection during somatic evolution for pathogenic drivers that allow cancer cells to continually proliferate, damaging passengers that destabilize proteins must also cause cancer cells to experience second-order indirect selection for alterations that allow tumors to overcome this proteostasis imbalance. This could occur through phenotypic plasticity, shifts in methylation and chromatin structure, or through compensatory point mutations and duplications, consistent with other studies[50,51]. Indeed, gene duplication, where one copy can still perform the required function while the other copy is non-functional, is another known mechanism that allows cells to maintain robustness to damaging mutations in many eukaryotic organisms[52,53]. In support of this, whole genome-duplication, which is common in cancer, has recently been shown as another potential mechanism that tumor cells could use to maintain robustness to deleterious passengers[54]. However, duplication events are also known to be deleterious due to gene dosage effects that cause protein imbalance[55], which could further exacerbate protein misfolding. Further experimental studies are needed to distinguish how cancer cells compensate for protein misfolding and the role that genome duplication may play in this process.

Second, the extra demands of proteostasis maintenance presents important vulnerabilities in high mutational load cancers that could be exploited. The clinical use of chaperone inhibitors for cancer treatment has been explored for over two decades[56–58] but no study, to our knowledge, has compared the efficacy of chaperone inhibitor use in tumors stratified by mutational load. Similarly, the clinical use of proteasome inhibitors, which are currently only approved for the treatment of multiple myeloma and mantle-cell lymphoma[59,60], has not been directed specifically to high mutational load tumors. While the efficacy of proteasome inhibitors in multiple myeloma patients is linked to the protein misfolding stress response[61,62] , it is currently unknown whether high mutational load tumors are more susceptible to these inhibitors. Outside of drugs in the clinic, the need for cancers to compensate for protein misfolding could also present additional vulnerabilities due to evolutionary trade-offs, where the improvement in fitness of one trait comes at the expense of another. Previous work in yeast has identified strong trade-offs between the adaptive mechanisms that allow for the tolerance of mistranslation and survival under conditions of starvation[49]. Whether similar conditions could be exploited in high mutational load cancer cells warrants additional further investigation.

Finally, our results contribute to an accumulating body of evidence that cancer and aging are different manifestations of related underlying evolutionary processes[63–65]. The same forces of mutation and inefficient selection in somatic evolution generates a persistent problem of deleterious mutation accumulation in normal somatic tissues and during tumor growth. Disruption of proteostasis is a known hallmark of aging in normal tissues[66]. Many transcriptional responses observed in high mutational load tumors — such as shifts in regulation of alternative splicing[67], protein degradation[68], and protein re-folding[69] — are also observed in normal aging tissues which contain somatic mutations. Despite this, aging tissues appear to utilize different strategies to deal with proteostasis disruption — such as up-regulation of chaperones in the Small HS family[70]

and autophagy[71] — which are not a pre-dominant response observed here in high mutational load tumors. Whether different combinations of strategies are used by high mutational load cancer cells use to overcome their mutational load or whether all the strategies identified here are needed to maintain proteostasis is unclear. Differences in these proteostasis strategies could be due to different selection pressures during somatic evolution, the degree of mutational load required to induce a stress response, differences in energetic costs of protein maintenance, or the interplay that exists between apoptosis and proteostasis. Further studies are needed to elucidate the precise dynamics and physiological consequences of inefficient negative selection in somatic evolution, how this impacts cellular growth, and the mechanisms somatic cells use to maintain robustness to proteostasis disruption.

Acknowledgements

We thank Kathleen Houlahan, Chuan Li, José Aguilar-Rodríguez and other members of Petrov and Curtis labs for their helpful comments and discussions. S.T. was supported in part by an NIH training grant T32-HG000044-21. This work was supported in part by the National Institutes of Health (NIH) Director's Pioneer Award: DP1CA238296 and the National Cancer Institute (NCI) Cancer Target Discovery and Development Center (U01CA217851) to C.C and by NIH grants R35GM118165 to D.A.P and GM74074 to J.F.

Conflicts of Interest

C.C. is an advisor and stockholder in Grail, Ravel, DeepCell and an advisor to Genentech, Bristol Myers Squibb, 3T Biosciences and NanoString. D.A.P. is a founder of, and stockholder equity in, D2G Oncology Inc.

Methods

Data availability and resources. Whole-exome, somatic mutation calls of 10,486 cancer patients across 33 cancer types in The Cancer Genome Atlas (TCGA) were downloaded from the Multi-Center Mutation Calling in Multiple Cancers (MC3) project[16] (https://gdc.cancer.gov/about-data/publications/mc3-2017). For the same patients in TCGA, RNA-seq data of $\log_2$ transformed RSEM normalized counts were downloaded from the UCSC Xena Browser[72] (https://xenabrowser.net/datapages/) and copy number alterations (CNAs), including amplifications and deletions, called via ABSOLUTE were downloaded from COSMIC (v91)[73] (https://cancer.sanger.ac.uk/cosmic/download). Tumor purity estimates for TCGA were downloaded from the Genomic Database Commons (GDC)[74] (https://gdc.cancer.gov/about-data/publications/pancanatlas). Data for all cancer cell lines in the Cancer Cell Line Encyclopedia (CCLE) were downloaded from DepMap[30] (https://depmap.org/portal/download/all/). Specifically, mutation calls (Version 21Q3) from whole-exome sequencing data, copy number alternations

quantified by ABSOLUTE (Version CCLE 2019), $\log_2$ transformed TPM normalized counts (Version 21Q3) from RNA-seq data, proteomics data quantified by mass spectrometry[75], shRNA data from project Achilles[32] normalized using DEMETER (DEMETER2 Data v6), and primary drug sensitivity screens of replicate collapsed log fold changes relative to DMSO from project PRISM[34] (Version 19Q4) were used.

Statistical analysis. The lmerTest and lmer package in R was used to apply a separate generalized linear mixed model (GLMM) for each gene in the genome to identify groups of genes whose expression is up-regulated in response to mutational load in TCGA. For each gene, expression values across all patients were z-score normalized in all analyses to ensure fair comparisons across genes. Known co-variates of tumor purity and cancer type were included in the GLMM. Tumor purity and mutational load were modeled as fixed effects, whereas cancer type was modeled as a random effect (i.e. random intercept) since it varies across groups of patients and can be interpreted as repeated measurements across groups. For all analyses, mutational load was defined as $\log_{10}$ of the number of synonymous, nonsynonymous and nonsense mutations per tumor. For each gene, the parameters used in the GLMM were as follows,

$$Y \sim \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \mathrm{v} + e$$

where Y is a vector of expression values of each tumor, β_0 is the fixed intercept, β_1 is the fixed slope for the predictor variable X_1 which is a vector of mutational load values for each tumor, β_2 is the fixed slope for the predictor variable X_2 which is a vector of the purity of each tumor, v is the random intercept for each cancer type, and e is a Gaussian error term. To examine expression responses to mutational load within a given protein complex and cancer type, the same normalization procedures were applied as above within cancer types and a separate GLM for each cancer type was ran as follows,

$$Y \sim \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e$$

where Y is a vector of expression values of each tumor in a given cancer type, β_0 is the fixed intercept, β_1 is the fixed slope for the predictor variable X_1 which is a vector of mutational load values for each tumor, β_2 is the fixed slope for the predictor variable X_2 which is a vector of the purity of each tumor, β_3 is a change in the intercept for X_3 which is a categorical variable of individual genes within each proteostasis complex and e is a Gaussian error term.

Unlike TCGA, samples within each cancer type in CCLE can be small and are unbalanced (i.e. some cancer types have <10 samples and others have >100 samples). In these cases, mixed effects models may not be able to estimate among-population variance accurately[31]. Thus, for all regression-based analyses in CCLE, a simple generalized linear model (GLM) was used instead. Cell viability values across all cell lines were z-score normalized by gene in all analyses to ensure fair comparisons across genes. To assess whether the same sets of genes are up-regulated in response to

mutational load in CCLE using the GLM, a similar procedure to the GLMM was performed. A separate GLM was applied for each gene with the following parameters,

$$Y \sim \beta_0 + \beta_1 X_1 + e$$

where Y is a vector normalized expression values of each cell line, β_0 is the fixed intercept, β_1 is the fixed slope for the predictor variable X_1 which is a vector of mutational load values for each tumor, and e is a Gaussian error term. To assess whether protein abundances are similarly up-regulated in response to mutational load in CCLE in proteostasis complexes, a separate GLM was applied to each gene with the following parameters,

$$Y \sim \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e$$

where Y is a vector of protein abundance values within each cell line, β_0 is the fixed intercept, β_1 is the fixed slope for the predictor variable X_1 which is a vector of mutational load values for each tumor, and e is a Gaussian error term. A similar GLM framework as above was used to estimate the association of mutational load and cell viability after shRNA knock-down of individual genes in proteostasis complexes with the following parameters,

$$Y \sim \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e$$

where Y is a vector of normalized cell viability estimates after expression knock-down of an individual gene across all cell lines, β_0 is the fixed reference intercept, β_1 is the fixed slope for the predictor variable X_1 which is a vector of mutational load values for each cell line, β_2 is a change in the intercept for X_2 which is a categorical variable of individual genes within each proteostasis complex, and e is a Gaussian error term. To estimate the association of mutational load and cell viability after pharmacologic inhibition of proteostasis machinery, the following GLM was applied to each relevant drug in PRISM:

$$Y \sim \beta_0 + \beta_1 X_1 + e$$

where Y is a vector normalized cell viability estimates after drug inhibition across all cell ines, β_0 is the fixed intercept, β_1 is the fixed slope for the predictor variable X_1 which is a vector of mutational load values for each tumor, and e is a Gaussian error term.

Model validation. For both the GLM and GLMM, model assumptions of homogeneity of variance were verified by plotting residuals versus fitted values in the model and residuals versus each covariate in the model. Multi-collinearity with other mutational classes (e.g. such as copy number alterations, CNAs) were considered but not found to correlate with point mutations (Supplemental Figure 1). A jackknife re-sampling procedure was used for outlier analysis and to determine whether specific cancer types are driving patterns of association within the GLM and GLMM. Briefly, each cancer type was removed from the regression model one at a time, and regression coefficients were re-estimated. Overall, regression coefficients were fairly stable across cancer types and trends remained the same (Supplemental Figure 5 and 8).

Proteostasis gene sets. Genes for chaperone complexes were identified from[76] and genes that are co-chaperones were not considered. Proteasome and ribosomal complexes were identified from CORUM[17].

Gene set enrichment analysis. All gene set enrichment analysis was performed using gprofiler2 with default parameters. For all sets of genes, significance was determined after correcting for multiple hypothesis testing (FDR < 0.05). For gene set enrichment analysis used to identify genes up-regulated in TCGA in response to mutational load, all terms in CORUM database were reported and enrichment terms in the KEGG database of diseases not related to cancer (e.g. 'Influenza A') were omitted from the main figures for clarity and space. For gene sets used to identify terms differentially splice in between high and low mutational load tumors, all terms in the CORUM and the REACTOME database were reported in the main figures. The full set of enrichment terms for all analyses is reported in Supplemental Table 1.

Alternative splicing analysis. Alternative splicing events were quantified through a previously established metric called PSI. PSI is calculated as the number of reads that overlap the alternative splicing event (e.g. for intron retention, either at intronic regions or those at the boundary of exon to intron junctions) divided by the total number of reads that support and don't support the alternative splicing event. PSI summarizes alternative splicing events at specific exonic boundaries in the entire transcript population without needing to know the complete underlying composition of each full length-transcript.

Alternative splicing calls for all tumors in TCGA were downloaded from TCGA SpliceSeq[25]. Default splice event filters (percentage of samples with PSI values >75%) from the database were applied. To test whether gene expression is down-regulated in high mutational load tumors through alternative splicing, we calculated whether alternative splicing events were present based on different threshold values of percent spliced in (PSI) from 90% to 50%. (Supplemental Figure 4). For each alternative splicing event in a gene, we quantified whether the expression of the same gene was under-expressed. Each gene was counted as under-expressed if it was one standard deviation below the mean expression within each cancer type. Genes that contained a point mutation within the same alternative splicing event were removed to control for eQTL effects. We note that intron retention events removed from this analysis represent only ~1% of intron retention events across all tumors and similar trends are found when this filtering scheme is not applied (Supplemental Figure 3). In addition, we evaluated whether this trend is robust to other alternative splicing events (i.e., Alternate Donor Sites, Alternate Promoters, Alternate Terminators, Exon Skipping Events, ME=Mutually Exclusive Exon; Supplemental Figure 4).

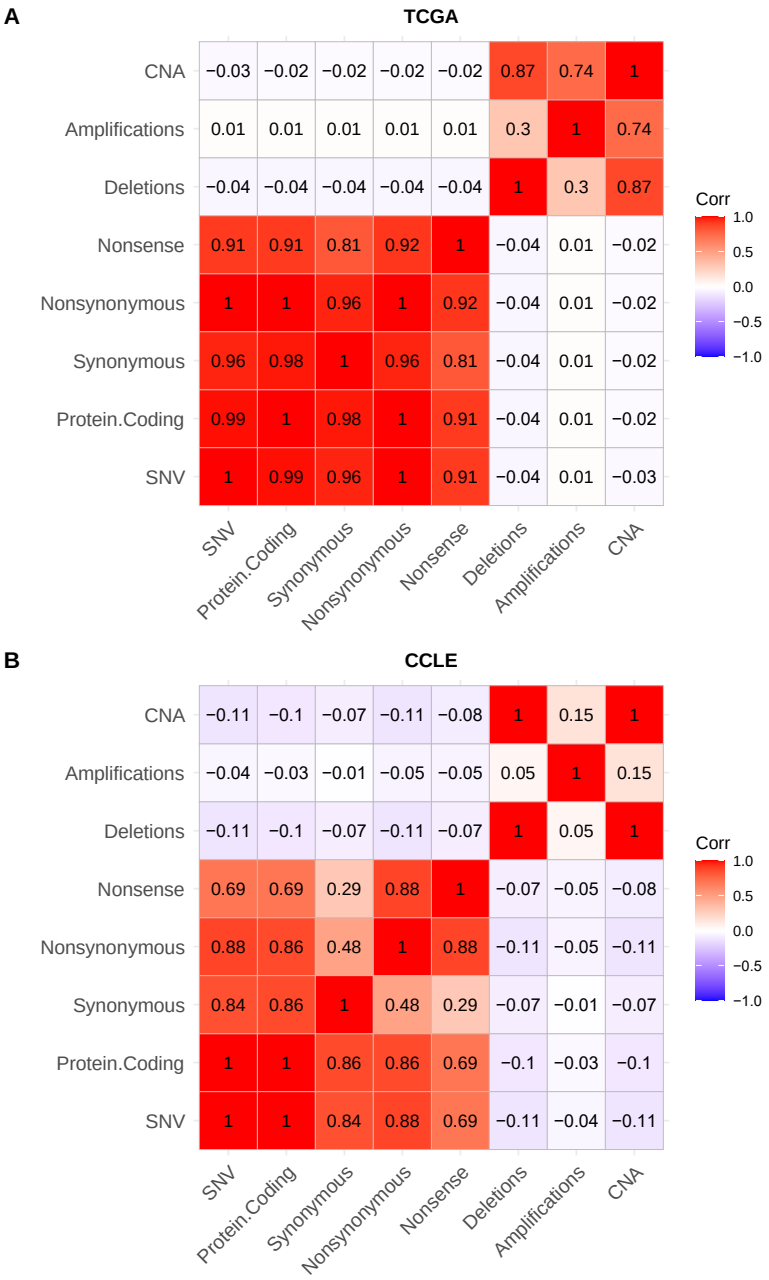
To investigate which genes are differentially spliced in between low and high mutational load tumors for specific alternative splicing events (i.e. intron retention), a t-test was used to calculate whether PSI values were significantly different in tumors with < 10 protein-coding mutations compared to tumors with > 1000 protein-coding mutations. Each alternative splicing event within a gene was required to have less than

25% of missing PSI values and a mean difference between the two groups of >0.01 to be considered. This approach identified 606 and 201 significant genes that have more and fewer intron retention events in high mutational load tumors, respectively, after correcting for multiple hypothesis testing (FDR < 0.05).

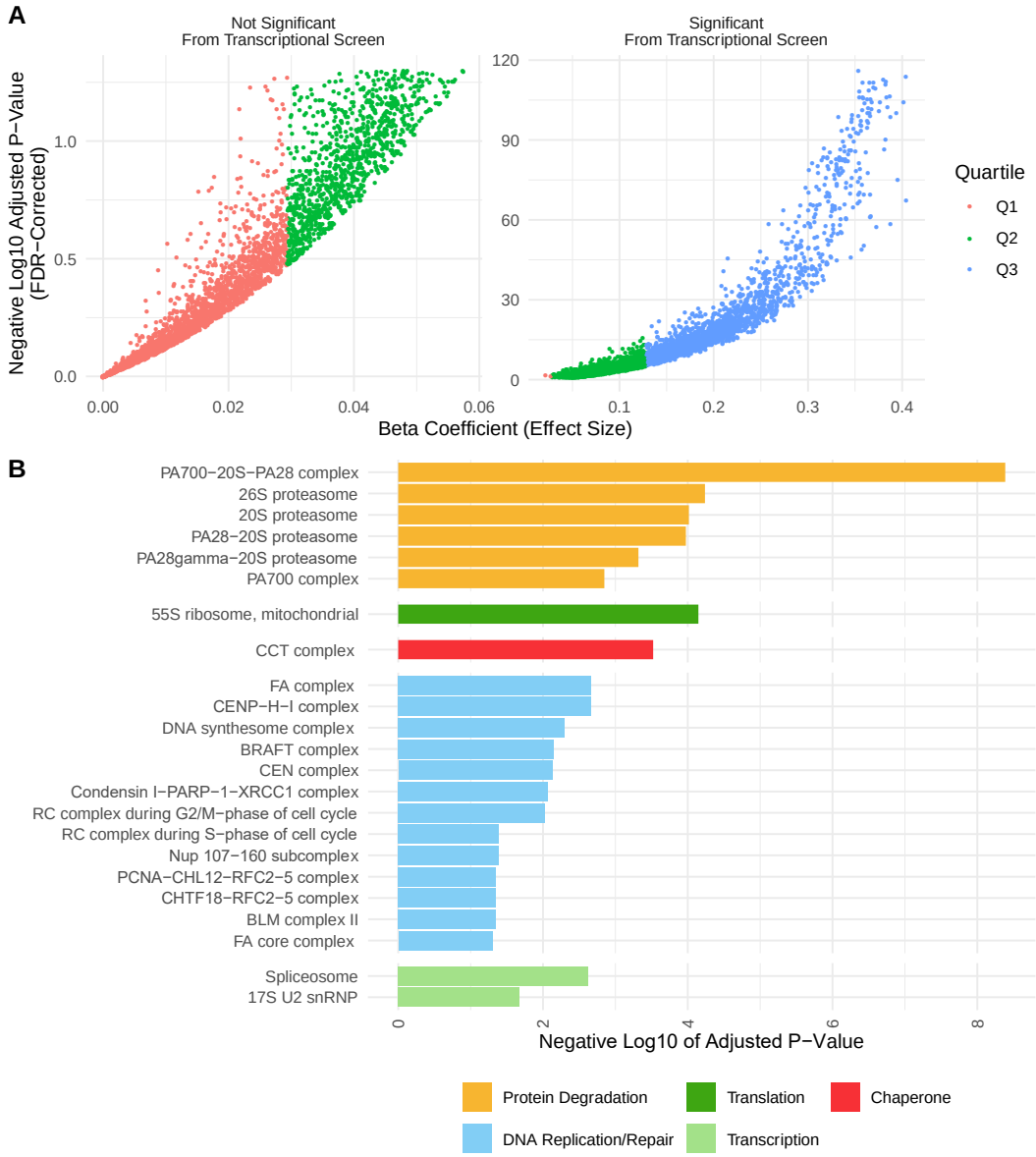
Drug category annotation and enrichment analysis. A separate GLM was ran for all drugs in the PRISM database to evaluate whether they are associated with mutational load and cell viability. All drugs that were negatively associated with mutational load and viability were queried on PubMed based on their reported mechanism of action in PRISM and grouped into broad categories (Supplemental Table 1). Categories of drug mechanism of action were first chosen based on their role in metabolism and known hallmarks of cancer. Additional categories not directly related to known cancer associated functional groups were made for drugs that could not otherwise be grouped (i.e. ‘Ion Channel Regulation’, Viral Replication Inhibitor', etc.). Drugs with ambiguous mechanism of action (e.g. 'cosmetic', 'coloring agent') were grouped into ‘Other’. The abstracts of up to 10 associated papers were used to examine for evidence connecting drug mechanisms of action to 33 broad categories. In total, 700 drug mechanism of action were grouped and annotated into 33 broad categories. These broad categories were used to assess whether high mutational load cancer cell lines are generically vulnerable to drugs or whether certain categories are more likely to contain drugs effective against high mutational load cell lines. To control for differences in the number of drugs within each category, 50 drugs were randomly sampled, and the fraction of drugs significantly associated with mutational load in each category was calculated 100 times to generate confidence intervals.

Code and software availability. All code used for analysis will be made publicly available on Github under the open-source MIT License upon publication.

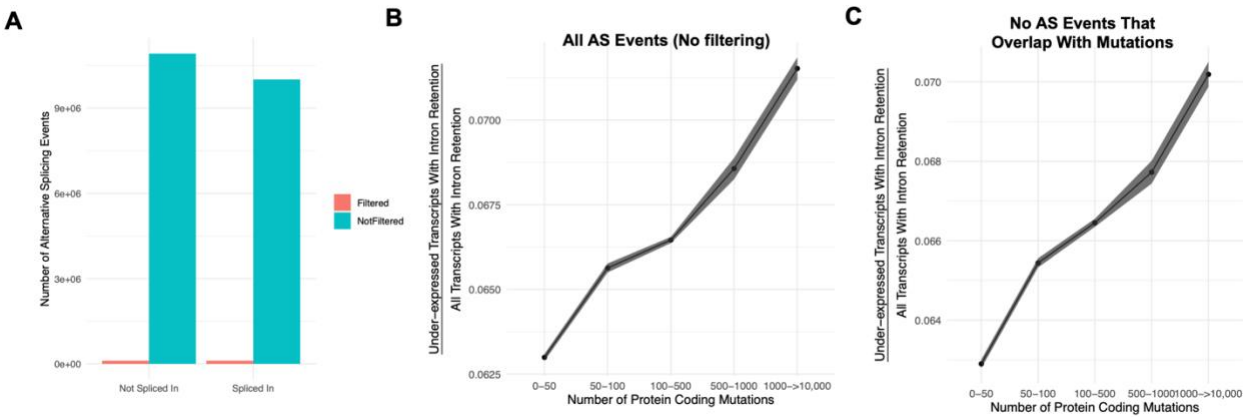
Supplemental Figures



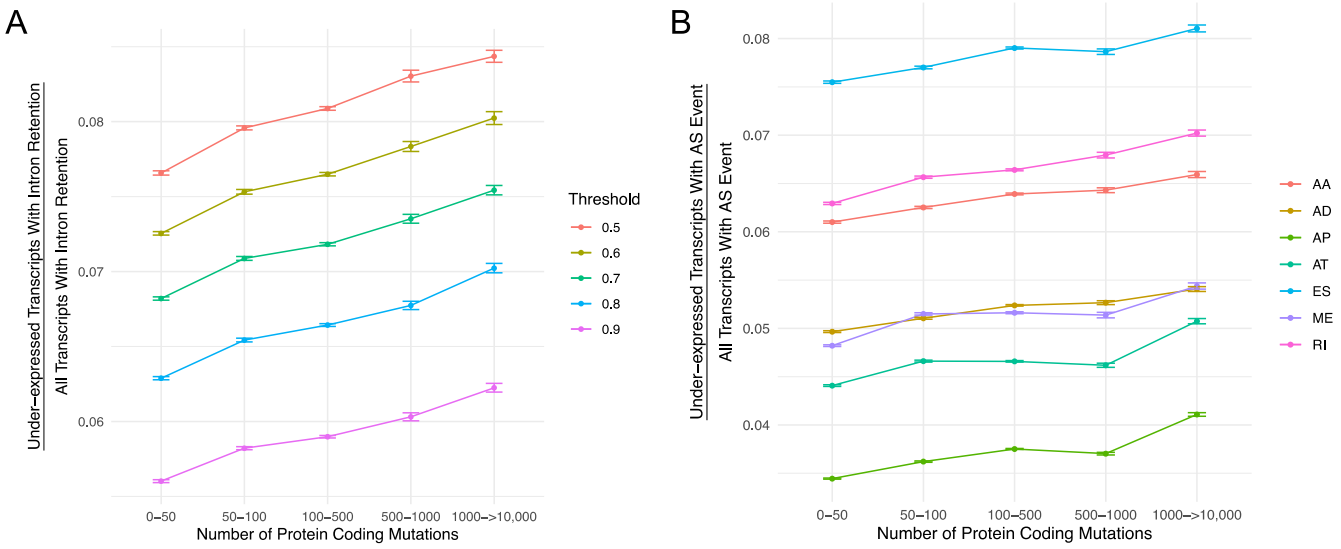
Supplemental Figure 1. No collinearity of point mutations and copy number alterations in human tumors (TCGA) and cancer cell lines (CCLE). Heatmap of Pearson's correlation coefficients between different classes of mutations in **A.** CCLE (cancer cell lines) and **B.** TCGA (human tumors). Colors denote magnitude of correlation coefficients and whether the relationship is positive (red), negative (blue) or negligible (white). CNAs are defined as the combined number of amplifications and deletions, while SNVs are the combined number of all point mutations.



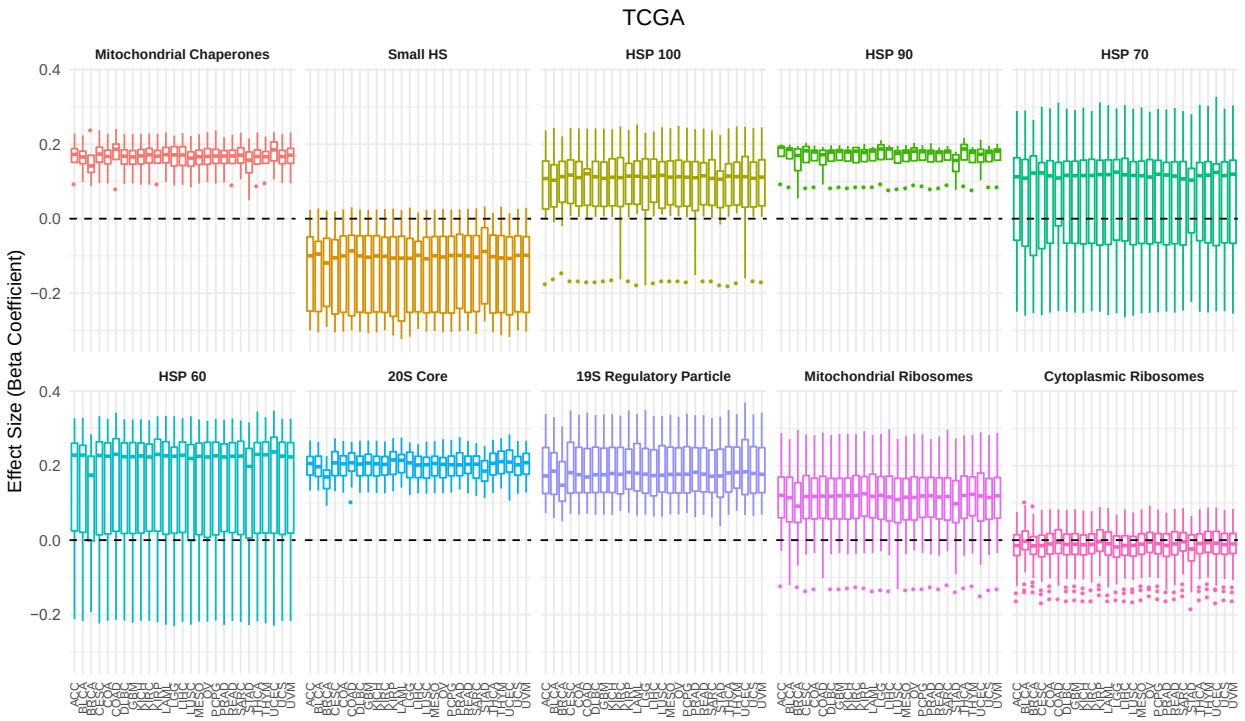
Supplemental Figure 2. Genes significantly expressed from the transcriptional screen mostly fall into the upper quartile of effect sizes, which are enriched for proteostasis complexes. A. Volcano plot of positive $\boldsymbol{\beta}_1$ regression coefficients and negative $\log_{10}$ adjusted *p*-values measuring the association of mutation load and the expression of individual genes from the transcriptional screen in Fig.1A. Colors denote the lower (Q1 in red), median (Q2 in green) and upper (Q3 in red) quartiles of each positive beta coefficient from the regression model. Genes that are significantly expressed from the transcriptional screen mostly fall into the upper quartile. **B.** Barplot of significant protein complexes in the CORUM database identified using gene set enrichment analysis only on genes that fall into the upper quartile of effect sizes. Genes in the upper quartile of effect sizes contain half of the genes that were identified as significant previously (n=2,152 vs n=5,330), yet still identify protein degradation, translation and chaperones as the top significant protein complexes.



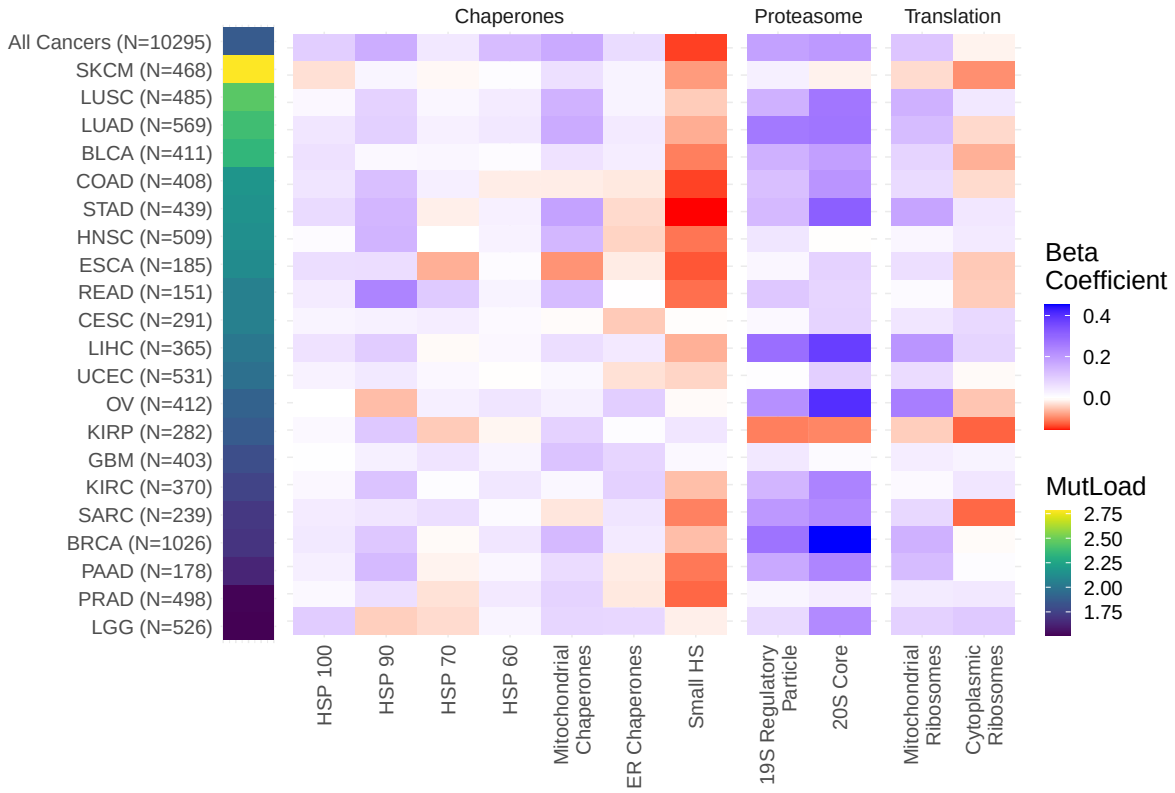
Supplemental Figure 3. Intron retention events that overlap with mutations do not account for the association of gene silencing in high mutational load tumors. A. Counts of the number of intron retention events filtered (in red) due to overlap with a mutation present in the same gene (and thus corresponding to potential eQTLs) compared the number of remaining alternative splicing events with no overlap with a mutation (in blue). Alternative splicing events filtered represent ~1% of all alternative splicing events across all tumors. **B-C.** Counts of the number of under-expressed transcripts with intron retention events, relative to counts of all intron retention events in tumors binned by the total number of protein-coding mutations. Shown are when trends when **(B)** not filtering alternative splicing events due to overlap with mutations and **(C)** when events are filtered (same as Fig. 2A). Intron retention events with PSI > 80% are counted. Error bars are 95% confidence intervals determined by bootstrap sampling. These results further support the prediction that gene silencing is elevated in high mutational load tumors and likely mediated by the coupling of intron retention with mRNA decay



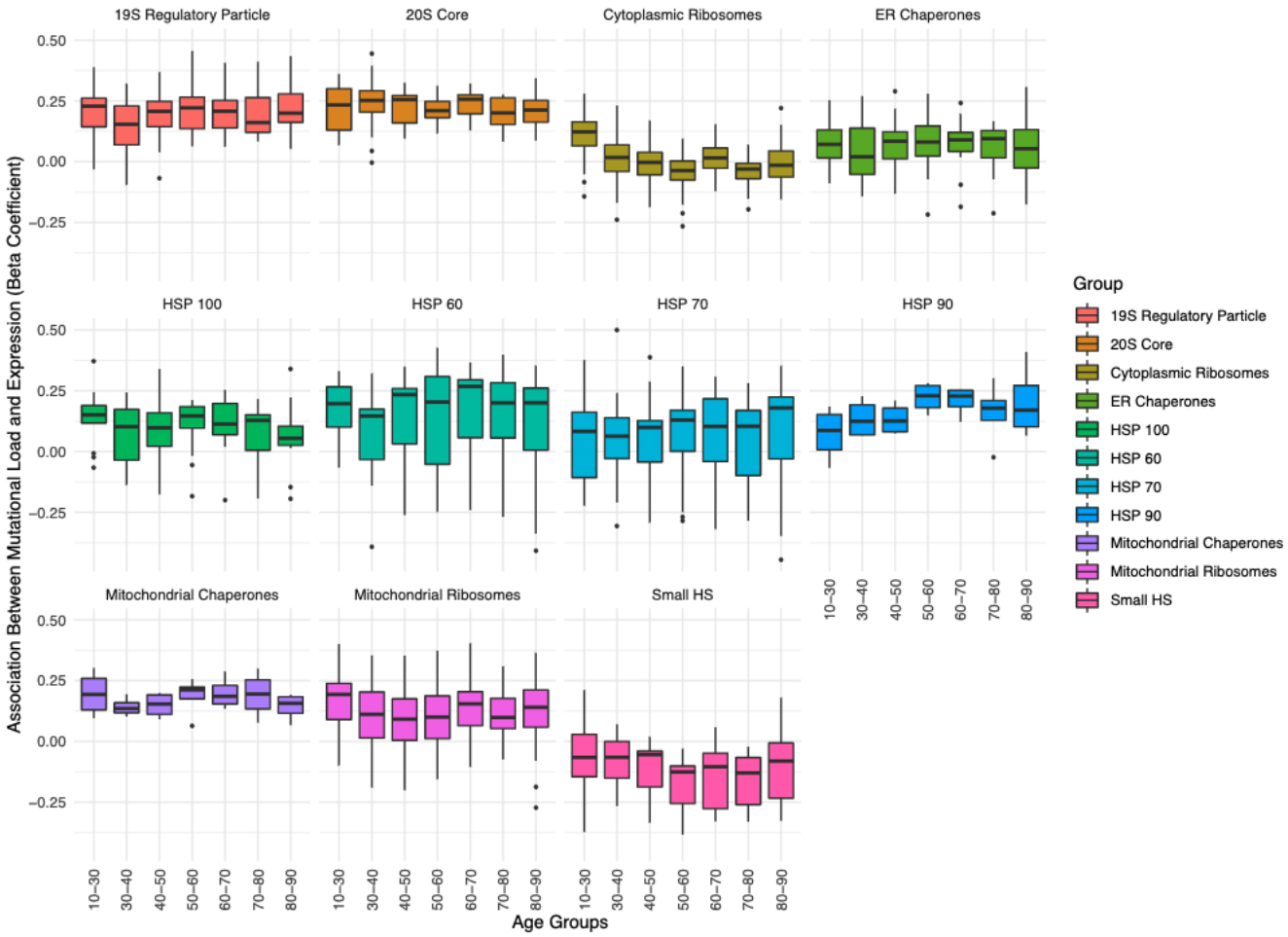
Supplemental Figure 4. The number of under-expressed transcripts increases with the mutational load of tumors for different PSI value thresholds and alternative splicing events. A. Counts of the number of under-expressed transcripts with intron retention events, relative to counts of all intron retention events in tumors binned by the total number of protein-coding mutations. Intron retention events with different PSI thresholds are shown colored. **B.** Counts of the number of under-expressed transcripts that contain different classes alternative splicing events, relative to counts of all alternative splicing events of the same class in tumors binned by the total number of protein-coding mutations. Alternative splicing events of different classes are shown colored (AA=Alternate Acceptor Sites, AD=Alternate Donor Sites, AP=Alternate Promoter, AT=Alternate Terminator, ES=Exon Skip, ME=Mutually Exclusive Exons, RI= Retained Intron). Error bars are 95% confidence intervals determined by bootstrap sampling.



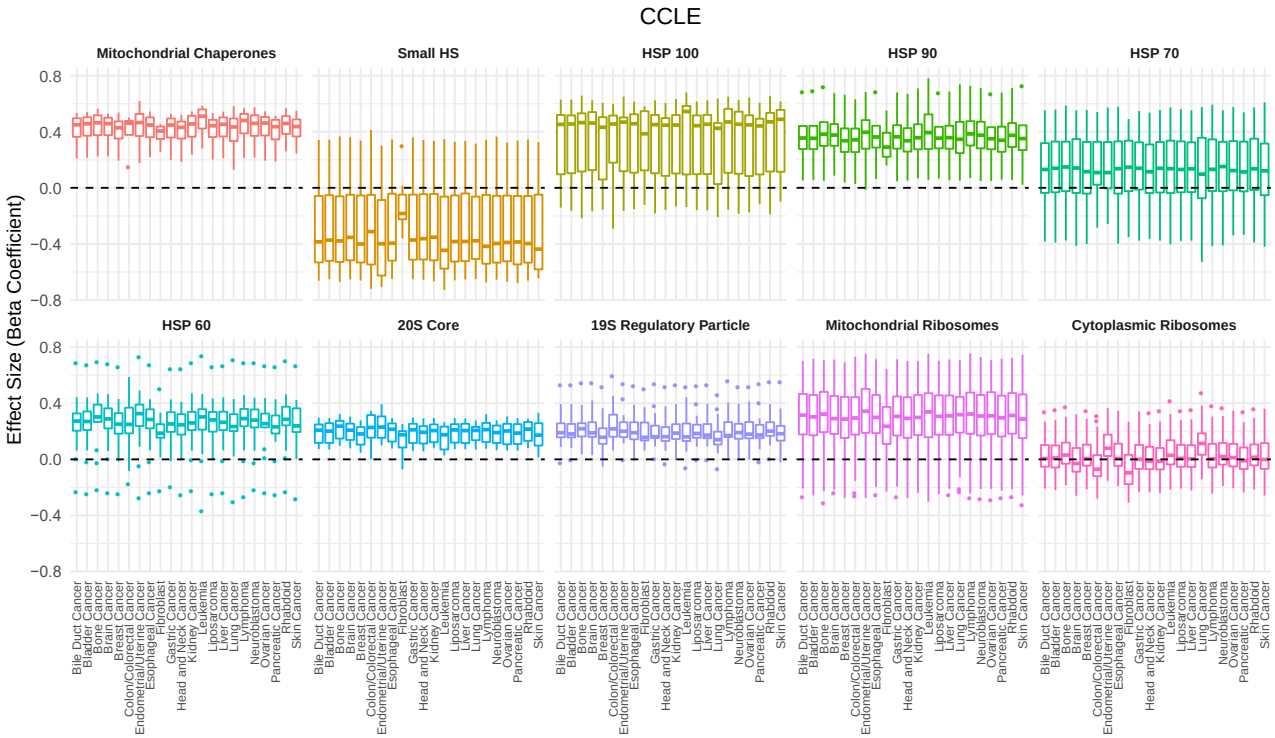
Supplemental Figure 5. Association between expression in proteostasis complexes and mutational load is not driven by a single cancer type in TCGA. Box plots of regression coefficients from the GLMM measuring the association of the expression of each individual gene with the mutational load of tumors in TCGA colored by different proteostasis complexes. Shown are regression estimates after removing each individual cancer type (x-axis) and re-running the GLMM.



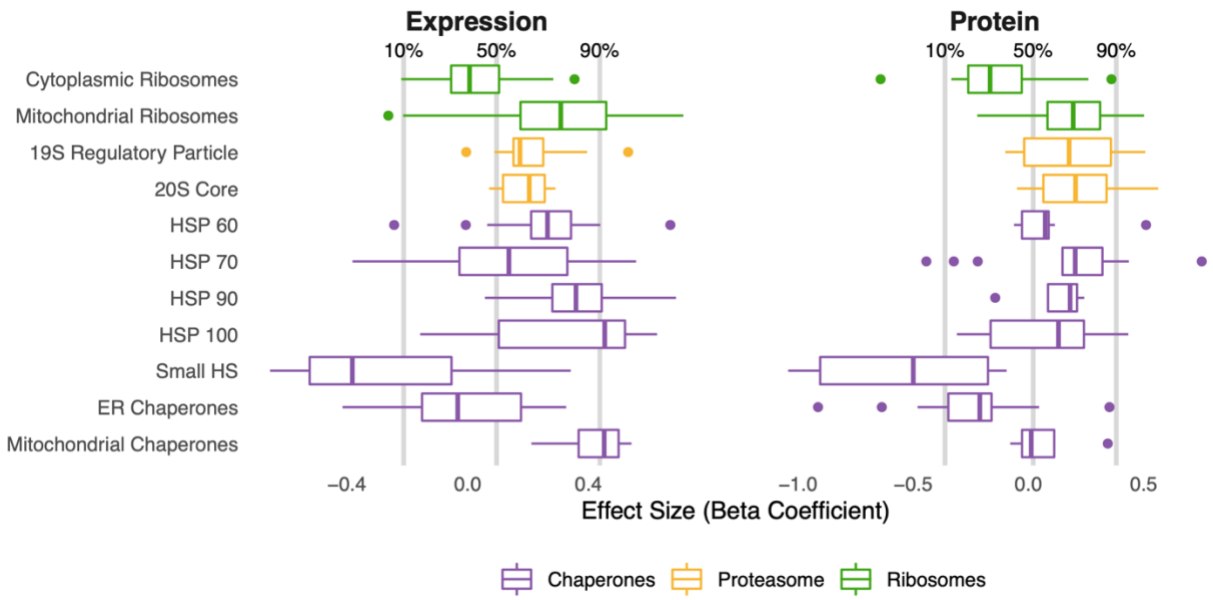
Supplemental Figure 6. Linear regression analysis within cancer types in TCGA captures similar expression responses to mutational load across proteostasis complexes. Heatmap of $\boldsymbol{\beta}_1$ regression coefficients measuring the effect of mutational load on gene expression in proteostasis complexes while controlling for tumor purity within cancer types which have enough samples to accurately measure effect sizes (N>150) and contain a sufficiently large enough mutational load to potentially generate a proteostasis response (median protein coding mutations > 25). 'MutLoad' shows log10 of the median number of protein coding mutations for each cancer type.



Supplemental Figure 7. Association between the expression in proteostasis complexes and mutational load is not driven by patient age. Boxplots of regression coefficients from the GLMM measuring the association of the expression of each individual gene with the mutational load of tumors from TCGA colored by different proteostasis complexes. Shown are regression coefficients when running the GLMM on tumors stratified by different age groups (x-axis).



Supplemental Figure 8. Association between the expression in proteostasis complexes and mutational load is not driven by a single cancer type in CCLE. Box plots of regression coefficients from the GLM measuring the association of the expression of each individual gene with the mutational load of tumors colored by different proteostasis complexes. Shown are regression estimates after removing each cancer type in CCLE (x-axis) and re-running the GLM.



Supplemental Figure 9. Similar patterns of expression and protein abundances in response to mutational load in CCLE within genes that regulate protein folding, degradation, and synthesis. Box plots of $\boldsymbol{\beta}_1$ regression coefficients measuring the association of mutation load and protein abundance (right) or gene expression (left) of individual genes in chaperone (purple), proteasome (yellow), and ribosome (green) complexes. Shown are regression coefficients from cancer cell lines (CCLE), which contains the largest dataset available of RNA (n=1377) and protein (n=373) abundances which are harmonized across samples. Percentages and grey lines on top panels show the quantile distribution of regression coefficients measuring the association of mutational load and expression for all genes in the genome within each dataset.

References

1. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* (2013) doi:10.1038/nature12213.

2. Bakhoum, S. F. & Landau, D. A. Cancer Evolution: No Room for Negative Selection. *Cell* (2017) doi:10.1016/j.cell.2017.10.039.

3. Martincorena, I. *et al.* Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* (2017) doi:10.1016/j.cell.2017.09.042.

4. Bozic, I., Paterson, C. & Waclaw, B. On measuring selection in cancer from subclonal mutation frequencies. *PLoS Comput. Biol.* (2019) doi:10.1371/journal.pcbi.1007368.

5. McFarland, C. D. *et al.* The damaging effect of passenger mutations on cancer progression. *Cancer Res.* (2017) doi:10.1158/0008-5472.CAN-15-3283-T.

6. McFarland, C. D., Korolev, K. S., Kryukov, G. V., Sunyaev, S. R. & Mirny, L. A. Impact of deleterious passenger mutations on cancer progression. *Proc. Natl. Acad. Sci. U. S. A.* (2013) doi:10.1073/pnas.1213968110.

7. Zapata, L. *et al.* Negative selection in tumor genome evolution acts on essential cellular functions and the immunopeptidome. *Genome Biol.* (2018) doi:10.1186/s13059-018-1434-0.

8. Williams, M. J., Werner, B., Barnes, C. P., Graham, T. A. & Sottoriva, A. Identification of neutral tumor evolution across cancer types. *Nat. Genet.* (2016) doi:10.1038/ng.3489.

9. Neutral Theory and the Somatic Evolution of Cancer. *Mol. Biol. Evol.* (2018) doi:10.1093/molbev/msy100.

10. Tilk, S., Curtis, C., Petrov, D. A. & McFarland, C. D. Most cancers carry a substantial deleterious load due to Hill-Robertson interference. *bioRxiv* (2019) doi:10.1101/764340.

11. Eyre-Walker, A. & Keightley, P. D. The distribution of fitness effects of new mutations. *Nature Reviews Genetics* (2007) doi:10.1038/nrg2146.

12. Martincorena, I. *et al.* Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* (2017) doi:10.1016/j.cell.2017.09.042.

13. Drummond, D. A. & Wilke, C. O. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* **134**, 341–52 (2008).

14. Hill, W. G. & Robertson, A. The effect of linkage on limits to artificial selection. *Genet. Res. (Camb).* (2008) doi:10.1017/S001667230800949X.

15. Weinstein, J. N. *et al.* The cancer genome atlas pan-cancer analysis project. *Nat. Genet.* (2013) doi:10.1038/ng.2764.

16. Ellrott, K. *et al.* Scalable Open Science Approach for Mutation Calling of Tumor Exomes Using Multiple Genomic Pipelines. *Cell Syst.* **6**, 271-281.e7 (2018).

17. Giurgiu, M. *et al.* CORUM: The comprehensive resource of mammalian protein complexes - 2019. *Nucleic Acids Res.* (2019) doi:10.1093/nar/gky973.

18. Tanabe, M. & Kanehisa, M. Using the KEGG database resource. *Curr. Protoc. Bioinforma.* (2012) doi:10.1002/0471250953.bi0112s38.

19. Peterson, H., Kolberg, L., Raudvere, U., Kuzmin, I. & Vilo, J. gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g: Profiler. *F1000Research* (2020) doi:10.12688/f1000research.24956.2.

20. Biamonti, G. & Caceres, J. F. Cellular stress and RNA splicing. *Trends in Biochemical Sciences* (2009) doi:10.1016/j.tibs.2008.11.004.

21. Dutertre, M., Sanchez, G., Barbier, J., Corcos, L. & Auboeuf, D. The emerging role of pre-messenger RNA splicing in stress responses: Sending alternative messages and silent messengers. *RNA Biology* (2011) doi:10.4161/rna.8.5.16016.

22. Ge, Y. & Porse, B. T. The functional consequences of intron retention: Alternative splicing coupled to NMD as a regulator of gene expression. *BioEssays* (2014) doi:10.1002/bies.201300156.

23. Lindeboom, R. G. H., Supek, F. & Lehner, B. The rules and impact of nonsense-mediated mRNA decay in human cancers. *Nat. Genet.* (2016) doi:10.1038/ng.3664.

24. Lareau, L. F., Inada, M., Green, R. E., Wengrod, J. C. & Brenner, S. E. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* (2007) doi:10.1038/nature05676.

25. Ryan, M. *et al.* TCGASpliceSeq a compendium of alternative mRNA splicing in cancer. *Nucleic Acids Res.* (2016) doi:10.1093/nar/gkv1288.

26. Sun, Y. & MacRae, T. H. The small heat shock proteins and their role in human disease. *FEBS Journal* (2005) doi:10.1111/j.1742-4658.2005.04708.x.

27. Ruan, L. *et al.* Cytosolic proteostasis through importing of misfolded proteins into mitochondria. *Nature* (2017) doi:10.1038/nature21695.

28. Shcherbakov, D. *et al.* Ribosomal mistranslation leads to silencing of the unfolded protein response and increased mitochondrial biogenesis. *Commun. Biol.* (2019) doi:10.1038/s42003-019-0626-9.

29. Stein, K. C. & Frydman, J. The stop-and-go traffic regulating protein biogenesis: How translation kinetics controls proteostasis. *Journal of Biological Chemistry* (2019) doi:10.1074/jbc.REV118.002814.

30. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer

drug sensitivity. *Nature* (2012) doi:10.1038/nature11003.

31. Harrison, X. A. *et al.* A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ* (2018) doi:10.7717/peerj.4794.

32. Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* (2017) doi:10.1016/j.cell.2017.06.010.

33. Zolkiewski, M., Zhang, T. & Nagy, M. Aggregate reactivation mediated by the Hsp100 chaperones. *Archives of Biochemistry and Biophysics* (2012) doi:10.1016/j.abb.2012.01.012.

34. Corsello, S. M. *et al.* Discovering the anticancer potential of non-oncology drugs by systematic viability profiling. *Nat. Cancer* (2020) doi:10.1038/s43018-019-0018-6.

35. Kristofich, J. C. *et al.* Synonymous mutations make dramatic contributions to fitness when growth is limited by a weak-link enzyme. *PLoS Genet.* (2018) doi:10.1371/journal.pgen.1007615.

36. Spencer, P. S., Siller, E., Anderson, J. F. & Barral, J. M. Silent substitutions predictably alter translation elongation rates and protein folding efficiencies. *J. Mol. Biol.* (2012) doi:10.1016/j.jmb.2012.06.010.

37. Zhang, G., Hubalewska, M. & Ignatova, Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat. Struct. Mol. Biol.* (2009) doi:10.1038/nsmb.1554.

38. Walsh, I. M., Bowman, M. A., Soto Santarriaga, I. F., Rodriguez, A. & Clark, P. L. Synonymous codon substitutions perturb cotranslational protein folding in vivo and impair cell fitness. *Proc. Natl. Acad. Sci. U. S. A.* (2020) doi:10.1073/pnas.1907126117.

39. Plotkin, J. B. & Kudla, G. Synonymous but not the same: The causes and consequences of codon bias. *Nature Reviews Genetics* (2011) doi:10.1038/nrg2899.

40. Bobula, J. *et al.* Why molecular chaperones buffer mutational damage: A case study with a yeast Hsp40/70 system. *Genetics* (2006) doi:10.1534/genetics.106.061564.

41. Maisnier-Patin, S. *et al.* Genomic buffering mitigates the effects of deleterious mutations in bacteria. *Nat. Genet.* (2005) doi:10.1038/ng1676.

42. Fares, M. A., Ruiz-González, M. X., Moya, A., Elena, S. F. & Barrio, E. Endosymbiotic bacteria: GroEL buffers against deleterious mutations. *Nature* (2002) doi:10.1038/417398a.

43. Elena, S. F., Carrasco, P., Daròs, J. A. & Sanjuán, R. Mechanisms of genetic robustness in RNA viruses. *EMBO Reports* (2006) doi:10.1038/sj.embor.7400636.

44. Pan, Q. *et al.* Quantitative microarray profiling provides evidence against widespread coupling of alternative splicing with nonsense-mediated mRNA decay to control gene expression. *Genes Dev.* (2006) doi:10.1101/gad.1382806.

45. Green, R. E. *et al.* Widespread predicted nonsense-mediated mRNA decay of alternatively-spliced transcripts of human normal and disease genes. in *Bioinformatics* (2003). doi:10.1093/bioinformatics/btg1015.

46. Cuccurese, M., Russo, G., Russo, A. & Pietropaolo, C. Alternative splicing and nonsense-mediated mRNA decay regulate mammalian ribosomal gene expression. *Nucleic Acids Res.* (2005) doi:10.1093/nar/gki905.

47. Mitrovich, Q. M. & Anderson, P. Unproductively spliced ribosomal protein mRNAs are natural targets of mRNA surveillance in C. elegans. *Genes Dev.* (2000) doi:10.1101/gad.819900.

48. Parenteau, J. *et al.* Introns within ribosomal protein genes regulate the production and function of yeast ribosomes. *Cell* (2011) doi:10.1016/j.cell.2011.08.044.

49. Kalapis, D. *et al.* Evolution of Robustness to Protein Mistranslation by Accelerated Protein Turnover. *PLoS Biol.* (2015) doi:10.1371/journal.pbio.1002291.

50. Tokheim, C. *et al.* Systematic characterization of mutations altering protein degradation in human cancers. *Mol. Cell* (2021) doi:10.1016/j.molcel.2021.01.020.

51. Martínez-Jiménez, F., Muiños, F., López-Arribillaga, E., Lopez-Bigas, N. & Gonzalez-Perez, A. Systematic analysis of alterations in the ubiquitin proteolysis system reveals its contribution to driver mutations in cancer. *Nat. Cancer* (2020) doi:10.1038/s43018-019-0001-2.

52. Conant, G. C. & Wagner, A. Duplicate genes and robustness to transient gene knock-downs in Caenorhabditis elegans. *Proc. R. Soc. B Biol. Sci.* (2004) doi:10.1098/rspb.2003.2560.

53. Gu, Z. *et al.* Role of duplicate genes in genetic robustness against null mutations. *Nature* (2003).

54. López, S. *et al.* Interplay between whole-genome doubling and the accumulation of deleterious alterations in cancer evolution. *Nat. Genet.* (2020) doi:10.1038/s41588-020-0584-7.

55. Torres, E. M. *et al.* Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science (80-.).* (2007) doi:10.1126/science.1142210.

56. Neckers, L. & Workman, P. Hsp90 molecular chaperone inhibitors: Are we there yet? *Clinical Cancer Research* (2012) doi:10.1158/1078-0432.CCR-11-1000.

57. Pacey, S., Banerji, U., Judson, I. & Workman, P. Hsp90 inhibitors in the clinic. *Handbook of experimental pharmacology* (2006).

58. Kim, Y. *et al.* Update on Hsp90 Inhibitors in Clinical Trial. *Curr. Top. Med. Chem.* (2009) doi:10.2174/156802609789895728.

59. Manasanch, E. E. & Orlowski, R. Z. Proteasome inhibitors in cancer therapy. *Nature Reviews Clinical Oncology* (2017) doi:10.1038/nrclinonc.2016.206.

60. Park, J. E., Miller, Z., Jun, Y., Lee, W. & Kim, K. B. Next-generation proteasome inhibitors for cancer therapy. *Translational Research* (2018) doi:10.1016/j.trsl.2018.03.002.

61. Bianchi, G. *et al.* The proteasome load versus capacity balance determines apoptotic sensitivity of multiple myeloma cells to proteasome inhibition. *Blood* (2009) doi:10.1182/blood-2008-08-172734.

62. Ling, S. C. W. *et al.* Response of myeloma to the proteasome inhibitor bortezomib is correlated with the unfolded protein response regulator XBP-1. *Haematologica* (2012) doi:10.3324/haematol.2011.043331.

63. Martincorena, I. *et al.* Somatic mutant clones colonize the human esophagus with age. *Science (80-.).* (2018) doi:10.1126/science.aau3879.

64. Martincorena, I. *et al.* High burden and pervasive positive selection of somatic mutations in normal human skin. *Science (80-.).* (2015) doi:10.1126/science.aaa6806.

65. Kennedy, S. R., Zhang, Y. & Risques, R. A. Cancer-Associated Mutations but No Cancer: Insights into the Early Steps of Carcinogenesis and Implications for Early Cancer Detection. *Trends in Cancer* (2019) doi:10.1016/j.trecan.2019.07.007.

66. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* (2013) doi:10.1016/j.cell.2013.05.039.

67. Adusumalli, S., Ngian, Z. K., Lin, W. Q., Benoukraf, T. & Ong, C. T. Increased intron retention is a post-transcriptional signature associated with progressive aging and Alzheimer's disease. *Aging Cell* (2019) doi:10.1111/acel.12928.

68. Löw, P. The role of ubiquitin-proteasome system in ageing. *General and Comparative Endocrinology* (2011) doi:10.1016/j.ygcen.2011.02.005.

69. Soti, C. & Csermely, P. Aging and molecular chaperones. *Exp. Gerontol.* (2003) doi:10.1016/S0531-5565(03)00185-2.

70. Charmpilas, N., Kyriakakis, E. & Tavernarakis, N. Small heat shock proteins in ageing and age-related diseases. *Cell Stress and Chaperones* (2017) doi:10.1007/s12192-016-0761-x.

71. Aman, Y. *et al.* Autophagy in healthy aging and disease. *Nat. Aging* (2021) doi:10.1038/s43587-021-00098-4.

72. Goldman, M. J. *et al.* Visualizing and interpreting cancer genomics data via the Xena platform. *Nature Biotechnology* (2020) doi:10.1038/s41587-020-0546-8.

73. Bamford, S. *et al.* The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br. J. Cancer* (2004) doi:10.1038/sj.bjc.6601894.

74. Grossman, R. L. *et al.* Toward a Shared Vision for Cancer Genomic Data. *N. Engl. J. Med.* **375**, 1109–1112 (2016).

75. Nusinow, D. P. *et al.* Quantitative Proteomics of the Cancer Cell Line Encyclopedia. *Cell* (2020) doi:10.1016/j.cell.2019.12.023.

76. Hadizadeh Esfahani, A., Sverchkova, A., Saez-Rodriguez, J., Schuppert, A. A. & Brehme, M. A systematic atlas of chaperome deregulation topologies across the human cancer landscape. *PLoS Comput. Biol.* (2018) doi:10.1371/journal.pcbi.1005890.