Gene expression analysis of heat-shock proteins and redox regulators reveals combinatorial prognostic markers in carcinomas of the gastrointestinal tract

Sebastian Öther-Gee Pohl[a,b,c], Shazib Pervaiz[b,d,e,f], Arun Dharmarajan[a,b,c], Mark Agostino[a,b,c,g,*]

[a] Stem Cell and Cancer Biology Laboratory, Curtin University, Perth, WA, Australia
[b] School of Pharmacy and Biomedical Sciences, Curtin University, Perth, WA 6102, Australia
[c] Curtin Health and Innovation Research Institute, Curtin University, Perth, WA 6102, Australia
[d] Department of Physiology and Medical Science Cluster Cancer Program, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117593, Singapore
[e] NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, Singapore 117456, Singapore
[f] National University Cancer Institute, National University Health System, Singapore 119074, Singapore
[g] Curtin Institute for Computation, Curtin University, Perth, WA 6102, Australia

ABSTRACT

Heat shock proteins (HSPs) are a large family of ubiquitously expressed proteins with diverse functions, including protein assembly and folding/unfolding. These proteins have been associated with the progression of various gastrointestinal tumours. Dysregulation of cellular redox has also been associated with gastrointestinal carcinogenesis, however, a link between HSPs and dysregulation of cellular redox in carcinogenesis remains unclear. In this study, we analysed mRNA co-expression and methylation patterns, as well as performed survival analysis and gene set enrichment analysis, on gastrointestinal cancer data sets (oesophageal, stomach and colorectal carcinomas) to determine whether HSP activity and cellular redox dysregulation are linked. A widespread relationship between HSPs and cellular redox was identified, with specific combinatorial co-expression patterns demonstrated to significantly alter patient survival outcomes. This comprehensive analysis provides the foundation for future studies aimed at deciphering the mechanisms of cooperativity between HSPs and redox regulatory enzymes, which may be a target for future therapeutic intervention for gastrointestinal tumours.

1. Introduction

Heat shock proteins (HSPs) are a highly conserved, ubiquitously expressed family of proteins, that act as molecular chaperones to correctly fold a new polypeptide chain, or support the refolding of damaged proteins [1]. HSPs have been implicated in a variety of cellular processes, including cell cycle regulation, organelle trafficking, and signal transduction [2]. Interestingly, experimental evidence indicates that stress-induced activation of HSPs endows cells with a survival advantage by mechanism(s) that, on the one hand, activate pro-survival networks, and on the other, inhibit apoptotic execution [3]. These attributes and the association with pro-inflammatory signalling lend credence to the involvement of HSPs in promoting carcinogenesis and its progression [4,5]. Along similar lines, the differential expression of various HSP family members and dysregulation of redox homeostasis have been individually implicated in tumorigenesis, particularly involving the gastrointestinal (GI) tract [6,7]. However, the interplay between HSPs and redox regulation in the setting of GI cancers is less widely understood.

HSPs are divided into five families, including the DNAJ (HSP40) family, the HSPB (small heat shock protein family), the HSP90/HSPC family, chaperonins (HSPD/E and CCT) and the HSP70 superfamily, containing the HSPA (HSP70) and HSPH (HSP110) subfamilies [8]. These families mainly differ in their functions and cellular localisations. The DNAJ/HSP40 family primarily function by stimulating the ATPase activity of HSP70s [9], while the HSP70 family has been implicated in most aspects of protein folding, import, recovery from aggregation and drug resistance in cancer [10]. HSP90 is involved in many cellular processes, including DNA repair [11], and has been implicated in a variety of neoplasia [12]. The HSPB family, in particular HSP27, has been shown to be associated with several human cancers, as it promotes epithelial to mesenchymal transition (EMT) [13,14], blocks apoptotic signalling [15], and triggers activation of oncogenic transcription factors such as NF-κB [13] and mitogen activated protein kinases (MAPK) [16], independent of their ATP hydrolysis function [17]. Chaperonins, or CCT (chaperonin containing TCP-1) proteins, are ring complexes encompassing a central cavity; this allows protein folding in a shielded environment. They can be divided into two groups: Group 1 (HSPD),

that are present in the mitochondria, and Group 2, that are localized to the cytosol [18].

An imbalance in intracellular redox conditions can promote or inhibit tumour growth [19]. This is dependent on a tight balance between the different reactive oxygen species present within the cell [20]. Increased superoxide (O_2^-) production, favouring a pro-oxidant milieu, has been demonstrated to act as a pro-survival signal in some cancers [21–23], while increased peroxides, particularly hydrogen peroxide (H_2O_2), can initiate death signalling, leading to apoptosis [24,25]. There is a close relationship between intracellular redox and HSPs. Protein folding, specifically, disulfide bond formation, is an oxidative process and generates ROS, primarily in the form of H_2O_2 [26,27]. Dysregulation of protein folding can cause oxidative stress and stimulate the unfolded protein response (UPR), which is dependent on HSP proteins [28]. Furthermore, ROS-mediated DNA oxidation and lipid peroxidation can result in the induction of HSP synthesis [29]. For example, a redox-regulated HSP family member, HSP33, functions as a cellular protective agent against oxidative stress in prokaryotes; a co-ordinated zinc ion of HSP33 is released from the protein following the oxidation of cysteine residues by H_2O_2 to activate the protein [30]. Mammalian HSP1 has been reported to exhibit a similar oxidative stress response to prokaryotic HSP33. Two conserved cysteine residues near the DNA-binding domain are crucial for the formation of higher order homomultimers, required for target gene transcription and protection from oxidative stress [31]. These redox-mediated reversible reactions are critical for the activity of HSP33 and HSP1 and demonstrate the importance of cellular redox state on HSP function.

Datasets describing molecular and clinical features of patient tumours have been made publicly available through consortia such as The Cancer Genome Atlas (TCGA), accessible through websites such as cBioPortal [32,33]. These datasets contain large numbers of patient samples across many cancer types, linking mutation, copy-number alterations, transcriptomic, proteomic and methylomic data with clinical attributes, including survival outcomes, age, gender and therapeutic interventions. These large datasets have provided the foundation for studies that have resulted in the successful identification of biomarkers, development of *in vivo* models and the identification of drug candidates [34–36]. The generation of a gene signature in response to different drugs using only public databases enabled the repurposing of niclosamide ethanolamine, previously used as an anti-helminthic, as an anti-tumour agent against hepatocellular carcinoma (HCC); this molecule disrupted the interaction between cell division cycle 37 and HSP90, and was effective at reducing tumour burden *in vivo* [34]. Similarly, an *in silico* approach using public datasets identified protein tyrosine kinase PTK7 as a potential oncogenic driver in non-small cell lung cancer (NSCLC); these findings were validated *in vitro* and *in vivo*, where PTK7 was found to regulate MKK7-JNK signalling, indicating its utility as a potential therapeutic target [35].

In this study, we identified a range of previously unreported genes in GI cancer patients – specifically, oesophageal, stomach and colorectal cancers – that resulted in significantly altered survival outcomes when differentially expressed. We have also identified new combinations of HSP and redox genes that, when differentially expressed, predicted overall and disease-free survival in the selected cohorts. These genes may represent clinically relevant targets for use as biomarkers in monitoring disease progression and severity, as well as offering potential interest for the development of new cancer therapeutics.

2. Methods

2.1. Data collection

Gastrointestinal carcinoma datasets were obtained from the relevant TCGA studies via cBioPortal [33] in November 2017, specifically the oesophageal carcinoma [37], stomach adenocarcinoma [38] and colorectal adenocarcinoma data sets [39]. Based on the families set out in the guidelines for the nomenclature for HSPs, we examined 96 HSP genes [8] and 30 cancer-associated genes involved in redox generation and detoxification [40] in a range of analyses, including co-expression of HSP and redox regulator genes, correlation of gene methylation and expression at different cancer stages, survival, and enrichment of other gene families in susceptible populations. For all analyses, the three cancer types were analysed separately. Analysis was facilitated via Perl scripts developed in-house.

2.2. Co-expression and methylation analysis

For co-expression analyses, Spearman's ρ values were calculated using mRNA expression *z*-scores (calculated from RNA Seq V2 RSEM data, compared to the expression distribution of each gene in tumours that are diploid for that gene, as provided by cBioPortal) for all gene combinations, using all patient data. Spearman's ρ was also used to assess correlation between methylation β-values (from HM450 data) and mRNA expression *z*-scores for all genes, separating patient data by tumour stage; positive ρ values indicate that increased gene methylation can be correlated to increased gene expression, while conversely, a negative ρ values that increased gene methylation can be correlated to decreased gene expression.

2.3. Survival analysis

Survival analysis was performed for selected patient groups meeting specific criteria. Patients were categorized as "high" or "low" expressers of genes, using an absolute mRNA expression *z*-score cutoff of 0.75; such groups required a minimum of 10 patients before further analysis was performed. The log-rank test was used to assess whether differences in survival were statistically significant; comparisons yielding *p*-values less than 0.05 were selected as statistically significant. Patient groups with single gene high/low phenotypes were assessed, as well as groups with two gene high/low phenotypes.

2.4. Gene set enrichment analysis

For patient groups with phenotypes affording significant differences in survival, gene set enrichment analysis (GSEA) was performed using the GSEA software [41,42]. The lists of patients corresponding to each patient group for comparison were manually provided to GSEA to create phenotypes for analysis. GO gene sets [43] were used in the analysis. Gene sets affording *p*-values less than 0.05 and *q*-values (false discovery rates) of less than 0.25 were selected as being significantly altered between the patient groups. The remaining gene sets were then ranked by the GSEA normalised enrichment score (NES), and up to the top 350 ranked gene sets were analysed using REVIGO [44] to further summarize the biological importance of the identified gene sets. NESs for each gene set were supplied to REVIGO as a measure of the relative significance of each gene set. The human database was used and allowed similarity of output gene sets was set to tiny (0.4). The SimRel semantic similarity measure was used [45].

3. Results

3.1. Analysis of co-expression patterns of heat shock proteins and redox regulator genes in GI carcinomas

To determine co-expression patterns of the 126 HSP and redox regulator genes, we calculated Spearman's rank correlation coefficient (ρ) for all gene combinations for the stomach adenocarcinoma, colorectal adenocarcinoma and oesophageal carcinoma datasets. Hierarchical clustering revealed numerous HSPs and redox regulators that were strongly positively and negatively correlated in all carcinoma datasets.

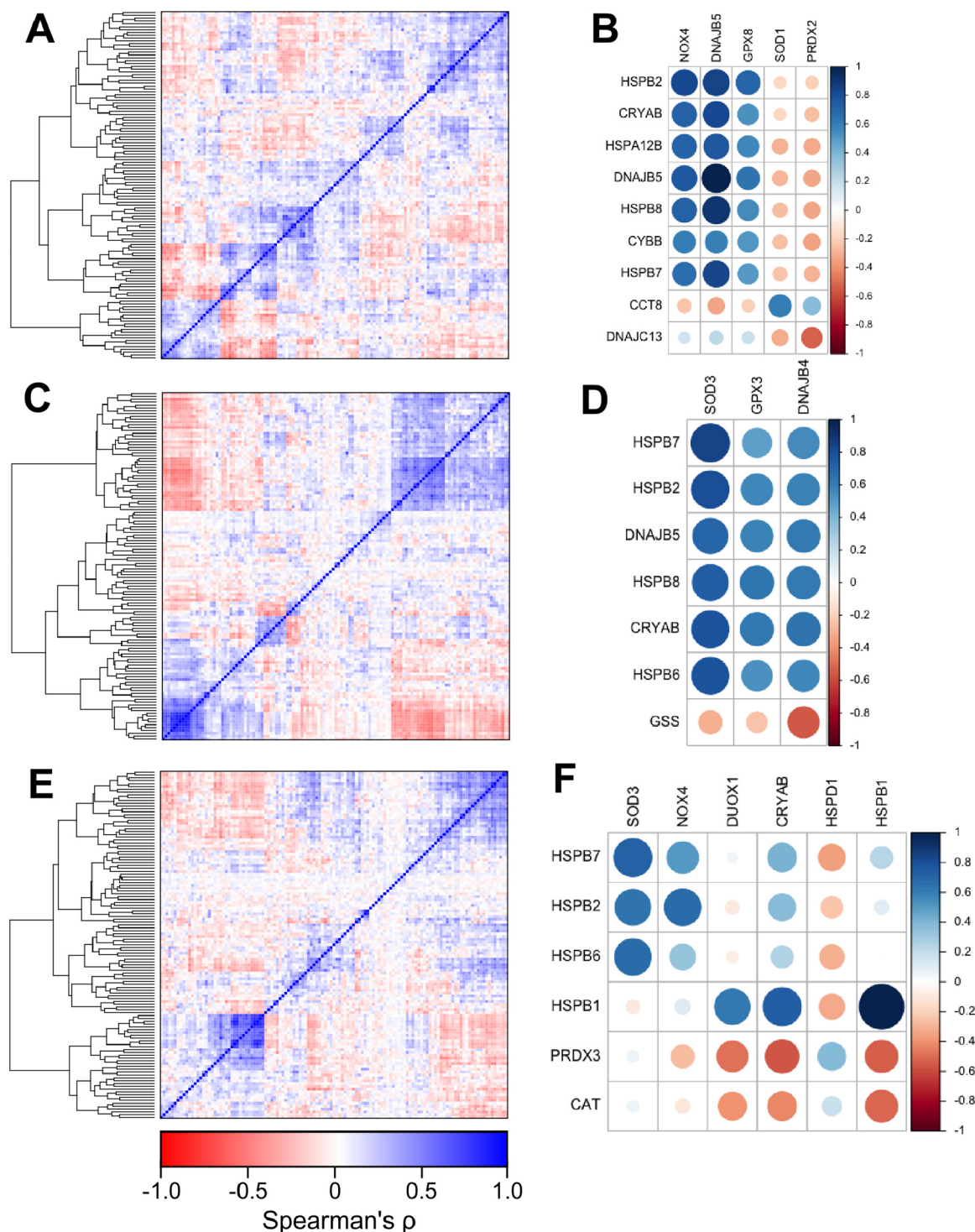The colorectal dataset showed a strong positive correlation in gene

Fig. 1. Co-expression analysis of HSPs and redox regulators in gastrointestinal tumours. Heatmaps of Spearman's rank correlation coefficient (ρ) for all HSP and redox genes (A, C,E) and selected gene combinations (B, D,F) in the following TCGA datasets: colorectal carcinoma (A, B), stomach adenocarcinoma (C, D), oesophageal carcinoma (E, F). In all panels, blue indicates negative ρ, red indicates positive ρ. In panels B, D and F, circle size indicates magnitude of ρ.

expression between NADPH oxidase 4 (*NOX4*) and numerous members of the small HSP family, including *HSPB2, HSPB5 (CRYAB), HSPB7 and HSPB8*, as well as members of the HSP40 family *DNAJB5* and *DNAJC5B. NOX4* also has a strong positive correlation with *HSPA12B*, a member of the HSP70 family (Fig. 1A & B). Other strong positive correlations in gene expression were identified between superoxide dismutase 1 (*SOD1*) with *CCT8* and glutathione peroxidase 8 (*GPX8*) with *HSPB2*. Peroxiredoxin 2 (*PRDX2*) gene expression was negatively correlated with *DNAJC13*.

In the stomach adenocarcinoma dataset, strong positive correlations in gene expression were identified between members of the small HSP family, including *HSPB2, HSPB5 (CRYAB), HSPB6, HSPB7 and HSPB8,* as well as the HSP40 family member *DNAJB5*, with superoxide dismutase (*SOD3*) (Fig. 1C & D). Glutathione peroxidase 3 (*GPX3*) gene expression was strongly positive correlated with *HSPB8*. The only redox/HSP combination that exhibited a noteworthy negative correlation at the level of gene expression was between glutathione synthetase (*GSS*) and *DNAJB4*.
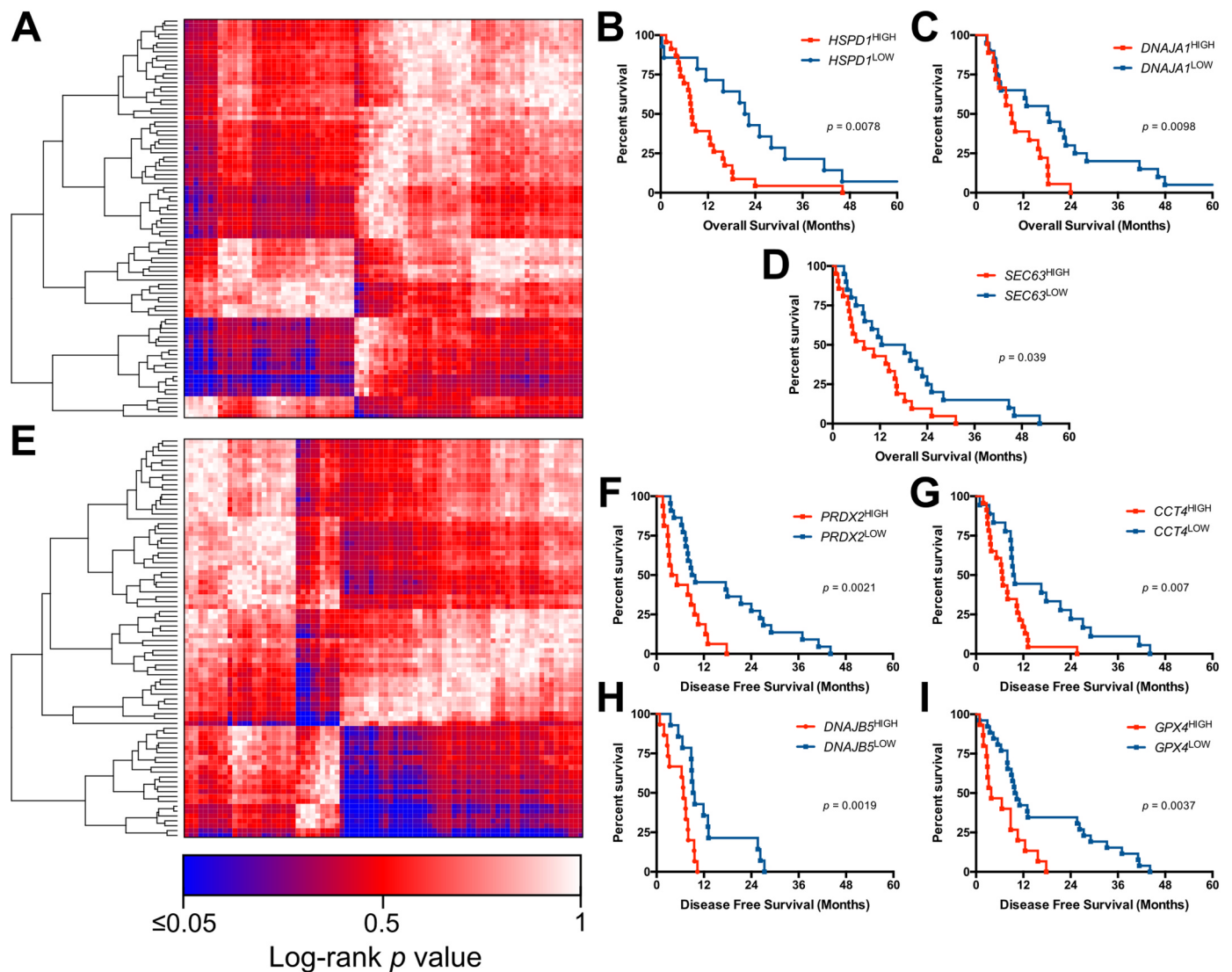
Fig. 2. Single gene survival analysis of HSPs and redox regulators in oesophageal carcinoma. A) Heatmap of log-rank p-values for overall survival differences of HSPs and redox regulators in oesophageal carcinoma. Blue indicates significant ($p < 0.05$) survival differences, Red indicates non-significant ($p > 0.05$) survival differences. Overall survival analysis for B) $HSPD1^{high}$ vs $HSPD1^{low}$ C) $DNAJA1^{high}$ vs $DNAJA1^{low}$ D) $SEC63^{high}$ vs $SEC63^{low}$. E) Heatmap of log-rank p-values for disease-free survival differences of HSPs and redox regulators in oesophageal carcinoma. Overall disease-free survival analysis for F) $PRDX2^{high}$ vs $PRDX2^{low}$ G) $CCT4^{high}$ vs $CCT4^{low}$ H) $DNAJB5^{high}$ vs $DNAJB5^{low}$ I) $GPX4^{high}$ vs $GPX4^{low}$.

Strong positive correlations in the expression of *SOD3* with the expression of *HSPB2*, *HSPB6* and *HSPB7* were found in the oesophageal carcinoma dataset (Fig. 1E & F). Gene expression of the NADPH oxidase family members *NOX4* and *DUOX1* was positively correlated with expression of *HSPB1* and *HSPB2*. Negative correlations were found between *PRDX3* with both *HSPB5* and *HSPD1*, as well as between *CAT* and *HSPB1*.

3.2. Survival analysis of patient groups exhibiting differential expression of HSP and redox genes

We then set to determine if there were novel predictors of overall and disease-free survival based on combinations of HSP or redox genes, which exhibited either high or low expression. Single gene and two-gene survival analyses were performed.

In the oesophageal carcinoma dataset, patients exhibiting high expression of *HSPD1*, *DNAJA1* and *SEC63* all resulted in significantly lower overall survival when compared to patients with low expression (Fig. 2A-D). High expression of *PRDX2* and *GPX4* reduced the disease-free survival significantly compared to patients with lower mRNA levels

of these genes. Furthermore, patients with higher expression of *DNAJB5* and *CCT4* had significantly reduced disease-free survival rates (Fig. 2E-I). Two-gene survival analysis indicated that patients with $CCT2^{high}$/ GSS^{high}, $DNAJB12^{high}/GSS^{high}$ and $DNAJC8^{high}/GSS^{high}$ phenotypes all exhibited poorer disease-free survival, compared to patients with $CCT2^{low}/GSS^{high}$, $DNAJB12^{low}/GSS^{high}$ and $DNAJC8^{low}/GSS^{high}$ expression (Fig. 5A–C).

The stomach adenocarcinoma dataset revealed that overall survival was significantly altered as a result of differential expression of four HSPs - *CCT8*, *DNAJC19*, *GAK* and *SEC63* (Fig. 3B-E). Low expression of *CCT8*, *DNAJC19* and *SEC63* significantly increased overall survival, whereas low expression of *GAK* significantly decreased overall survival (Fig. 3D). Significantly poorer disease-free survival rates were found in patients with high *DNAJA4*, *DNAJC13* and *HSPA4* expression compared to patients with low expression (Fig. 3G-I). Two-gene survival analysis revealed significantly poorer survival in patients with $CCT5^{high}/GPX1^{low}$ expression compared to $CCT5^{low}/GPX1^{low}$ expression. Conversely, patients with $DNAJC7^{high}/GPX1^{low}$ expression had a significantly better disease free survival rate than those with $DNAJC7^{low}/GPX1^{low}$ expression. High expression of both *CCT4* and *CCT7* resulted in significantly
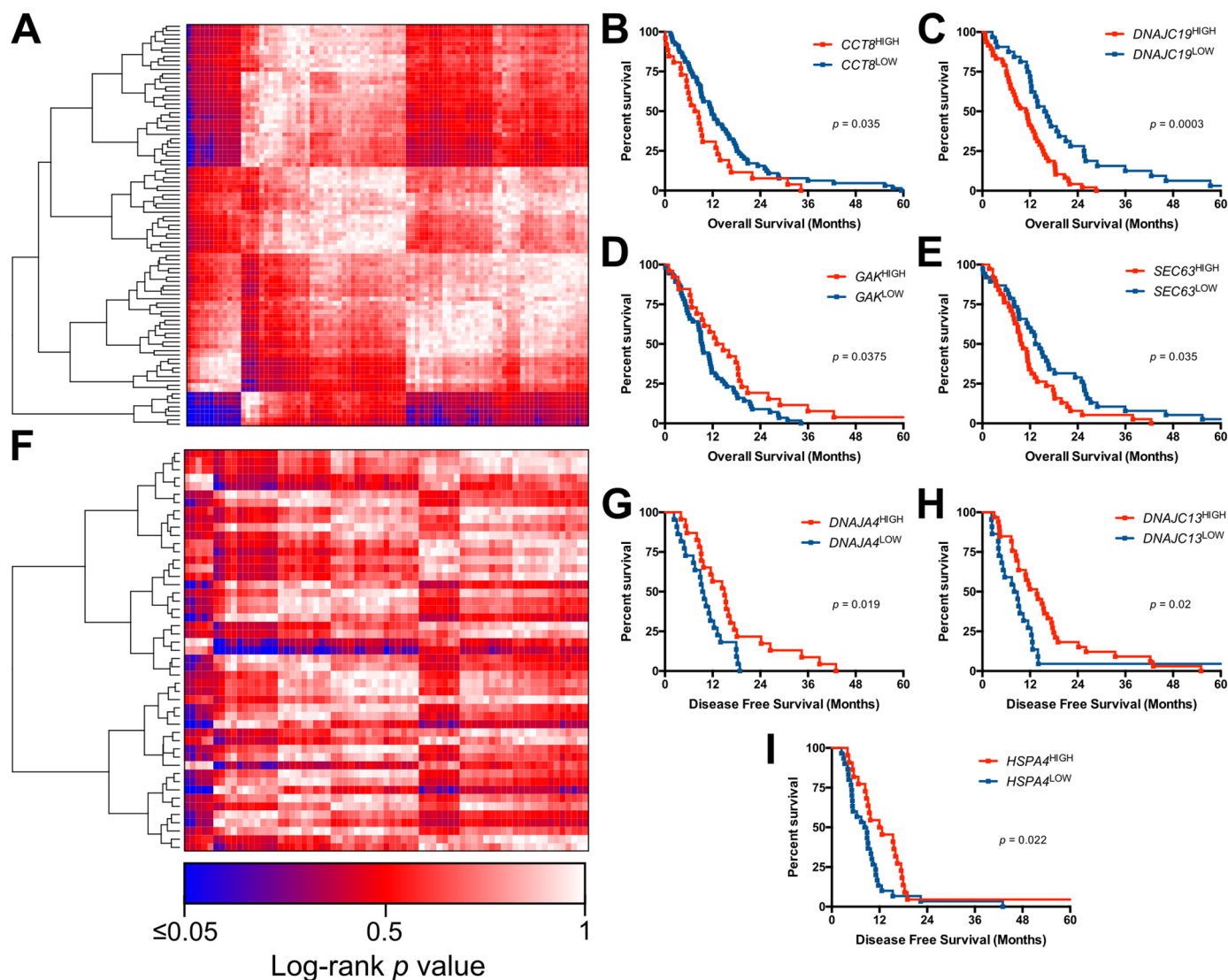
Fig. 3. Single gene survival analysis of HSPs and redox regulators in stomach adenocarcinoma. A) Heatmap of log-rank p-values for overall survival differences of HSPs and redox regulators in stomach adenocarcinoma. Blue indicates significant (p < 0.05) survival differences, Red indicates non-significant (p > 0.05) survival differences. Overall survival analysis for B) *CCT8*^high^ vs *CCT8*^low^ C) *DNAJC19*^high^ vs *DNAJAC19*^low^ D) *GAK*^high^ vs *GAK*^low^ E) *SEC63*^high^ vs *SEC63*^low^. F) Heatmap of log-rank p-values for disease-free survival differences of HSPs and redox regulators in stomach adenocarcinoma. Disease-free survival analysis for G) *DNAJA4*^high^ vs *DNAJA4*^low^ H) *DNAJC13*^high^ vs *DNAJC13*^low^ I) *HSPA4*^high^ vs *HSPA4*^low^

poorer survival than those with low expression of *CCT4* and *CCT7* (Fig. 5D–F).

The colorectal adenocarcinoma dataset revealed that the increased expression of *CCT2* resulted in a significantly decreased disease-free survival (Fig. 4C). When two-gene survival analysis was performed combinations of HSP and redox regulators, patients with $HSPA9^{high}$/$PRDX1^{high}$, $DNAJA2^{high}$/TXN^{high} and $DNAJA2^{high}$/$PRDX6^{high}$ genotypes all had significantly decreased overall survival rates when compared to the same combinations with low expression (Fig. 5G–I).

3.3. Correlations between gene methylation and expression at different pathological stages

We set to determine Spearman correlation coefficients between gene methylation and expression for 126 genes in our HSP and redox regulator set, and to draw connection between this, pathological stage and survival (Fig. 6A-C). The majority of genes in oesophageal carcinoma unexpectedly exhibited a negative correlation between methylation and mRNA expression at stages T1, T2 and T3 (Fig. 6D). *HSPB9* showed a negative correlation across these stages (Fig. 6D). *GPX4* demonstrated a

weak negative correlation that remained mainly unchanged across different pathological stages. *CCT6A* exhibited the strongest negative correlation at T1, which decreased with tumour stage progression. *DNAJC18* showed a strong negative correlation at T2, which progressed to a weak negative correlation in stage T3. The loss of a strong negative correlation during the progression of tumours suggests that there may be decreased methylation and increased mRNA expression during the more progressed stages. This is consistent with survival data for *DNAJC18* (Fig. S4A) and *CCT6A* (Fig. S4B), both of which, when highly expressed, result in significantly poorer overall patient survival compared to patients with low expression. Conversely, consistently low negative correlations between stages in *HSPB9* suggest no loss of methylation and low mRNA expression, consistent with low levels of *HSPB9* reflecting significantly poorer patient survival.

When the analysis was performed on the stomach adenocarcinoma data, a strong negative correlation between gene expression and methylation was observed for *DNAJA4*. This correlation was stronger in the later stages (T2, T2b, T3, T4, T4a and T4b) than in T1 (Fig. 6E). The survival analysis revealed that patients with $DNAJA4^{low}$ had a significantly poorer disease free survival than those with the $DNAJA4^{high}$
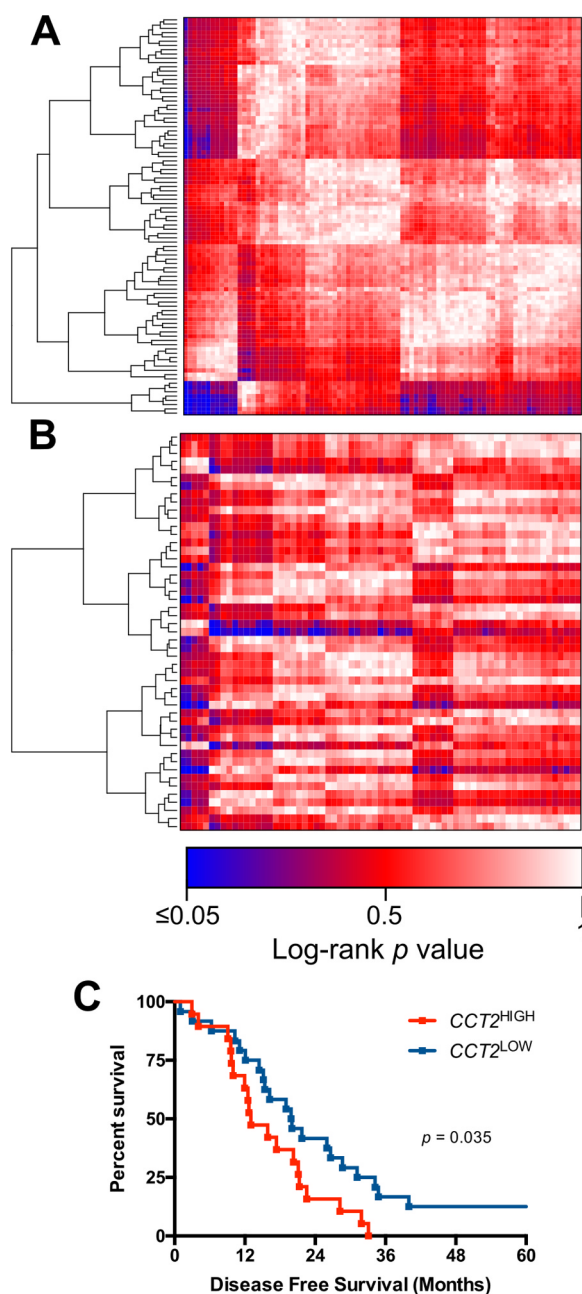
Fig. 4. Single gene survival analysis of HSPs and redox regulators in colorectal carcinoma. A) Heatmap of log-rank p-values for overall survival differences of HSPs and redox regulators in colorectal carcinoma. B) Heatmap of log-rank p-values for disease-free survival differences of HSPs and redox regulators in colorectal carcinoma. Blue indicates significant (p < 0.05) survival differences, Red indicates non-significant (p > 0.05) survival differences. C) Disease-free survival analysis for *CCT2*^high vs *CCT2*^low.

phenotype, suggesting that increased gene methylation of *DNAJA4* may predict poorer survival. A similar trend of a strong negative correlation was seen in *GAK* in all stages (T1-T4b). Again, patients with a *GAK*^low phenotype had a significantly poorer overall survival than those with a *GAK*^high phenotype, suggesting that the methylation status, corresponding to overall gene expression, plays an important role in overall patient survival.

In the colorectal dataset, two genes were identified where methylation may play an important role in gene expression at different tumour stages. *GSTM1* exhibited a very strong negative correlation between methylation and gene expression in stages T1 through to T4a,

however, this correlation was nearly completely lost in stage T4b (Fig. 6F). Methylation of *DNAJC8* was strong negatively correlated in stage T1, suggesting gene silencing through methylation, although in T4a and T4b, this correlation was reversed, suggesting that methylation was correlated with an increase in gene expression.

3.4. Gene set enrichment analysis in patient groups associated with significantly altered survival

To determine whether the genotypes associated with differing disease-free or overall survival outcomes were also broadly associated with other biological processes and molecular functions, we performed gene set enrichment analysis (GSEA). The Gene Ontology (GO) gene sets were used for this purpose. Across the three gastrointestinal tumour types, grouped individuals with high expression of HSP family members were, unsurprisingly, enriched in gene sets covering processes involved in canonical HSP function, for example, protein folding, chaperone-mediated protein folding and heat shock protein binding (Fig. 7A–C).

The colorectal adenocarcinoma dataset, which had only one statistically significant survival phenotype (*CCT2*^high vs *CCT2*^low), was primarily enriched in GO gene sets involved in mitochondrial transport and organisation. (Fig. 7D). *CCT2*^low individuals were enriched for gene sets involved in clathrin-mediated endocytosis compared to *CCT2*^high individuals.

GSEA on the oesophageal carcinoma phenotypes associated with differential survival outcomes were enriched in GO terms associated with mitochondrial matrix, transport and genome maintenance (*CCT4/CCT7*^high vs *CCT4/CCT7*^low), mitochondrial morphogenesis and oxidoreductase activity (*HSPD1*^high vs *HSPD1*^low) and protein localisation to the membrane, organelles and ER (*GPX4*^high vs *GPX4*^low). The *CCT4/CCT7*^low vs *CCT4/CCT7*^high phenotypes were enriched for gene sets involved in the negative regulation of IL-6 production (Fig. 7E).

In the stomach adenocarcinoma dataset, GO gene sets for mitochondrial membrane transport and structure, as well as mitochondrial respiration components, were significantly enriched. Gene sets involved in hydrogen peroxide metabolic processes and cell redox homeostasis were enriched in patients with a *DNAJC13*^low vs *DNAJC13*^high phenotypes. *HSPA9*^high/*PRDX*1^high individuals were significantly enriched in genes mediating mitochondrial respiratory chain complex I biogenesis and protein localisation to the mitochondria, while conversely, *HSPA9*^low/*PRDX*1^low individuals were significantly enriched for gene sets involved in the negative regulation of growth. Gene sets for mitochondrial matrix, transport and aerobic respiration were all enriched in the *CCT8*^high individuals, while genes mediating negative regulation for locomotion were enriched in grouped individuals demonstrating a *CCT8*^low phenotype (Fig. 7F).

4. Discussion

We analysed publicly available TCGA datasets from three gastrointestinal tumours utilising a list of 96 HSP family genes and 30 genes that have been implicated in cellular redox functions. We analysed gene correlations, correlations between methylation status in tumour stages with gene expression, and overall and disease free survival in patients. We then selected the genotypes that demonstrated significant differences in patient survival and performed gene set enrichment analysis to reveal up or downregulated gene sets associated with these phenotypes. The colorectal and oesophageal dataset revealed positive gene correlations between *NOX4* and a variety of HSP family members. *NOX4* has been previously shown to mediate the expression of *HSP27* mediated by TGFβ [46]. This is in line with previous reports that have demonstrated co-localisation of NOX4 with HSP27 and HSP70 in vascular smooth muscle cells [47] and myoblasts [48] respectively. In all datasets, members of the superoxide dismutase family, which are responsible for the production of H_2O_2 through the dismutation of O_2^-, have a positive correlation with many HSPs. This is consistent with previous
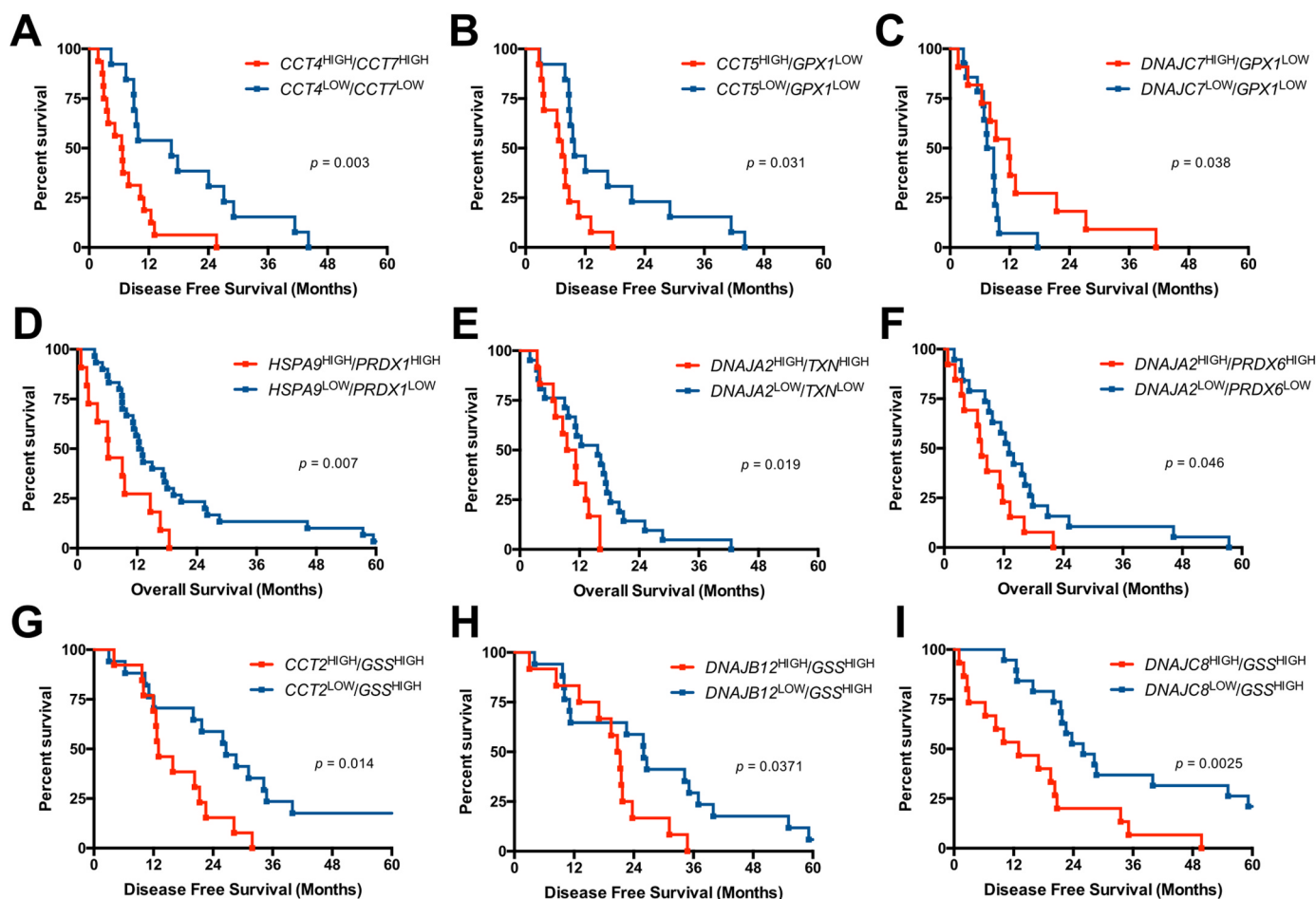
Fig. 5. Combinatorial gene survival analysis of HSPs and redox regulators in gastrointestinal cancers. Disease-free survival analysis for A) *CCT4/CCT7*[high] vs *CCT4/CCT7*[low], B) *CCT5/GPX1*[high] vs *CCT5/GPX1*[low] C) *DNAJC7/GPX1*[high] vs *DNAJC7/GPX1*[low] in oesophageal carcinoma dataset. Overall survival analysis for D) *HSPA9/PRDX1*[high] vs *HSPA9/PRDX1*[low] E) *DNAJA2/TXN*[high] vs *DNAJA2/TXN*[low] and F) *DNAJA2/PRDX6*[high] vs *DNAJA2/PRDX6*[low] in stomach adenocarcinoma dataset. Disease-free survival analysis for G) *CCT2*[high]*/GSS*[high] vs *CCT2*[low]*/GSS*[low] H) *DNAJB12*[high]*/GSS*[high] vs *DNAJB12*[low]*/GSS*[low] I) *DNAJC8*[high]*/GSS*[high] vs *DNAJC8*[low]*/GSS*[low] in colorectal carcinoma dataset.

observations that have demonstrated the induction of a variety of HSPs, including HSP30, HSP90 and HSP70, through H_2O_2 [49–51]. HSP70 itself directly interacts with SOD2 and may chaperone the enzyme to the mitochondria, although the exact mechanism remains unclear [52]. Interestingly, enzymes involved in ROS detoxification, such as catalase and peroxiredoxins, exhibited a negative correlation with HSPs. This aligns with the observation that *HSP* gene expression is induced by H_2O_2. Higher levels of intracellular hydrogen peroxide, as a result of decreased catalase or peroxidredoxin expression, may facilitate cellular redox conditions permissible for the induction of high levels of HSP family members. Strong positive correlations were shown between various members of the small HSP family and redox-related genes. It has previously been shown that HSP27 plays roles in the pathological state of cancer cells in response to oxidative stress through modulation of the ROS-glutathione pathway [53]. HSP27 has been shown to act as a protectant against the oxidation of proteins [53]. Overexpression of HSP27 has also been shown to increase extracellular O_2^- production to produce a pro-oxidant state beneficial for cell survival [53,54]. Furthermore, phosphorylation-mediated downregulation of HSP27 was shown to sensitize cancer cells to TRAIL-mediated apoptosis upon exposure to a small molecule [55] that signals through extracellular H_2O_2 production [56]. These observations align with results demonstrated in all these gastrointestinal cancer datasets.

Gene methylation is widely known to influence gene expression [57]. We observed a strong negative correlation between *DNAJA4* gene

methylation and expression in stomach adenocarcinoma, which suggested that methylation of *DNAJA4* may suppress gene expression. This is consistent with reports that *DNAJA4* expression is strongly influenced by its methylation status [58]. Our data also revealed that patients with low *DNAJA4* expression had poorer survival outcomes. Together, this data suggests that gene methylation of *DNAJA4* may suppress its expression, leading to poorer patient outcomes. *DNAJA4* has been shown to promote metastasis and angiogenesis in melanoma [59], but may act as a tumour suppressor in stomach cancer. *DNAJA4* and its methylation status may be a new biomarker for stomach adenocarcinoma, although this may not be applicable to other carcinomas. Gene expression of *GAK* (cyclin G-associated kinase) was also found to be negatively correlated in stomach cancer, and patients with a *GAK*[low] phenotype had poorer overall survival. While *GAK* has been shown to promote oncogenesis in some cancers [60], in stomach cancer it appears that higher expression of *GAK* may be beneficial. *GAK* has been shown to be essential for clathrin-mediated endocytosis [61], which, if impaired, can lead to prolonged cell signalling and higher rates of proliferation [62]. In the colorectal dataset, *GSTM1* demonstrated a high negative correlation in all stages prior to stage T4a, whereas in stage T4b this was completely lost, suggesting that low levels of GSTM1 are associated with the onset and progression of colorectal cancer. This has been confirmed in different populations, where the *GSTM1* null variant was correlated with increased colorectal cancer risk [63,64]. The mechanisms of *GSTM1* gene methylation may provide interesting insights into the onset of
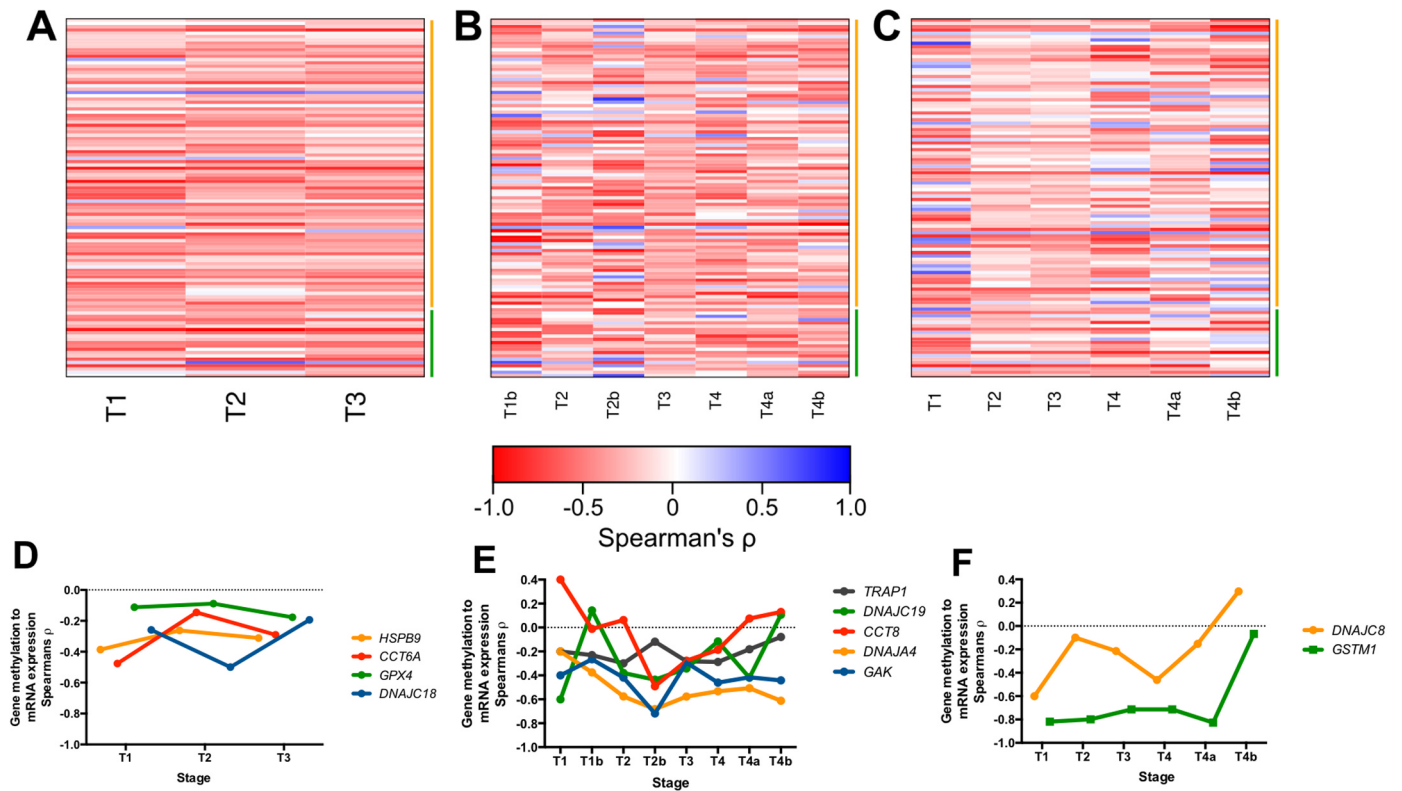
Fig. 6. Correlation of methylation and gene expression of HSPs and redox regulators in gastrointestinal cancers. Heatmap of Spearman's rank correlation coefficient (ρ) for TCGA datasets of A) oesophageal carcinoma B) stomach adenocarcinoma and C) colorectal carcinoma for various tumour stages (T1-T4b). Blue indicates negative Spearman's coefficient, Red indicates positive Spearman's coefficient. Green line indicates redox regulator genes, yellow line indicates HSP genes. D) Line graph plotting gene methylation to mRNA expression through tumour stages T1-T3 of *HSPB9, CCT6A, GPX4* and *DNAJC18* for oesophageal carcinoma dataset. E) Line graph plotting gene methylation to mRNA expression through tumour stages T1-T4b of *TRAP1, DNAJC19, CCT8, DNAJA4* and *GAK* for stomach adenocarcinoma dataset. F) Line graph plotting gene methylation to mRNA expression through tumour stages T1-T4b of *DNAJC8*, and *GSTM1* for colorectal carcinoma dataset.
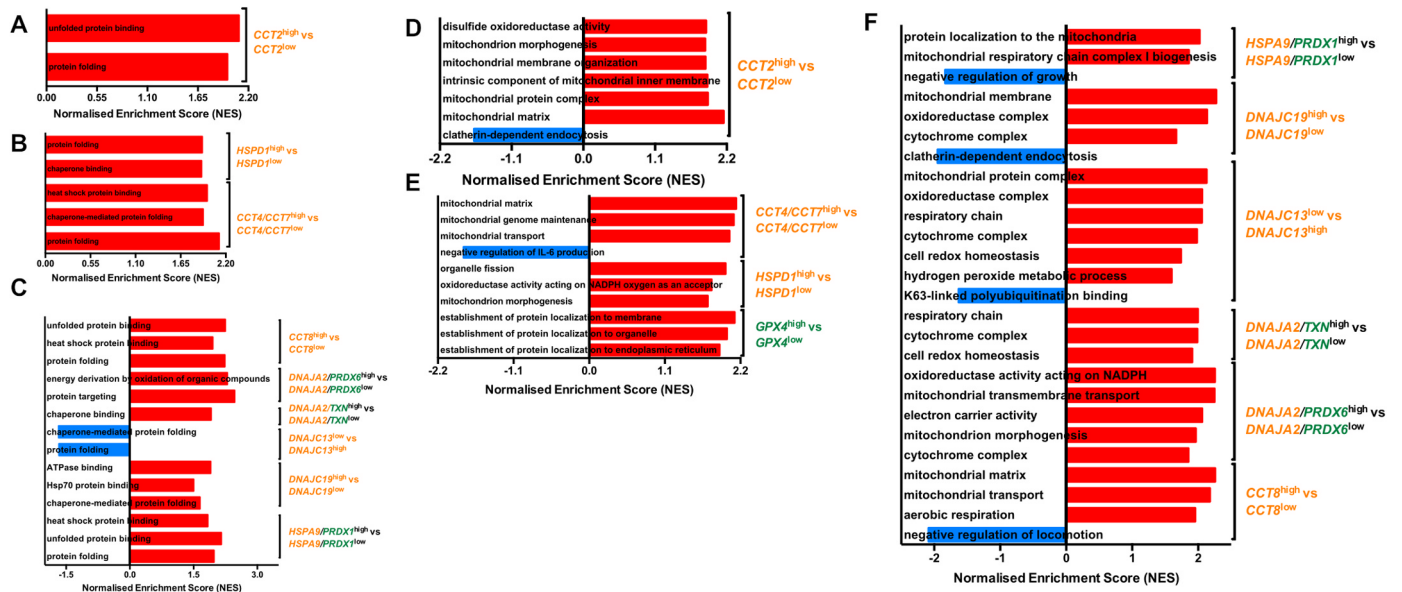

Fig. 7. Gene set enrichment analysis (GSEA) of HSPs and redox regulators in gastrointestinal cancers. Bar graphs of normalised enrichment scores (NES) from GSEA of various HSP^high vs HSP^low phenotypes in A) colorectal carcinoma B) oesophageal adenocarcinoma and C) stomach adenocarcinoma for validation of HSP related processes. Bar graphs of normalised enrichment scores (NES) from GSEA of various redox regulators^high/low/HSP^high/low vs redox regulators^high/low/HSP^low/high phenotypes in D) colorectal carcinoma E) oesophageal carcinoma and F) stomach adenocarcinoma. Heat shock protein genes indicated by yellow writing, redox regulators indicated by green writing.

colorectal cancer.

Our gene set enrichment data implicated many biological processes that are crucial for tumorigenesis in these cancer types surveyed, many being involved in mitochondrial functions, an organelle crucial in the progression of cancer [65]. In the colorectal dataset, the *CCT2*^high phenotype, which displayed poorer survival, was enriched in gene sets involved in mitochondrial organisation and membrane processes. While CCT2 overexpression has previously been shown to adversely affect colorectal patient survival [66], its role in mitochondrial membrane organisation is less clearly understood. A recent study in colorectal cancer identified *CCT2* as a synthetic lethal partner to mutant *KRAS*, which was found to be dependent on mitochondrial translation for viability [67]. Similarly, in oesophageal cancer, patients with *HSPD1*^high and *CCT4*^high/*CCT7*^high phenotypes displayed enrichment for gene sets involved in mitochondrial processes, such as organelle fission, transport and genome maintenance, which were associated with poorer survival. HSPD1 was identified as an enriched gene in the ESCC4 subtype of oesophageal carcinoma, which also has the poorest survival outcomes [68]. HSPD1 has been previously shown to be critical for mitochondrial protein folding [69]. It also functions as a protectant against pharma-cologically-induced oxidative stress [70]; therefore, combination therapy that initiates oxidative stress-induced apoptosis, such as myr-tucommulone while simultaneously inhibiting HSP60 may result in better outcomes [71]. Glutathione peroxidase family member GPX4 protects against lipid peroxidation and ferroptosis, and is pre-dominantly expressed in the mitochondria of the testis [72,73]. Over-expression of GPX4 in oesophageal cancer patients correlates with the upregulation of gene sets involved in protein localisation to the ER and membrane, a process that, if dysregulated, can contribute to carcino-genesis [74]. *GPX4* overexpression may represent a novel biomarker and a potential therapeutic target for the treatment of oesophageal cancer.

The stomach adenocarcinoma dataset revealed many combinations of HSPs and redox regulators that influenced patient survival. These were also associated with many biological processes involved in mi-tochondrial transport, metabolism and cellular redox homeostasis. The overexpression of *DNAJC19*, which encodes for the protein TIM14, has yet to be shown to influence oncogenesis or patient survival in stomach cancer. It functions as a translocase on the inner mitochondrial mem-brane [75], and GSEA on *DNAJC19*^high patients revealed enrichment of gene sets for mitochondrial membrane and oxidoreductase activity. Conversely, *DNAJC13* overexpression may benefit patient survival outcomes. The primary role of *DNAJC13* is in early endosome trans-portation, [76] and in *ERBB2* positive breast cancers has been im-plicated in the trafficking of epidermal growth factor receptor (EGFR) to the plasma membrane [77]. In this context, *DNAJC13* is seen as a potential therapeutic target, whereas the situation is reversed in sto-mach cancer; the enriched gene sets for the *DNAJC13*^low phenotype primarily included cell redox homeostasis and those involved in mi-tochondrial complexes and metabolism. Therefore, *DNAJC13* depletion may facilitate mitochondrial complex function and pro-oxidant state necessary to promote tumorigenesis. *CCT8*^high individuals were found to have a significantly poorer overall survival, and interestingly, the GSEA revealed that *CCT8*^low individuals were enriched for gene sets involved in the negative regulation of locomotion. This is consistent with the reduction in cell migration and invasion observed in glioma cells following siRNA-mediated knockdown of *CCT8* [78].

This study has uncovered a variety of novel HSP and redox genes that may influence patient outcomes and survival in gastrointestinal cancers. This high-throughput method of data extraction and analysis will become more valuable as more patient data becomes available. This study also sets a foundation for the analysis and development of multi-gene signatures that can predict patient survival, and ex-ponentially expand our list of biomarkers for the onset and progression of a variety of cancers.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.redox.2018.11.018.

References

[1] D. Whitley, S.P. Goldberg, W.D. Jordan, Heat shock proteins: a review of the mo-lecular chaperones, J. Vasc. Surg. 29 (1999) 748–751.

[2] Z. Li, P. Srivastava, Heat-shock proteins, Curr. Protoc. Immunol. (2004) (**Appendix 1**, Appendix 1T).

[3] M. Romanucci, T. Bastow, L. Della Salda, Heat shock proteins in animal neoplasms and human tumours–a comparison, Cell Stress Chaperon. 13 (2008) 253–262.

[4] S.K. Calderwood, J. Gong, Heat shock proteins promote cancer: it's a protection racket, Trends Biochem. Sci. 41 (2016) 311–323.

[5] G.D. Lianos, et al., The role of heat shock proteins in cancer, Cancer Lett. 360 (2015) 114–118.

[6] A. Bhattacharyya, R. Chattopadhyay, S. Mitra, S.E. Crowe, Oxidative stress: an es-sential factor in the pathogenesis of gastrointestinal mucosal diseases, Physiol. Rev. 94 (2014) 329–354.

[7] V. Dudeja, S.M. Vickers, A.K. Saluja, The role of heat shock proteins in gastro-intestinal diseases, Gut 58 (2009) 1000–1009.

[8] H.H. Kampinga, et al., Guidelines for the nomenclature of the human heat shock proteins, Cell Stress Chaperon-. 14 (2009) 105–111.

[9] X.B. Qiu, Y.M. Shao, S. Miao, L. Wang, The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones, Cell Mol. Life Sci. 63 (2006) 2560–2570.

[10] J. Radons, The human HSP70 family of chaperones: where do we stand? Cell Stress Chaperon-. 21 (2016) 379–404.

[11] F.H. Schopf, M.M. Biebl, J. Buchner, The HSP90 chaperone machinery, Nat. Rev. Mol. Cell Biol. 18 (2017) 345–360.

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

[13] L. Wei, et al., Hsp27 participates in the maintenance of breast cancer stem cells through regulation of epithelial-mesenchymal transition and nuclear factor-kappaB, Breast Cancer Res 13 (2011) R101.

[14] T. Cordonnier, et al., Hsp27 regulates EGF/beta-catenin mediated epithelial to mesenchymal transition in prostate cancer, Int. J. Cancer 136 (2015) E496–E507.

[15] C.G. Concannon, A.M. Gorman, A. Samali, On the role of Hsp27 in regulating apoptosis, Apoptosis 8 (2003) 61–70.

[16] A. Cuenda, S. Rousseau, p38 MAP-kinases pathway regulation, function and role in human diseases, Biochim. Biophys. Acta 1773 (2007) 1358–1375.

[17] A.P. Arrigo, B. Gibert, HspB1, HspB5 and HspB4 in human cancers: potent onco-genic role of some of their client proteins, Cancers 6 (2014) 333–365.

[18] H.L. Smith, W. Li, M.E. Cheetham, Molecular chaperones and neuronal proteostasis, Semin. Cell Dev. Biol. 40 (2015) 142–152.

[19] V. Sosa, et al., Oxidative stress and cancer: an overview, Ageing Res. Rev. 12 (2013) 376–390.

[20] S. Pervaiz, M.V. Clement, Superoxide anion: oncogenic reactive oxygen species? Int. J. Biochem. Cell Biol. 39 (2007) 1297–1304.

[21] Z.X. Chen, S. Pervaiz, Bcl-2 induces pro-oxidant state by engaging mitochondrial respiration in tumor cells, Cell Death Differ. 14 (2007) 1617–1627.

[22] M.V. Clement, J.L. Hirpara, S. Pervaiz, Decrease in intracellular superoxide sensi-tizes Bcl-2-overexpressing tumor cells to receptor and drug-induced apoptosis in-dependent of the mitochondria, Cell Death Differ. 10 (2003) 1273–1285.

[23] S. Pervaiz, J. Cao, O.S. Chao, Y.Y. Chin, M.V. Clement, Activation of the RacGTPase inhibits apoptosis in human tumor cells, Oncogene 20 (2001) 6263–6268.

[24] H. Hug, et al., Reactive oxygen intermediates are involved in the induction of CD95 ligand mRNA expression by cytostatic drugs in hepatoma cells, J. Biol. Chem. 272 (1997) 28191–28193.

[25] K.A. Ahmad, M.V. Clement, I.M. Hanif, S. Pervaiz, Resveratrol inhibits drug-in-duced apoptosis in human leukemia cells by creating an intracellular milieu non-permissive for death execution, Cancer Res 64 (2004) 1452–1459.

[26] E. Zito, ERO1: a protein disulfide oxidase and H2O2 producer, Free Radic. Biol. Med 83 (2015) 299–304.

[27] E. Gross, et al., Generating disulfides enzymatically: reaction products and electron acceptors of the endoplasmic reticulum thiol oxidase Ero1p, Proc. Natl. Acad. Sci. USA 103 (2006) 299–304.

[28] W.C. Chong, M.D. Shastri, R. Eri, Endoplasmic reticulum stress and oxidative stress: a vicious nexus implicated in bowel disease pathophysiology, Int. J. Mol. Sci. 18 (2017).

[29] M.P. Minisini, S. Kantengwa, B.S. Polla, DNA damage and stress protein synthesis induced by oxidative stress proceed independently in the human premonocytic line

U937, Mutat. Res. 315 (1994) 169–179.

[30] U. Jakob, W. Muse, M. Eser, J.C. Bardwell, Chaperone activity with a redox switch, Cell 96 (1999) 341–352.

[31] S.G. Ahn, D.J. Thiele, Redox regulation of mammalian heat shock factor 1 is essential for Hsp gene activation and protection from stress, Genes Dev. 17 (2003) 516–528.

[32] J. Gao, et al., Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal, Sci. Signal 6 (2013) (pl1).

[33] E. Cerami, et al., The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data, Cancer Discov. 2 (2012) 401–404.

[34] B. Chen, et al., Computational discovery of niclosamide ethanolamine, a repurposed drug candidate that reduces growth of hepatocellular carcinoma cells in vitro and in mice by inhibiting cell division cycle 37 signaling, Gastroenterology 152 (2017) 2022–2036.

[35] R. Chen, et al., A meta-analysis of lung cancer gene expression identifies PTK7 as a survival gene in lung adenocarcinoma, Cancer Res. 74 (2014) 2892–2902.

[36] A.J. Butte, S. Ito, Translational bioinformatics: data-driven drug discovery and development, Clin. Pharmacol. Ther. 91 (2012) 949–952.

[37] N. Cancer Genome Atlas Research, et al., Integrated genomic characterization of oesophageal carcinoma, Nature 541 (2017) 169–175.

[38] N. Cancer Genome Atlas Research, Comprehensive molecular characterization of gastric adenocarcinoma, Nature 513 (2014) 202–209.

[39] N. Cancer Genome Atlas, Comprehensive molecular characterization of human colon and rectal cancer, Nature 487 (2012) 330–337.

[40] S.O. Pohl, M. Agostino, A. Dharmarajan, S. Pervaiz, Cross talk between cellular redox state and the antiapoptotic protein Bcl-2, Antioxid. Redox Signal. (2018).

[41] A. Subramanian, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Natl. Acad. Sci. USA 102 (2005) 15545–15550.

[42] V.K. Mootha, et al., PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes, Nat. Genet 34 (2003) 267–273.

[43] M. Ashburner, et al., Gene ontology: tool for the unification of biology. The gene ontology consortium, Nat. Genet 25 (2000) 25–29.

[44] F. Supek, M. Bosnjak, N. Skunca, T. Smuc, REVIGO summarizes and visualizes long lists of gene ontology terms, PLoS One 6 (2011) e21800.

[45] A. Schlicker, F.S. Domingues, J. Rahnenfuhrer, T. Lengauer, A new measure for functional similarity of gene products based on Gene Ontology, BMC Bioinform. 7 (2006) 302.

[46] I. Fernandez, et al., Hic-5 mediates TGFbeta-induced adhesion in vascular smooth muscle cells by a Nox4-dependent mechanism, Arterioscler. Thromb. Vasc. Biol. 35 (2015) 1198–1206.

[47] A.F. Gil Lorenzo, V. Bocanegra, M.E. Benardon, V. Cacciamani, P.G. Valles, Hsp70 regulation on Nox4/p22phox and cytoskeletal integrity as an effect of losartan in vascular smooth muscle cells, Cell Stress Chaperon-. 19 (2014) 115–134.

[48] L.P. Desai, et al., Negative regulation of NADPH oxidase 4 by hydrogen peroxide-inducible clone 5 (Hic-5) protein, J. Biol. Chem. 289 (2014) 18270–18278.

[49] M. Muller, J. Gauley, J.J. Heikkila, Hydrogen peroxide induces heat shock protein and proto-oncogene mRNA accumulation in Xenopus laevis A6 kidney epithelial cells, Can. J. Physiol. Pharmacol. 82 (2004) 523–529.

[50] A.M. Courgeon, E. Rollet, J. Becker, C. Maisonhaute, M. Best-Belpomme, Hydrogen peroxide (H2O2) induces actin and some heat-shock proteins in Drosophila cells, Eur. J. Biochem. 171 (1988) 163–170.

[51] S. Motoyama, et al., Hydrogen peroxide induces midzonal heat shock protein 72 and apoptosis in sinusoidal endothelial cells of hypoxic rat liver, Crit. Care Med. 28 (2000) 1509–1514.

[52] A.J. Afolayan, et al., Inducible HSP70 regulates superoxide dismutase-2 and mitochondrial oxidative stress in the endothelial cells from developing lungs, Am. J. Physiol. Lung Cell Mol. Physiol. 306 (2014) L351–L360.

[53] A.P. Arrigo, Hsp27: novel regulator of intracellular redox state, IUBMB Life 52 (2001) 303–307.

[54] J.E. Souren, H. Van Aken, R. Van Wijk, Enhancement of superoxide production and protection against heat shock by HSP27 in fibroblasts, Biochem. Biophys. Res. Commun. 227 (1996) 816–821.

[55] G. Mellier, D. Liu, G. Bellot, A.L. Holme, S. Pervaiz, Small molecule sensitization to TRAIL is mediated via nuclear localization, phosphorylation and inhibition of chaperone activity of Hsp27, Cell Death Dis. 4 (2013) e890.

[56] K. Shenoy, Y. Wu, S. Pervaiz, LY303511 enhances TRAIL sensitivity of SHEP-1 neuroblastoma cells via hydrogen peroxide-mediated mitogen-activated protein kinase activation and up-regulation of death receptors, Cancer Res. 69 (2009) 1941–1950.

[57] S.B. Baylin, DNA methylation and gene silencing in cancer, Nat. Clin. Pract. Oncol. 2 (Suppl 1) (2005) S4–S11.

[58] P.O. Esteve, et al., Binding of 14-3-3 reader proteins to phosphorylated DNMT1 facilitates aberrant DNA methylation and gene expression, Nucleic Acids Res. 44 (2016) 1642–1656.

[59] N. Pencheva, et al., Convergent multi-miRNA targeting of ApoE drives LRP1/LRP8-dependent melanoma metastasis and angiogenesis, Cell 151 (2012) 1068–1082.

[60] M.R. Ray, et al., Cyclin G-associated kinase: a novel androgen receptor-interacting transcriptional coactivator that is overexpressed in hormone refractory prostate cancer, Int. J. Cancer 118 (2006) 1108–1119.

[61] D.W. Lee, X. Zhao, Y.I. Yim, E. Eisenberg, L.E. Greene, Essential role of cyclin-G-associated kinase (Auxilin-2) in developing and mature mice, Mol. Biol. Cell 19 (2008) 2766–2776.

[62] S. Floyd, P. De Camilli, Endocytosis proteins and cancer: a potential link? Trends Cell Biol. 8 (1998) 299–301.

[63] J. Li, W. Xu, F. Liu, S. Huang, M. He, GSTM1 polymorphism contribute to colorectal cancer in Asian populations: a prospective meta-analysis, Sci. Rep. 5 (2015) 12514.

[64] M.N. Khabaz, et al., GSTM1 gene polymorphism and the risk of colorectal cancer in a Saudi Arabian population, Genet Mol. Res. 15 (2016).

[65] S. Vyas, E. Zaganjor, M.C. Haigis, Mitochondria and cancer, Cell 166 (2016) 555–566.

[66] C. Coghlin, et al., Characterization and over-expression of chaperonin t-complex proteins in colorectal cancer, J. Pathol. 210 (2006) 351–357.

[67] T.D. Martin, et al., A role for mitochondrial translation in promotion of viability in K-Ras mutant cells, Cell Rep. 20 (2017) 427–438.

[68] T. Xiong, et al., An esophageal squamous cell carcinoma classification system that reveals potential targets for therapy, Oncotarget 8 (2017) 49851–49860.

[69] J. Ostermann, A.L. Horwich, W. Neupert, F.U. Hartl, Protein folding in mitochondria requires complex formation with hsp60 and ATP hydrolysis, Nature 341 (1989) 125–130.

[70] U. Sarangi, et al., Hsp60 chaperonin acts as barrier to pharmacologically induced oxidative stress mediated apoptosis in tumor cells with differential stress response, Drug Target Insights 7 (2013) 35–51.

[71] K. Wiechmann, et al., Mitochondrial chaperonin HSP60 Is the apoptosis-related target for myrtucommulone, Cell Chem. Biol. 24 (2017) 614–623 (e616).

[72] W.S. Hambright, R.S. Fonseca, L. Chen, R. Na, Q. Ran, Ablation of ferroptosis regulator glutathione peroxidase 4 in forebrain neurons promotes cognitive impairment and neurodegeneration, Redox Biol. 12 (2017) 8–17.

[73] H. Imai, et al., Depletion of selenoprotein GPx4 in spermatocytes causes male infertility in mice, J. Biol. Chem. 284 (2009) 32522–32532.

[74] M.C. Hung, W. Link, Protein localization in disease and therapy, J. Cell Sci. 124 (2011) 3381–3392.

[75] D. Mokranjac, M. Sichting, W. Neupert, K. Hell, Tim14, a novel key component of the import motor of the TIM23 protein translocase of mitochondria, EMBO J. 22 (2003) 4945–4956.

[76] A. Fujibayashi, et al., Human RME-8 is involved in membrane trafficking through early endosomes, Cell Struct. Funct. 33 (2008) 35–50.

[77] M. Girard, P.S. McPherson, RME-8 regulates trafficking of the epidermal growth factor receptor, FEBS Lett. 582 (2008) 961–966.

[78] X. Qiu, et al., Overexpression of CCT8 and its significance for tumor cell proliferation, migration and invasion in glioma, Pathol. Res. Pract. 211 (2015) 717–725.