Redox Biology 81 (2025) 103565

Contents lists available at ScienceDirect

Redox Biology

journal homepage: www.elsevier.com/locate/redox

Cleavage site heterogeneity at the pre-mRNA 3′-untranslated region regulates gene expression in oxidative stress response

Feba Shaji [a,b], Jamshaid Ali [a], Rakesh S. Laishram [a,*]

[a] Rajiv Gandhi Centre for Biotechnology, Cardiovascular Biology Group, Trivandrum, 695014, India
[b] Regional Centre for Biotechnology, Faridabad, Haryana, 121001, India

ARTICLE INFO

Keywords:
3′-end processing
Cleavage and polyadenylation
Cleavage site
Cleavage heterogeneity
Oxidative stress response
Star-PAP
NQO1
Antioxidant response genes
Gene expression

ABSTRACT

The endonucleolytic cleavage step of the eukaryotic mRNA 3′-end processing is considered imprecise, which leads to heterogeneity of cleavage site (CS) with hitherto unknown function. Contrary to popular belief, we show that this imprecision in the cleavage is tightly regulated, resulting in the CS heterogeneity (CSH) that controls gene expression in antioxidant response. CSH centres around a primary CS, followed by several subsidiary cleavages determined by CS's positions. Globally and using reporter antioxidant mRNA, we discovered an inverse relationship between the number of CS and the gene expression, with the primary CS exhibiting the highest cleavage efficiency. Strikingly, reducing CSH and increasing primary CS usage induces gene expression. Under oxidative stress (we employ three conditions that induce antioxidant response, tBHQ, H_2O_2, and $NaAsO_2$) conditions, there is a decrease in the CSH and an increase in the primary CS usage to induce antioxidant gene expression. Key oxidative stress response genes (*NQO1*, *HMOX1*, *PRDX1*, and *CAT*) also show higher CSH compared to the non-stress response genes and that the number of CSs are reduced to impart cellular response to oxidative stresses. Concomitantly, ectopic expression of one of the key antioxidant response gene (*NQO1*) driven by the primary CS but not from other subsidiary CSs, or reduction in CSH imparts tolerance to cellular oxidative stresses (H_2O_2, and $NaAsO_2$). Genome-wide CS analysis of stress response genes also shows a similar result. Compromised CSH or CSH-mediated gene control hampers cellular response to oxidative stress. We establish that oxidative stress induces affinity/strength of cleavage complex assembly, increasing the fidelity of cleavage at the primary CS, thereby reducing CSH inducing antioxidant response. Together, our study reports a novel cleavage imprecision- or CSH-mediated anti-oxidant response mechanism that is distinct and operates downstream but in concert with the transcriptional pathway of oxidative stress induction.

1. Introduction

Processing at the 3′-end of a precursor messenger RNA (pre-mRNA) plays a pivotal role in gene expression[1,2]. The processing reaction involves two steps: cleavage and polyadenylation (CPA). These steps require cleavage and polyadenylation specificity factor (subunits CPSF160, WDR33, CPSF100, hFIP1, CPSF73, and CPSF30), cleavage stimulation factor (subunits CstF77, CstF64, and CstF50), cleavage factor (CFIm and CFIIm), scaffolding protein symplekin, poly(A) binding protein (PABPN1), and poly(A) polymerase (PAP) as core components [1,3,4]. While CPSF, CstF, PAP, CFIm, and IIm are involved in the cleavage reaction, polyadenylation requires PAP, CPSF, and PABP[5–8]. CPSF subunits WDR33 and CPSF30 recognises the poly(A) signal (PA-signal) hexamer (AAUAAA) to assemble the CPA complex at the poly(A) site (PA-site)[9–11]. CPSF160 cooperates with CstF64 that binds the downstream sequence element (DSE, GU/U-rich ~15–30 nucleotides downstream of the PA-signal)[6,12,13]. Along with CFIm, CFIIm, and other core components, CPSF assembles the CPA complex [14–16]. CPSF73 then cleaves the pre-mRNA at the cleavage site (CS), followed by PAP-mediated PA-tail addition and PABPN1 coating of the PA-tail[17–20]. Canonical PAPα/γ is the primary PAP for mRNA polyadenylation in the nucleus[21–24]. The discovery of a variant PAP, Star-PAP, has revealed alternative PAPs for nuclear polyadenylation[25,26].

CPA has a crucial role in gene expression primarily through 1. control of mRNA metabolism and translation, 2. co-ordination with transcription and other RNA processing events, and 3. alternative polyadenylation[1,27–31]. The PA-tail serves at least three functions:

* Corresponding author. author. Rajiv Gandhi Centre for Biotechnology, Melaranoor Road, Jagathy, Trivandrum, Kerala, 695014, India.
E-mail address: laishram@rgcb.res.in (R.S. Laishram).

https://doi.org/10.1016/j.redox.2025.103565
Received 17 February 2025; Accepted 23 February 2025
Available online 24 February 2025

first, PA-tail, through its coating with PABP, protects the mRNA from 3′- to 5′-exonucleases along with the 5′-end m7G cap; second, it serves as the initiation site for mRNA turnover by deadenylation; and third, PA-tail shortening precedes decapping for mRNA decay[31–33]. In addition to the stability, PA-tail coating by PABP facilitates mRNA nuclear exit and also acts as an enhancer of mRNA translation initiation by association with transcription initiation factors[34–37]. Additionally, many of the core CPA components are detected at the promoter along with PolII and also couple with cap-binding protein, forming a closed loop to coordinate with transcription initiation[38–41]. Furthermore, several CPA components assemble with spliceosomal components, which stimulates intron splicing[42–46]. Apart from these mechanisms, CPA also regulates gene expression through alternative polyadenylation (APA), where a gene encodes for more than one transcript with different lengths due to the presence of more than one PA-site[47–49]. APA alters the primary structure of a protein or transcript length, affecting stability, translation, localisation, and encoded protein interaction[47–49].

While cleavage and polyadenylation are two coupled steps in the 3′-end processing, the cleavage step is considered imprecise and typically occurs within 10–25 nucleotides downstream of the PA-signal [50,51]. The positioning of the PA-signal and the DSE limits the region where cleavage can occur [51]. Nevertheless, the predominant nucleotide at the CS appears as adenosine followed by uridine, cytosine, and guanosine (A > U > C ≫ G), and the dinucleotide composition at the CS (CA > AA > TA ≫ GA) in the order[50]. Intriguingly, the imprecision in the cleavage leads to multiple cleavage events instead of cleaving at only one CS between the PA-signal and the DSE, resulting in cleavage site heterogeneity (CSH) [50,52,53]. Unlike in APA, where more than one PA-site is alternatively used at the 3′-UTR leading to distinct mRNA isoforms with different UTR lengths, in CSH, cleavage occurs at multiple nucleotide sequences between the PA-signal hexamer and the DSE of a single PA-site [50,53](Schematics to clarify CSH, APA and PA-site is shown in Supplementary Fig. 1A). CSH neither leads to discernible difference in the UTR length nor generate significant mRNA isoforms[50, 52,53]. It is considered stochastic, and its mechanism, regulation, or cellular ramifications of the imprecision in cleavage is unclear. Nevertheless, both gain and loss of function mutations of CS are reported in various diseases[54–56]. Therefore, we investigate the cellular and mechanistic implications of CSH in gene expression.

Global analysis shows a tight regulation of CSH wherein cleavage centres around a primary CS followed by several subsidiary CSs. This CSH is linked with the fidelity of cleavage such that mutation of subsidiary CSs increases cleavage at the primary CS. In a global analysis of total cellular mRNAs and reporter mutational analysis of an antioxidant mRNA (*NQO1*), we discovered an inverse relationship between the number of CS and the gene expression. The primary CS shows the highest cleavage efficiency, such that reducing CSH and increasing primary CS usage induces gene expression. Cleavage at the primary CS, and hence, cleavage efficiency is determined by the position of the CS downstream of the PA-signal and not by the sequence composition. This mechanism regulates oxidative stress response where increased antioxidant protein expression involves a reduction in the CSH and an enhanced primary CS usage. Strikingly, key oxidative stress response genes (*NQO1*, *HMOX1*, *PRDX1*, and *CAT*) show higher CSH compared to the non-stress response genes and that the number of CSs are reduced to impart cellular response to oxidative stresses. In the cellular assays, cells expressing one of the key antioxidant response gene (*NQO1*) driven by the primary CS but not from other subsidiary CSs, or with reduced CSH imparts cellular tolerance to H_2O_2 and $NaAsO_2$ stress. We then generated a CSH-mutant in the poly(A) polymerase, Star-PAP that compromises cellular CSH or CSH-mediated gene control, and demonstrated that compromised CSH hampers cellular response to oxidative stress. We show that oxidative stress induces affinity/strength of cleavage complex assembly, increasing the fidelity of cleavage at the primary CS, thereby reducing CSH-inducing antioxidant response. The CSH-mediated stress response mechanism is distinct but operates downstream and in concert with the transcriptional pathway of oxidative stress induction. Thus, our study reports a novel mechanism of cellular oxidative stress response that operates through the regulated imprecision of the endonucleolytic cleavage reaction at the mRNA 3′-end. To our knowledge, this is the first example of a biological significance of CS imprecision or heterogeneity that regulates gene expression in antioxidant response.

2. Results

2.1. Global cleavage site analysis reveals the existence of cleavage site heterogeneity in a majority of the cellular mRNAs covering nearly half of the PA-sites genome-wide

To understand the regulation and ramifications of imprecisions in the endonucleolytic cleavage step leading to CSH in gene expression, we first analysed the global 3′-READS (3′-Region Extraction and Deep Sequencing) data (Accession Number: GSE84461) for the CS selection in each PA-site genome-wide in HEK293 cells [57]. It was also confirmed in Jurkat cells (GSE138290) and NIH3T3 cells (GSE101851). We define a PA-site as a region at the 3′-UTR where cleavage and polyadenylation occur, encompassing the PA-signal and other *cis*-elements required for CPA reaction; CS is defined as the actual nucleotide where cleavage occurs, and PA-tail is added such that a transcript from a particular a PA-site can have either one or more CS (Supplemental Fig. 1A). For clarity, schematics of CSH and comparison with that of APA is shown in Fig. 1A and Supplemental Fig. 1A. In CSH, cleavage of a PA-site occurs at more than one CS within 10–30 nucleotides downstream of the AAUAAA PA-signal resulting in heterogenous cleavages. Thus, a single PA-site exhibits multiple CSs with no significant difference in the apparent mRNA length. Whereas, in APA, different PA-sites at the 3′-UTR are alternatively used, which leads to distinct mRNA isoforms with differences in the UTR lengths [48]. Each PA-site in APA can exhibit CSH. Altogether, our analysis obtained >175,000 CSs clustered across >70, 000 PA-sites on 12,000 genes, indicating the existence of CSH globally.

Our analysis revealed two classes of PA-sites on mRNAs, the first having only one CS and the other mRNAs having more than one CS (confirming the occurrence of CSH). Circos plot of three experimental replicate datasets showed the genome-wide presence of CSH on numerous PA-sites (Fig. 1B). The outermost track of the plot represents the chromosome ideogram showing the physical location of each gene on the chromosome. The second track represents the location and the number of PA-sites of different mRNAs from the PolyA DB database corresponding to respective chromosomal locations. The three inner tracks correspond to the CSs on respective PA-sites from three replicate datasets plotted against the genomic locations of the outer tracks. Each line in the plot represents a PA-site and the radial length of the line in each track corresponds to the number of CS, thus representing heterogeneity. We observed that the position and length of the lines from various replicates occurred (overlapped) at similar chromosomal positions (Fig. 1B). Similarly, a jitter plot also showed distribution of CSH across different chromosomes of the same datasets. In this plot, each dot represents a PA-site with distinct number of CSs, and the size and the density of the dot shows the number CSs associated with different PA-sites on each chromosome (X-axis shows the chromosome number while Y-axis represents the number CSs detected on each PA-site). There was a decrease in the number of PA-sites associated with increasing number of CS on different chromosomes (three different datasets are shown in different colours) (Supplemental Fig. 1B). CS clusters ranged primarily from 1 to 20 on different PA-sites, where a majority of the PA-sites had 3 to 6 CS with a median of ~2.8 CS per PA-site (Fig. 1C, Supplemental Fig. 1B). The number of CSs determine the degree of heterogeneity of cleavage. We observed more PA-sites with low heterogeneity (fewer CS) than those with higher heterogeneity (more CS) (Fig. 1C, Supplemental Fig. 1B). Overall, CSH was observed on 10944 mRNAs (majority of mRNAs) at different PA-sites of which around half (~40 %, 30,500 out of 70,000 PA-sites on different mRNAs) exhibited

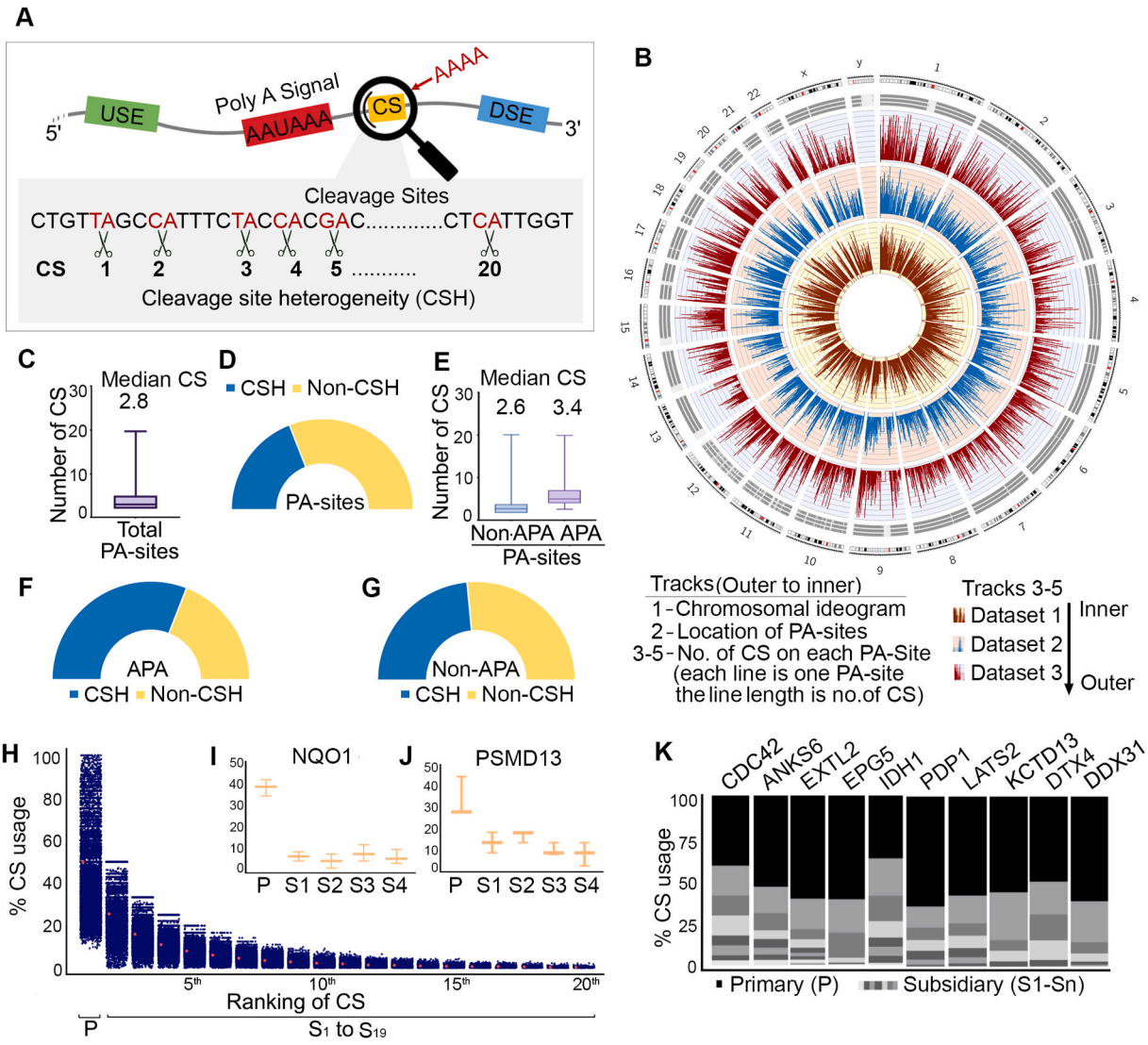


Fig. 1. Global cleavage site analysis reveals heterogeneity in the majority of the cellular mRNAs, covering nearly half of the PA-sites genome-wide (*A*) Schematics illustrating cleavage site heterogeneity (CSH) of a particular PA-site. Various *cis*-elements in the 3′-UTR of RNA [PA-signal, downstream sequence element (DSE), and cleavage site (CS), upstream elements (USE) are shown]. A zoomed-in view highlights different CSs that exhibit CSH. (*B*) Circos plot to visualise the genome-wide distribution of CSH of different PA-sites against the chromosomal locations. This plot examines the chromosomal positions of the PA-sites and the corresponding CSs across three different replicate 3′-READS datasets (materials and methods). The plot consists of five concentric tracks: outermost track of the plot shows the chromosome ideogram with the physical location of each gene across the genome; the second track represents the location and the number of PA-sites from the PolyA DB database corresponding to respective chromosomal locations of each gene; and the remaining three inner tracks represent experimental data of CSs from three datasets that are plotted against the genomic location of the outer tracks. Each line in the three inner tracks represents each PA-site and the length of the line corresponds to the number of CSs, thus representing heterogeneity. The circumference (X-axis) shows the location of the PA-site and the radial length (Y axis) represents the number of CS on respective PA-sites. (*C*) Box plot showing the distribution pattern of number of CS clusters (>175,000) across total PA-sites (>70,000) genome-wide. The median number of CS is indicated at the top. (*D*) Half donut plot showing % of the total number of PA-sites that show CSH (having more than one CS) (blue) versus those that do not show CSH (having only one CS) (yellow). (*E*) Box plot, as in C, comparing the CS distribution pattern on PA-sites of APA-dependent and -independent mRNAs. The median number of CS of PA-sites from APA-dependent and-independent mRNAs is indicated. (*F*) Half donut plot of PA-sites from APA-dependent mRNAs. Blue and yellow colours represent the percentage PA-sites that show CSH and that do not show CSH (non-CSH), respectively. (*G*) Half donut plot of PA-sites from non-APA-regulated mRNAs. Blue and yellow colours represent the % of PA-sites with CSH and non-CSH, respectively. (*H*) Dot plot showing usage distribution pattern of different CSs (ranked from 1 to 20 based on the percentage usage of each CS relative to the total reads from all CSs of each PA-site) (materials and methods). The CS with the highest usage is assigned 1st (primary CS) followed by CS numbers based on their decreasing usage (2nd to 20th) (subsidiary CS). (*I-J*) Box plot of % usage of different CS of terminal PA-site of *NQO1* mRNA (chromosomal locations of each CS are P: 69709403, S1: 69709401, S2: 69709407, S3: 69709409 and S4: 69709417) and *PSMD1*3 mRNA (chromosomal locations of each CS are P: 252984, S1: 252979, S2: 252980, S3: 252981 and S4: 252982). (*K*) Stacked column plots showing the percentage usage of primary versus subsidiary CSs of 10 select mRNAs (*CDC42, ANKS6, EXTL2, EPG5, IDH1, PDP1, LATS2, KCTD13, DTX4, DDX31*). The primary CS is indicated in black, and the different subsidiary CS are shown in various shades of grey.

CSH (Fig. 1D). Existence of global CSH was also confirmed in two additional cells: Jurkat cells and NIH3T3 cells and is detailed in the oxidative stress response section below. Intriguingly, PA-sites of alternatively polyadenylated mRNAs (APA-dependent) showed more heterogeneity than that of the APA-independent mRNAs (Fig. 1E). The median number of CS (cleavage per PA-site) of the APA-dependent mRNAs was 3.4 compared to the median CSH of 2.6 of the non-APA independent mRNAs (Fig. 1E). While 60 % of the PA-sites on the APA-dependent mRNAs exhibited CSH, only around ~45 % of the PA-sites on non-APA mRNAs showed CSH (Fig. 1F and G). As an example of

CSH, *EXOC3* mRNA with nucleotide details and the number of CSs is shown in Supplemental Fig. S1C. Further, we examined the heterogeneity pattern of target PA-sites of two major PAPs in the nucleus (canonical PAPα and the non-canonical Star-PAP) that polyadenylate distinct target mRNAs. We did not see any significant difference in the CSH pattern between the PA-sites regulated by Star-PAP and that of PAPα (Supplemental Figs. S1D–E). The median CS among the targets of the two PAPs was ~2.8, indicating a similar CSH pattern between the target PA-sites of the two PAPs. Together, these results reveal a global CSH with a higher CSH of PA-sites of APA-dependent than non-APA-dependent mRNAs.

2.2. Cleavage site heterogeneity is tightly regulated and prefers one primary site over other subsidiary sites

Strikingly, among the heterogeneous CS, cleavage occurs predominantly at a single CS that surpasses cleavages at other sites genome-wide (Fig. 1H). We refer to this preferential CS as the "primary CS" and the other lesser-used CS of a PA-site as the "subsidiary CS" such that a PA-site can have a primary site and several subsidiary sites. Dot plot analysis of CSs ranked according to their usage shows a predominance of the primary CS site over other subsidiary sites, with a marked decrease in the usage from primary CS to various subsidiary sites (Fig. 1H). The distribution of primary CS usage ranged from ~20 % to 99 % with a median usage of 65 %. In contrast, the subsidiary CS usage ranged from 1 % to 50 % with an overall median usage of ~14 % for the subsidiary CS. Analysis of the CSH on individual genes also showed similar observations. For example, on *NQO1* mRNA 3′-UTR, the terminal PA-site has five major CSs, one primary and four subsidiary sites (Fig. 1I). The primary CS was cleaved >40 % on average, while the cleavage on the other subsidiary sites was <5 % each. Similarly, *PSMD13* mRNA showed 40 % primary CS usage among the total 5 CSs (Fig. 1J). CSH analysis of 10 select mRNAs (*CDC42*, *ANKS6*, *EXTL2*, *EPG5*, *IDH1*, *PDP1*, *LATS2*, *KCTD13*, *DTX4*, and *DDX31*) also showed similar preferential cleavage at the primary CS over the subsidiary sites (Fig. 1K). Further, we validated our genome-wide data of CSH by 3′-RACE assays of 3 genes (*HMOX1*, *NQO1*, and *PRDX1*) followed by Sanger sequencing of the PA-tailed mRNA (Supplemental Fig. S1F). We observed a similar CSH pattern with consistent primary and subsidiary CS positions in both 3′-RACE and 3′-READS data (Supplemental Fig. S1G). Together, these results confirm that CSH is tightly regulated for CS selection in the cell.

2.3. Cleavage site heterogeneity on target PA-sites is closely linked with the expression of the corresponding mRNA

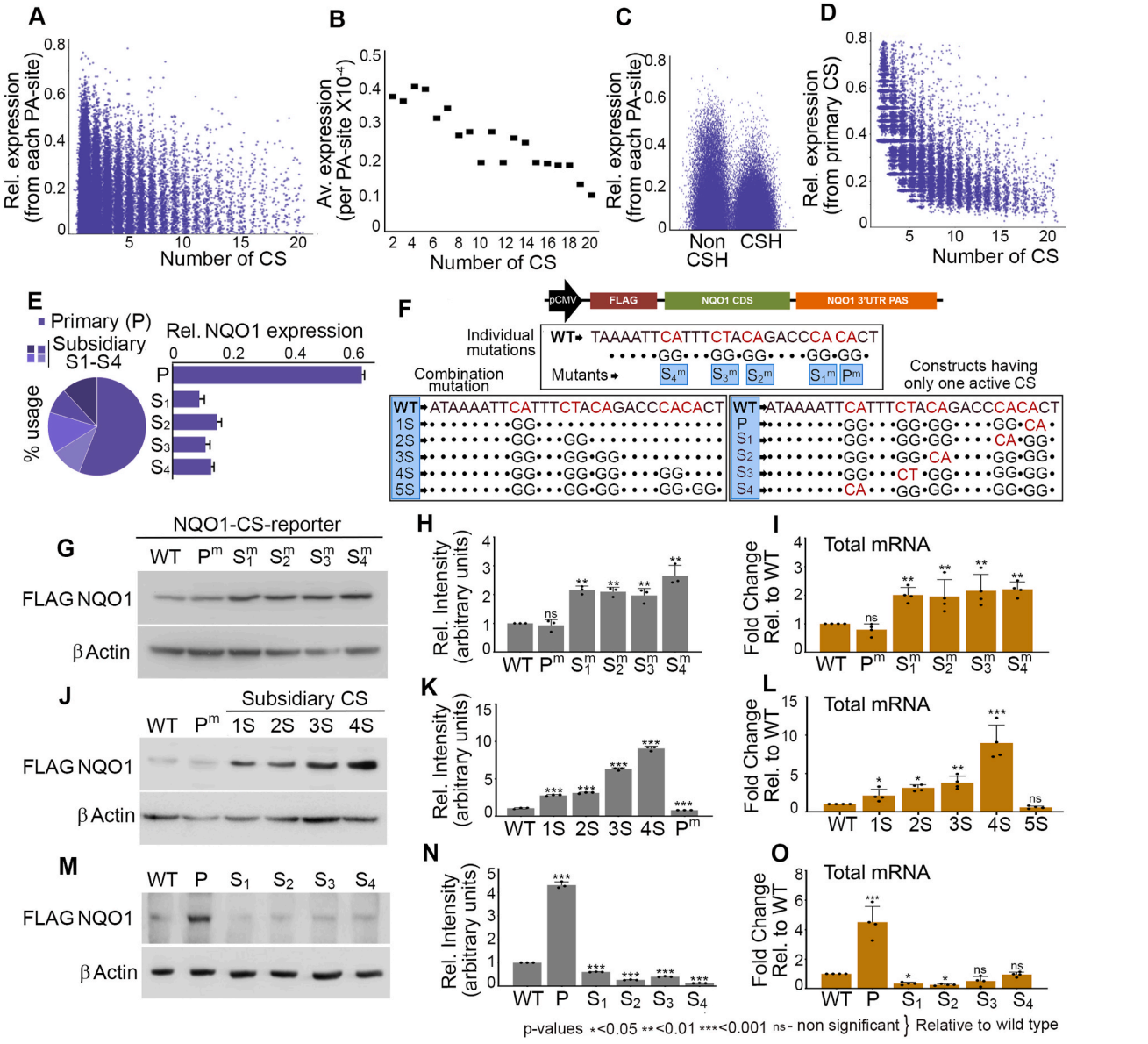
To understand the significance of CSH, we analyse the relative gene expression from each CS. We assessed the total read counts per million (CPM) of each CS-specific transcript of a PA-site relative to constitutive *GAPDH* level and expressed over the total expression level of the particular mRNA. Remarkably, we observed an inverse correlation between CSH and gene expression. There was a gradual decrease in the expression with an increase in the number of CS (Fig. 2A). There was a higher relative expression from mRNA isoforms with lower heterogeneity (<10 CS per PA-site) than that of higher heterogeneity, and the expression from the PA-sites with CS beyond 15 were marginal relative to total mRNA expression (Fig. 2A). A plot of average relative expression per PA-site also showed the highest expression with mRNAs with lower CSH (2–6 CS), a moderate expression with medium CSH (7–15 CS) and lowest expression with highest CSH (>18 CS) (Fig. 2B). Consistently, we observed a higher relative expression from transcripts with no CSH (having one CS) compared to those transcripts with multiple CSs (Fig. 2C). Further, randomly selected mRNA isoforms with different CSH level showed a linear decrease in the expression with the increasing number of CS (Supplemental Fig. S2A). Additionally, since primary CS is predominantly used, we then analysed the expression from the primary CS of different PA-sites. The expression from the primary CS of PA-sites having low CSH was higher than that of PA-sites having high CSH, and a gradual decrease in the expression was observed with increasing number of CS (Fig. 2D). Analysis of individual genes (*NQO1*, *HMGA2*, and *EIF3A*) also showed overwhelmingly higher expression from the primary CS compared to the subsidiary CS (Fig. 2E, Supplemental Figs. S2B–C). Together, these results reveal an inverse correlation between CSH and gene expression.

2.4. Mutational analysis of heterogeneous cleavage sites confirms an inverse correlation between the cleavage site heterogeneity and protein expression

To understand the effect of CSH on gene expression, we analysed the expression of *NQO1* mRNA (that encodes an essential antioxidant protein) and specific CSs. There are five major CS (one primary, P, and four subsidiaries, S_1 to S_4) on the terminal PA-site of *NQO1* pre-mRNA. The primary CS usage was >55 %, and there was a corresponding higher expression from the primary CS (Fig. 2E). We employed a reporter mini gene construct of FLAG-tagged *NQO1* CDS under the constitutive CMV promoter and driven by the *NQO1* 3′-UTR with the terminal PA-site as described earlier[58]. We first introduced mutations in each CS (CS was mutated from CA to GG, the least efficient sequence for cleavage) [56] individually by site-directed mutagenesis (Fig. 2F, top panel). The list of mutations and sequences of mutations are listed in Supplemental Table S1. Reporter expression in HEK293 cells was then analysed by Western blotting using an anti-FLAG antibody and qRT-PCR using a pair of primers (a forward primer from the FLAG-coding sequence and an *NQO1* CDS-specific reverse primer). To our surprise, individual mutations of each subsidiary CS (S_1^m to S_4^m) induced the reporter NQO1 protein and mRNA expressions (Fig. 2G–I). However, there was no marked induction of *NQO1* reporter expression after the mutation of the primary CS (P^m). Next, we analysed the reporter NQO1 expression from the combination of mutations with an increasing number of subsidiary CS mutations (1S, 2S, 3S, and 4S) (Fig. 2F bottom left panel). Strikingly, we observed a gradual increase in the reporter NQO1 protein and mRNA expression with the increasing number of subsidiary CS mutations (Fig. 2J–L). Mutation of all five sites, including the primary CS, resulted in the insignificant expression of the NQO1 reporter construct (Fig. 2J–L, Supplemental Fig. S2D). Further, to assess the expression from primary CS versus subsidiary CS, we generated CS mutants that have only one active CS where the other four CSs were mutated [reporter driven by only one CS, either primary CS (P) or individual subsidiary CS (S_1–S_4)] (Fig. 2F bottom right panel, Supplemental Table S1). We saw the highest reporter NQO1 protein and mRNA expression from the primary CS driven *NQO1* reporter construct, whereas the expressions from other subsidiary CS-driven constructs were minimal (Fig. 2M–O). Consistently, there was no discernible induction of the reporter NQO1 expression with an increasing number of subsidiary CS mutations after the primary CS was mutated (Supplemental Figs. S2D–E). Together, these results confirm an inverse correlation between the number of CS and the gene expression and the predominance of primary CS for the gene expression.

2.5. A decrease in the cleavage site heterogeneity and a concomitant increase in cleavage at the primary cleavage site induces gene expression

To understand the mechanism of the inverse correlation between the gene expression and the CSH, we analysed the cellular cleavage of the reporter *NQO1* pre-mRNA construct. We saw an increase in the cleavage efficiency with each subsidiary CS mutation when the primary CS was active in agreement with the increased reporter expression (Fig. 3A). Moreover, there was a gradual decrease in the uncleaved pre-mRNA on increasing number of subsidiary CS mutations (Fig. 3B). However, in the presence of the primary CS mutation, subsidiary CS mutations did not show significant effect on the cleavage efficiency (Supplemental Figs. S2D–E). Moreover, we also observed the highest cleavage



(caption on next page)

Fig. 2. Cleavage site heterogeneity at the target PA-site regulates gene expression exhibiting an inverse correlation between the number of CS and gene expression (*A***)** Dot plot showing relative expression from different PA-sites versus the number of CSs on respective PA-sites. Gene expression from each PA-site is obtained from the total read counts per million (CPM) of CS-specific transcripts for all CSs associated with a PA-site normalised to the CPM value of the internal *GAPDH* control. Relative expression from each PA-site is then calculated by dividing the CPM value of a PA-site by the total CPM values of the corresponding mRNA accounting for all PA-sites associated with the gene. Here, the relative expression from each PA-site is plotted against an increasing number of CSs on respective PA-sites. (*B*) Plot showing average expression per PA-site versus the number of CSs on the respective PA-sites. The mean relative expression from all the PA-sites (total relative CPMs of all PA-sites divided by the number of PA-sites) is plotted against the increasing number of CSs. (*C*) Relative expression, as in A, from PA-sites that show CSH versus those that do not show CSH. (*D*) Relative expression from the primary CS plotted against the increasing number of CSs on different PA-sites. Relative expression level from a CS is determined by dividing the CPM value of a CS-specific transcript normalised to *GAPDH* by the total CPM value of all CSs of the corresponding PA-site-specific mRNA isoform. Then, the relative expression values of the primary CS is plotted against the increasing number of CSs associated with respective PA-sites. (*E*) Pie chart showing relative usage of different CS (primary versus subsidiary CS) of *NQO1* mRNA (left) and the bar chart of expressions from different *NQO1* CSs relative to total *NQO1* expression as indicated (right). (*F*) Schematics of *NQO1* terminal PA-site region showing different CSs and individual and combination mutations (see Supplemental Table S1). Mutant sequences are written below the WT sequence. The dot (•) represents an unchanged sequence. Active CS is represented in red, and mutated CSs are represented in black. Mutant nomenclatures are highlighted in blue. (*G*) Immunoblotting of the NQO1 reporter expression from HEK293 cells using anti-FLAG-antibody with individual mutation of the primary and different subsidiary CSs as indicated. Control β-Actin is shown below. (*H*) Quantification of the NQO1 blot in G. Relative intensities of FLAG-NQO1 bands in the blot is quantified using Image J software in arbitrary units and normalised with control β-Actin blot and expressed as fold-change relative to wild-type. (*I*) qRT-PCR analysis of *NQO1* reporter expression using a pair of primers (forward from the FLAG-epitope tag coding region and a reverse primer specific to *NQO1* CDS) from HEK293 cells after overexpression of wild-type and different individual CS mutations as in G. mRNA levels are expressed as fold-change relative to wild-type. The error bar represents the standard error mean (SEM) of n = 3 independent experiments. (p values * <0.05, ** <0.01, *** <0.001, ns - non significant; p-values are relative to control). (*J*) Immunoblotting of the *NQO1* reporter expression from HEK293 cells using anti-FLAG-antibody with the CS combination mutations as indicated (also see Supplemental Table S1). Control β-Actin is shown below. Quantification of the blot in arbitrary units normalised with control β-Actin and expressed as fold-change over wild-type is shown in K. (*L*) qRT-PCR analysis of *NQO1* reporter expression using a pair of primers as in I from HEK293 cells after overexpression of wild-type and different combination CS mutations. mRNA levels are expressed as fold-change relative to the wild-type. The error bar represents the standard error mean (SEM) of n = 3 independent experiments. (p values * <0.05, ** <0.01, *** <0.001, ns - non significant; p-values are relative to control). (*M-O*) Immunoblotting followed by quantification and qRT-PCR analysis as in G-I of NQO1 reporter expression from primary and different subsidiary sites (only one CS present; p - only primary site is present and other subsidiary sites are mutated, S_1 to S_4 – only one respective subsidiary CS is active as indicated and other CSs are mutated) in HEK293 cells. The error bar represents the standard error mean (SEM) of n = 3 independent experiments. (p values * <0.05, ** <0.01, *** <0.001, ns - non significant; p-values are relative to control).

efficiency from the primary CS compared to the other subsidiary CS (Fig. 3C). Concomitantly, wherever primary CS is present, predominant cleavage was observed from the primary CS irrespective of the number of subsidiary CS mutations (Supplemental Fig. S2F). Without the primary CS, there was no predominant CS among the subsidiary CSs. There was a gradual increase in the cleavage at the primary CS with an increase in the number of subsidiary CS mutations (Supplemental Figs. S2G–H). However, when primary CS was mutated, there was no marked increase in the cleavage at any subsidiary CS, indicating that primary CS essentially drives the cleavage at the 3′-end. Together, these results reveal that an increase in the primary CS usage and a decrease in subsidiary CS induce cleavage efficiency and, subsequently, the *NQO1* reporter expression.

To confirm this finding, we then carried out an in vitro cleavage assay using 5′-capped *NQO1* 3′-UTR RNA encompassing the sequences required for cleavage and polyadenylation (~120 nucleotides upstream and downstream of the CS of the terminal PA-site) and HeLa nuclear extract as reported earlier[25,58]. Consistently, we observed the highest cleavage efficiency from the primary CS containing *NQO1* 3′-UTR RNA, whereas the cleavage from the subsidiary CS was negligible in vitro (Fig. 3D, Supplemental Fig. S3A). Akin to that in cellular assays, we also saw a gradual increase in the cleavage efficiency with a decreasing number of subsidiary CS (Fig. 3E, Supplemental Fig. S3B). Since cleavage is coupled with polyadenylation, we assessed the downstream polyadenylation efficiency using in vitro transcribed cleaved templates with each CS. We found the highest polyadenylation efficiency from the templates cleaved at the primary CS than the other subsidiary CS that showed negligible polyadenylation efficiency in vitro (Fig. 3F, Supplemental Fig. S3C). To confirm this, we then performed 3′-RACE to measure the 3′-end formation with wild-type and mutant *NQO1* reporter constructs (Fig. 3G–I). We saw the highest polyadenylation efficiency of the primary CS-driven *NQO1* reporter transcript with a gradual enhancement of polyadenylation with each subsidiary CS mutations (Fig. 3G–I). However, we did not see any significant difference in the length of the PA-tail addition on the primary and subsidiary CS-specific transcripts (Supplemental Fig. S3D). These results confirm that primary CS exhibits the highest cleavage and polyadenylation efficiency that drives gene expression.

2.6. The position of the CS downstream of the PA-signal, but not the sequence composition, determines the primary CS

Next, we investigated what determines the primary CS; we looked into two critical determinants: sequence composition and the nucleotide position from the PA-signal. Earlier studies established the predominant nucleotide at the CS (A > T > C ≫ G) and the dinucleotide (CA > AA > TA ≫ GA) in the order[50]. We first analysed the terminal nucleotide sequence of primary and subsidiary CS genome-wide (Fig. 3J). We observed a similar occurrence of the terminal nucleotide and the dinucleotide at the CS as shown earlier for general CS (terminal nucleotide: A > T > C > G, dinucleotide CA > AA > TA > GA) in both primary and subsidiary CS (Fig. 3J, Supplemental Fig. S3E). Therefore, sequence composition is unlikely to determine the primary versus subsidiary CS. Consistently, when we mutated the primary CS dinucleotide (-CA-) to other dinucleotide combinations (CA to -TA-, -GA-, and -AA-), there was no significant difference in the reporter NQO1 protein and mRNA expressions between the wild-type sequence and the mutant sequences of the primary CS (Fig. 3K and L). Likewise, there was no change in the 3′-end formation and the cleavage between the wild-type and mutant sequences in the reporter analysis (Supplemental Figs. S3F–G). We then analysed the position of occurrence of the primary CS and subsidiary CS for all the PA-sites genome-wide. We observed primary CS largely clustered between positions 12 to 18 nucleotides downstream of the PA-signal, whereas subsidiary CS did not show any specific clustering and were randomly distributed (Fig. 3M). In the majority of the PA-sites, the primary CS was located between 14 and 18 nucleotides (a median value of ~15.6 nucleotides) downstream of the PA-signal (Fig. 3M), showing a positional bias for the occurrence of primary CS relative to the PA-signal. To confirm this positional bias of the primary CS, we employed *NQO1* reporter construct with only the primary CS at the 3′-UTR. We introduced mutations to alter the position of the primary CS by moving the CS region (a sequence of 5 nucleotides, two nucleotides flanking the CS, CCACT where the CS is italicised) to a different position from the original position (16 nucleotides downstream of the PA-signal). We moved the CS sequence to 10 nucleotides upstream in one construct and 10 nucleotides downstream of the original position in another construct (Fig. 3N) (Supplemental Table S1). Surprisingly, we saw a

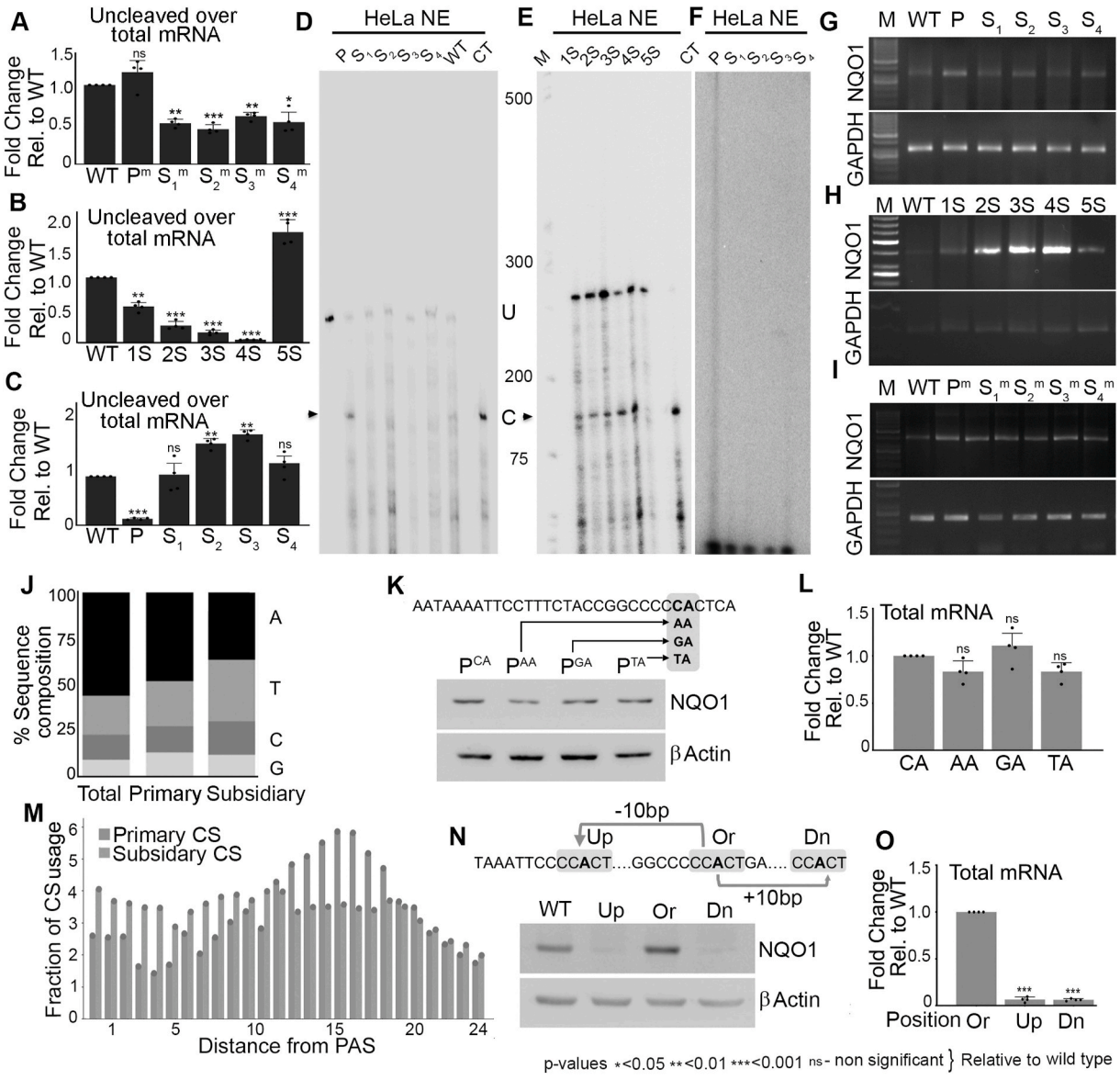


Fig. 3. Decrease in cleavage site heterogeneity and a concomitant increase in the cleavage at the primary CS induces gene expression. (*A-C*) Measurement of in vivo cleavage efficiency of *NQO1* reporter expression from HEK293 cells after the overexpression of wild-type and different CS mutants as in Fig. 2I–L, and O. Cleavage efficiency is measured by analysis of the uncleaved pre-mRNA level expressed over total mRNA level after normalising with internal control *GAPDH*. The error bar represents the standard error mean (SEM) of three independent experiments. (p values * <0.05, ** <0.01, *** <0.001, ns - non significant; p-values are relative to control). (*D*) In vitro cleavage assay of radiolabelled *NOQ1* 3′-UTR substrates (different mutants with primary or individual subsidiary sites as in Fig. 2M) with active HeLa nuclear extracts. The control-cleaved template (CT) is shown on the right. (U - uncleaved, C - cleaved fragment). (*E*) In vitro cleavage assay of radiolabelled *NOQ1* 3′-UTR substrates (wild-type and different combination mutants as in Fig. 2J) with active HeLa nuclear extracts. The control-cleaved template (CT) is shown on the right. (U - uncleaved, C - cleaved fragment). (*F*) In vitro polyadenylation assay of different *NQO1* cleaved templates (different mutants having either primary or individual subsidiary sites as in D) using active HeLa nuclear extract and α-^{32}P radiolabelled ATP. (*G-I*) 3′-RACE assays of *NQO1* reporter expression from HEK293 cells after overexpression of wild-type and different CS mutants as in A-C, respectively. (*J*) Percent sequence composition and frequency of occurrence of the terminal CS nucleotide (A, T, G, C) of primary and subsidiary CSs genome-wide. Four nucleotides are represented with four different colours. (*K*) Immunoblotting of the NQO1 reporter expression after overexpression of the CS sequence mutants (P^{CA}, P^{AA}, P^{GA}, P^{TA}) in HEK293 cells. (L) qRT-PCR analysis of *NQO1* reporter expression after overexpression of the CS sequence mutants (P^{CA}, P^{AA}, P^{GA}, P^{TA}) in the HEK293 cell. mRNA levels are expressed as fold-change relative to wild-type (P^{CA}). The error bar represents the standard error mean (SEM) of three independent experiments. (p values * <0.05, ** <0.01, *** <0.001, ns - non significant; p-values are relative to control). (*M*) Distance of the primary versus subsidiary CS occurrence relative to the PA-signal on each PA-site. The primary and subsidiary CSs are indicated. (*N*) Immunoblotting of the NQO1 reporter expression (wild-type and position mutations of the primary CS) and control β-Actin from HEK293 cells after overexpression of the respective mutants. The original position of the primary CS (Ori) was moved 10 nucleotides upstream (Up) or downstream (Dn) as indicated. Quantification of the blots with mutants relative to the original is shown in Supplemental Fig. 3F. (*O*) qRT-PCR analysis of *NQO1* reporter expression (with forward primer from the FLAG-epitope tag coding region and a reverse primer specific to *NQO1* CDS) from HEK293 cells after overexpression of wild-type and position mutations as in L. mRNA levels are expressed as fold-change relative to wild-type (original, Or). The error bar represents the SEM of three independent experiments. (p values * <0.05, ** <0.01, *** <0.001, ns - non significant; p-values are relative to control).

drastic reduction in the NQO1 reporter expression when the position of CS was changed to upstream and downstream positions relative to the original position (Fig. 3N and O, Supplemental Fig. S3H). There was a >5-fold difference between the reporter expression from the original primary CS position and the newly changed positions. Consistently, we saw an increased accumulation of uncleaved pre-mRNA and compromised 3′-end formation when the position of the primary CS was altered from its original position (Supplemental Figs. S3I–J). Together, these results demonstrate that the position of the CS relative to the PA-signal and not the sequence composition determines the primary CS of a PA-site.

2.7. *Reduced cleavage heterogeneity is directly associated with oxidative stress response gene expression*

To understand the biological relevance of CSH, we first performed enrichment analysis of genes that show CSH in different cellular pathways in silico (Metascape). Interestingly, we observed gene enrichment

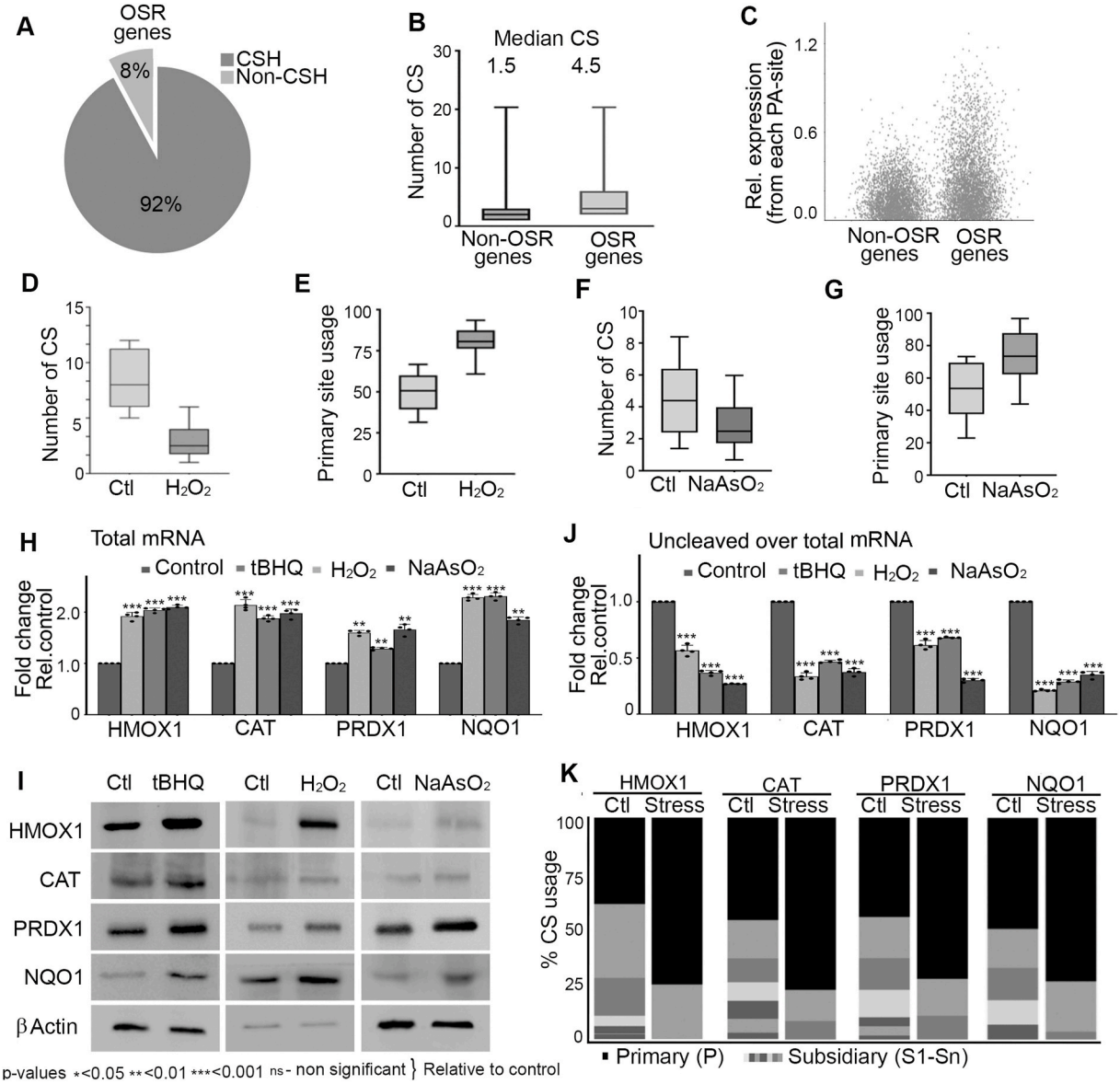


Fig. 4. Reduced cleavage heterogeneity regulates increased antioxidant gene expression during oxidative stress response (*A*) Pie chart showing fraction of mRNAs with CSH and non-CSH among genes encoding oxidative stress response proteins. (OSR - Oxidative stress response genes) (*B*) Box plot showing the distribution of number of CS clusters on respective PA-sites from oxidative stress-related (OSR) and random (non-oxidative stress-related, non-OSR) genes. Median CS on PA-sites of OSR and non-OSR genes is indicated. (*C*) Relative expression from PA-sites from OSR and non-OSR genes as in Fig. 2C. (*D-E*) Box plot showing a reduction in the CSH and increase in % usage of the primary CS of stress response genes in the presence of H_2O_2 induction. Data was analysed from earlier sequencing data (GSE138290)[60]. (*F-G*) Box plot showing reduction in the CSH and increase % usage of primary CS as in E-F but in the presence of $NaAsO_2$ treatment. Data was analysed from earlier sequencing data (GSE101851)[61]. (*H*) qRT-PCR analysis of stress response mRNAs *HMOX1, CAT, PRDX1,* and *NQO1* in HEK293 cells in the presence and absence of oxidative stress induction in HEK293 cells with different stressors (tBHQ, H_2O_2, $NaAsO_2$ as indicated). mRNA levels are expressed as fold-change relative to the control condition for each gene. (p values * <0.05, ** <0.01, *** <0.001, ns - non significant; p-values are relative to control). (*I*) Immunoblotting of the HMOX1, CAT, PRDX1, NQO1, and control β-Actin under the conditions as indicated. (*J*) Measurement of in vivo uncleaved pre-mRNA levels of stress response mRNAs from HEK293 cells expressed over total mRNA after normalisation with internal control GAPDH as in Fig. 2A. Cleavage efficiency is expressed as fold-change relative to control condition. (p values * <0.05, ** <0.01, *** <0.001, ns - non significant; p-values are relative to control). (*K*) Stacked column plot of % usage of primary versus subsidiary CS of stress response mRNAs *HMOX1, CAT, PRDX1, NQO1* in the presence and absence of oxidative stress induction (similar results were obtained for all stressors tBHQ, H_2O_2, $NaAsO_2$). Primary CS is indicated in black and different subsidiary CS are indicated in different shades of grey.

in pathways directly or indirectly involved in the cellular response to various stresses, including oxidative stress, ER stress, chemical-induced stress, mitochondrial stress, nutrient stress, unfolded protein response, etc. (Supplemental Fig. S4A) (Supplemental Dataset 2). Next, we checked important genes related to oxidative stress response (>800 genes detected in earlier studies) (Supplemental Table S2) [59]. Strikingly, >90 % of the PA-sites of these genes showed heterogeneity (Fig. 4A). The median number of CS of the oxidative stress response genes was 4.5 compared to the 1.5 of the randomly selected equivalent number of non-stress response genes (Fig. 4B). Eighteen key oxidative stress response genes, along with their number of CSs are shown in Supplemental Fig. S4B. Concomitantly, mRNA isoforms of stress response genes showed lower relative expression than that of non-stress response genes under normal conditions (Fig. 4C) consistent with the higher CSH. To understand the implication of CSH in oxidative stress response gene expression, we then analysed gene expression versus CSH after treatment with two stressors that induces antioxidant response - hydrogen peroxide (H_2O_2) and sodium arsenite ($NaAsO_2$) from earlier studies [60,61](Fig. 4D–G). There was a global induction of oxidative stress response genes under the two stress conditions. Consistently, we observed a global decrease in the CSH and an increase in the primary CS usage on both H_2O_2 and $NaAsO_2$ treatment (Fig. 4D–G). Under H_2O_2 treatment, the median number of CS was reduced from 6.8 to 1.5. In contrast, the primary CS usage was increased from 50 % average usage to >80 % average usage after treatment with H_2O_2 (Fig. 4D and E). Likewise, under $NaAsO_2$ treatment, there was a similar decrease in the CSH from 4.5 median CS to 2.2 median CS, whereas the primary site usage was increased from 55 % to 75 % (Fig. 4F and G), indicating that CSH and primary CS usage is directly linked with the oxidative stress response gene expression. Concomitantly, when we look at the individual stress response genes (*PRDX1, GSTCD, HMOX1*, and *ALDH3A2*), we see an equal reduction in the CSH and an increase in the primary CS usage on stress ($NaAsO_2$) treatment (Supplemental Fig. S4C). Together, these results reveal that induction of antioxidant gene expression is associated with decreased CSH and increased primary CS usage.

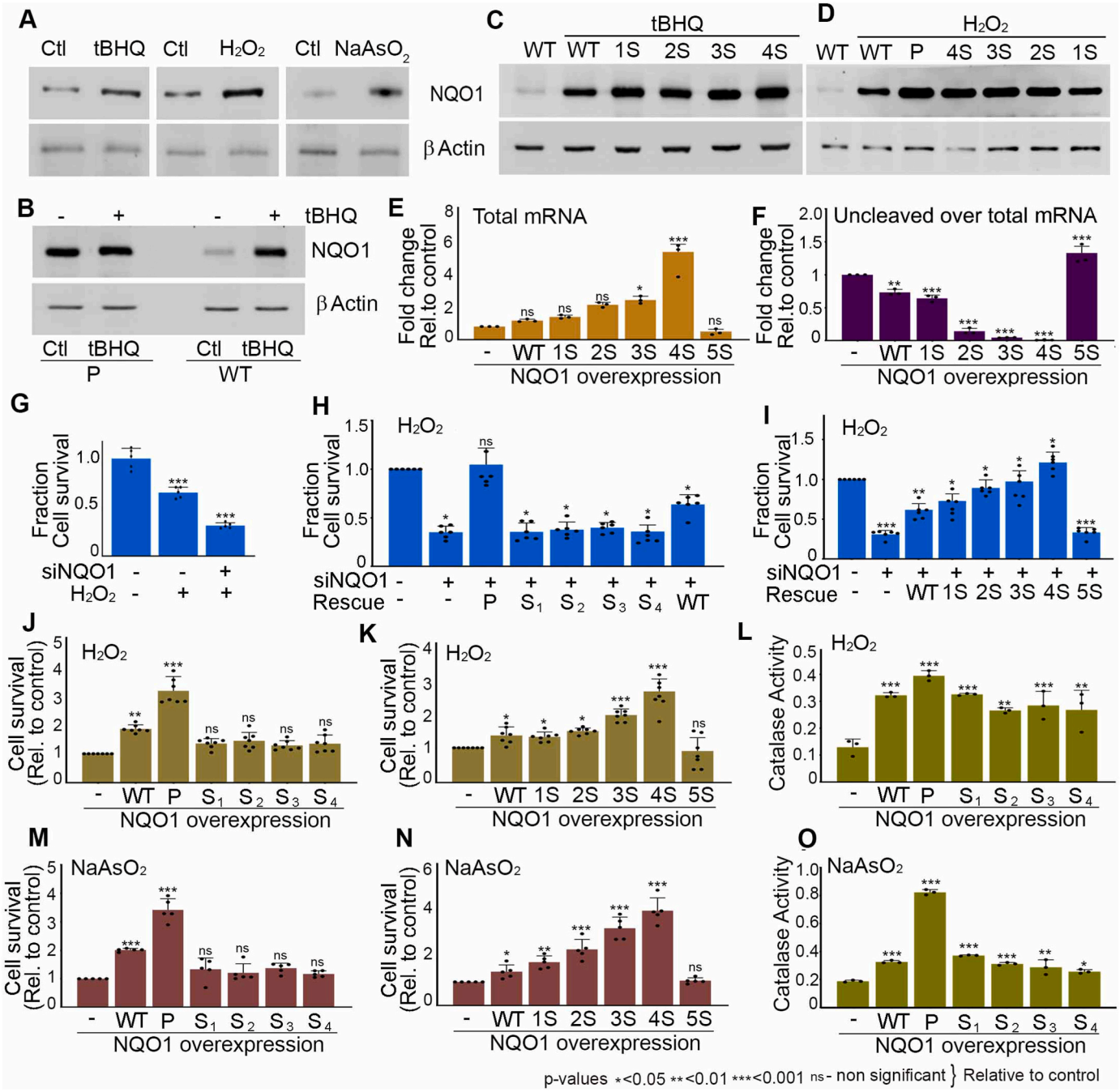
2.8. Cleavage heterogeneity regulates the expression of antioxidant response proteins under oxidative stress

Next, to further assess the role of CSH in antioxidant gene expression, we employed three stressors: H_2O_2, $NaAsO_2$, and the oxidative stress agonist *tert*-butyl hydroquinone (tBHQ), which induced antioxidant response in HEK293 cells. We then analysed the expression and CS usage of four key oxidative stress response genes (*HMOX1*, *CAT*, *PRDX1*, and *NQO1*). We consistently observed induced protein and mRNA expressions of the four-stress response regulator after treatment with the three stressors (Fig. 4H and I). Analysis of uncleaved pre-mRNA over total mRNA also showed increased cleavage efficiency of the selected four stress response pre-mRNAs on treatment with the three different stressors (Fig. 4J). In agreement with this, we also saw a reduced CSH of the four genes on stress induction (Fig. 4K). The number of CS was 6 for *HMOX1*, 8 for *CAT* and *PRDX1*, and 5 for *NQO1,* respectively, under the normal conditions that was reduced to 2 CS for *HMOX1* and *NQO1*, and 3 CS for *CAT* and *PRDX1* under all the three stress conditions (Fig. 4K). Concomitantly, there was also an increase in the primary CS usage from ~45 to 50 % under normal conditions to >70 % under all stress conditions (Fig. 4K, Supplemental Fig. S5A), confirming the association of CSH with antioxidant gene expression.

Next, we used *NQO1* reporter mini gene construct in the presence and absence of three stressors (H_2O_2, $NaAsO_2$, and tBHQ). First, we validated the induced *NQO1* reporter expression by Western blot and qRT-PCR analysis along with in vivo cleavage after treatment with the three stressors (Fig. 5A, Supplemental Figs. S5B–C). All three stressors showed stimulations of *NQO1* reporter expression and cleavage efficiency. Moreover, we saw a similar reduction in the CSH and increased primary CS usage on treatment with the three stressors (Supplemental Fig. S5D). To confirm the role of CSH in the stress response gene expression, we employed the wild-type reporter construct that shows CSH and the mutant construct that lacks heterogeneity (having only the primary CS and other subsidiary sites mutated). Strikingly, we observed no induction of the reporter NQO1 expression from the mutant construct that lacks CSH, whereas there was an induced expression from the wild-type construct on stress treatment (Fig. 5B). Similarly, no induction of NQO1 expression was seen from the reporter construct that has only one individual subsidiary CS on stress treatment in Western and qRT-PCR analysis (Supplemental Figs. S5E and F) indicating that CSH is involved in the antioxidant gene expression. Consistently, expression from the primary CS was highest compared to the subsidiary CSs under all three stress conditions (Supplemental Fig. S5G). To further assess the role of CSH, we checked the reporter expression using *NQO1* constructs with an increasing number of subsidiary CS mutations. We observed a gradual increase in the induced NQO1 protein expression with increasing subsidiary site mutations on stress treatment (Fig. 5C and D, Supplemental Figs. S5H–I). Likewise, there was an increase in the mRNA level and cleavage efficiency from the reporter expression with the subsidiary CS mutations (Fig. 5E and F). However, the highest difference in the expression level between control and stress treatment was seen from the wild-type reporter construct (having four subsidiary CSs), and least when all subsidiary CSs were mutated (Supplemental Figs. S5J–K). Together, these results showed that stress response gene expression involves a reduction in the CSH and an increase in primary CS usage. And that when only primary CS is present, the expression becomes constitutive.

2.9. Cleavage site heterogeneity versus primary cleavage site usage regulates cellular tolerance to oxidative stress induction

Since NQO1 is a crucial antioxidant response protein, we tested the effect of *NQO1* expression from reduced CSH in cellular oxidative stress tolerance in HEK293 cells using two stressors, H_2O_2 and $NaAsO_2$. We first knocked down *NQO1* in HEK293 cells, then treated cells with H_2O_2, and rescued them with wild-type and *NQO1* CS mutant constructs that are insensitive to the *NQO1* siRNA employed. Treatment with H_2O_2 resulted in a significant cellular sensitivity with around 40 % cell death in control cells in the MTT assay (Fig. 5G). *NQO1* knockdown resulted in further sensitivity with an overall >60 % cell death compared to control cells on the H_2O_2 treatment (Fig. 5G). Intriguingly, *NQO1* expression from the primary CS but not the other subsidiary CS rescued the cell death from the knockdown of *NQO1* on H_2O_2 treatment (Fig. 5H). Expression from the wild-type *NQO1* construct also increased the stress tolerance, yet the highest tolerance to cellular stress was obtained from the primary CS (Fig. 5H). Similarly, the rescue of the H_2O_2 sensitivity gradually increased with an increasing number of subsidiary CS mutations on stress treatment (Fig. 5I) in agreement with the increased *NQO1* expression. Further, we overexpressed *NQO1* from different ectopic constructs with individual and combination of subsidiary CS mutations and assessed the cellular tolerance to H_2O_2 treatment. Akin to earlier results, ectopic overexpression of *NQO1* from the wild-type construct increased cellular tolerance to H_2O_2 stress with the highest tolerance from *NQO1* expression from the primary CS. NQO1 expression from subsidiary CS did not affect the tolerance to H_2O_2 stress (Fig. 5J). We also saw a gradual increase in the cellular H_2O_2 tolerance with more subsidiary CS mutations (Fig. 5K). Similar cell sensitivity was also scored on trypan blue staining after H_2O_2 treatment in the presence and absence of ectopic *NQO1* expression from wild-type and mutant CS constructs. There was a similar higher cell survival from *NQO1* expression from the primary CS, and a gradual increase in the stress tolerance with decreasing CSH (Supplemental Figs. S6A–B). Moreover, the assessment of catalase activity after H_2O_2 treatment also showed higher catalase activity on *NQO1* expression from the primary CS having no CSH compared to the control cells (Fig. 5L), revealing the role of CSH in the cellular antioxidant response.



(caption on next page)

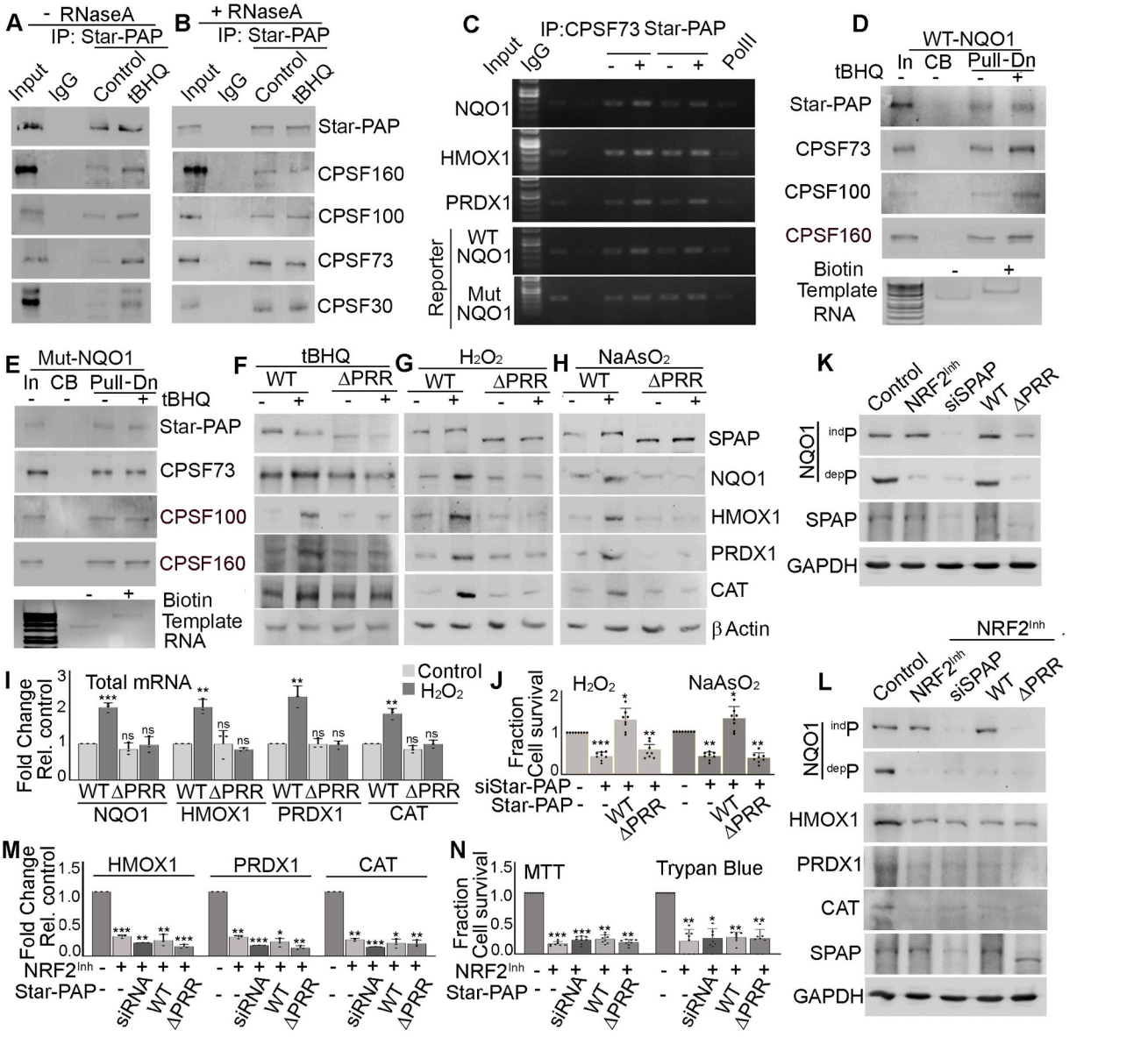
Fig. 5. Cleavage heterogeneity regulates cellular oxidative stress response (*A*) Immunoblotting of the wild-type FLAG-NQO1 reporter expression and control β-Actin in HEK293 cell after induction of oxidative stress using tBHQ, H_2O_2, and $NaAsO_2$. (*B*) Immunoblotting of wild-type and mutant *NQO1* reporter (having only the primary CS) expression and control β-Actin in HEK293 cell after treatment with tBHQ. (*C-D*) Immunoblotting of the NQO1 reporter expression and control β-Actin after overexpression of reporter constructs with increasing subsidiary CS mutations as indicated in HEK293 cell in the presence of induction of oxidative stress using tBHQ and H_2O_2 respectively. A blot of untreated control is also indicated. (*E*) qRT-PCR analysis of induced *NQO1* reporter expression relative to uninduced expression from HEK293 cells with WT and increasing subsidiary site mutations (1S to 4S) upon H_2O_2 treatment. mRNA levels are expressed as fold-change relative to control untransfected cells (−). The error bar represents the SEM of n = 3 independent experiments. (p values * <0.05, ** <0.01, *** <0.001, ns - non significant; p-values are relative to control). (*F*) In vivo cleavage assay of the *NQO1* reporter expression from HEK293 cells as in E. Cleavage efficiency is expressed as uncleaved pre-mRNA level over total mRNA and expressed as fold change relative to control untransfected cells (−) as in E. Error bar represents standard error mean (SEM) of n = 3 independent experiments. (p values * <0.05, ** <0.01, *** <0.001, ns - non significant; p-values are relative to control). (*G*) MTT assays to assess cell viability in HEK293 cells after H_2O_2 treatment in the presence and absence of knockdown of *NQO1*. Cell survival is expressed as fraction relative to control cells. We have used control scrambled siRNA wherever (−) siRNA is indicated. (*H*) MTT assays as in G after the knockdown of *NQO1* and rescue with wild-type and different *NQO1* CS mutants that have only primary or individual subsidiary sites, as indicated. Cell survival is expressed as fraction survival relative to control cells. We have used control scrambled siRNA wherever (−) siRNA is indicated. (*I*) MTT assays as in H but in rescue with wild-type and *NQO1* CS mutants with increasing subsidiary site mutations as indicated. (*J*) MTT assay in HEK293 cells after overexpression of wild-type and different *NQO1* CS mutants having only primary or individual subsidiary sites as indicated after induction of oxidative stress by H_2O_2. Cell survival is expressed as fold difference relative to control untransfected cells (−). (*K*) MTT assay in HEK293 cells after overexpression of wild-type and different *NQO1* CS mutants with increasing subsidiary CS mutations as indicated after induction of oxidative stress by H_2O_2. Cell survival is expressed as fold-difference relative to control untransfected cells (−). (*L*) Catalase assay from HEK293 cells under the similar conditions as in J. Activity is expressed as fold-difference relative to control untransfected cells (−). (*M-N*) MTT assay as in J-K respectively but after induction of oxidative stress with $NaAsO_2$ treatment. Cell survival is expressed as fold-change relative to control untransfected cells (−). (*O*) Catalase assay as in L after induction of oxidative stress with $NaAsO_2$. For all MTT and catalase assays, the error bar represents the standard error mean (SEM) of n = 3 independent experiments (p values * <0.05, ** <0.01, *** <0.001, ns - non significant; p-values are relative to control).

Concomitantly, stress tolerance was tested after $NaAsO_2$ treatment in the presence of ectopic *NQO1* expression from individual CS and combination CS mutations that showed a gradual decrease in the CSH. We found a similar increase in the stress tolerance with *NQO1* expression from the primary CS-driven construct than the other subsidiary CS-driven constructs in the MTT assay (Fig. 5M). There was a >1.8-fold increase in cell survival in the MTT assay from *NQO1* expression from the primary CS relative to the control cells. Moreover, there was a gradual increase in the $NaAsO_2$ stress tolerance with an increasing number of mutations of the subsidiary CS (Fig. 5N), as seen in H_2O_2 treatment. Similar results were obtained in the trypan blue staining after $NaAsO_2$ treatment, as in the case of the MTT assay. There was a >2.5-fold increase in the cell survival on *NQO1* expression compared to the control cells with no ectopic expression. The tolerance was highest with the NQO1 expression from the primary CS compared to other individual subsidiary CSs (Supplemental Fig. S6C). Moreover, there was also a gradual increase in the $NaAsO_2$ tolerance with NQO1 expression with decreasing CSH (Supplemental Fig. S6D). Furthermore, we also measured the catalase activity after $NaAsO_2$ treatment, and we observed the highest catalase activity in the presence of NQO1 expression from the primary CS compared to the wild-type or other subsidiary CSs (Fig. 5O). Together, these results demonstrate a correlation between CSH and cellular tolerance to oxidative stress. The antioxidant response involves a decrease in the CSH and an increase in the primary CS usage.

2.10. Increased affinity of cleavage complex assembly stimulates fidelity of cleavage and reduces cleavage heterogeneity to regulate anti-oxidant gene expression

Earlier studies showed that oxidative stress response gene expression is regulated by the non-canonical PAP, Star-PAP, and not by the canonical PAPα and that Star-PAP activity is induced during oxidative stress through phosphorylation [26,62]. We also confirmed the expression of critical oxidative stress response regulators (*NQO1*, *HMOX1*, *PRDX1*, *CAT*) by qRT-PCR and Western blot. We observed loss of both protein and mRNA expressions on Star-PAP knockdown but not on PAPα knockdown (Supplemental Figs. S6E–F). To understand the mechanism of CSH and how CSH is reduced and the primary CS usage is stimulated during oxidative stress, we analysed the Star-PAP cleavage complex assembly in the presence and absence of oxidative stress agonist, tBHQ treatment. We first immunoprecipitated (IP'ed) Star-PAP and examined the association of crucial CPA components (CPSF subunits 73, 30, 100, and 160). Intriguingly, we observed an induced association of Star-PAP with CPSF30, CPSF73, CPSF100, and CPSF160 suggesting an increased affinity of Star-PAP CPA complex under oxidative stress (Fig. 6A). However, the assembly of the 3′-end processing complex occurs on the target RNAs where CPA components are associated. Moreover, Star-PAP is known to bind target mRNAs and help recruit cleavage components CPSF73 and CPSF160 through direct interaction to assemble a stable CPA complex [25,63]. Therefore, to verify the involvement of RNA in the induced association of Star-PAP CPA complex, we carried out IP of Star-PAP after RNaseA digestion of the cellular lysates from the tBHQ-treated and -untreated cells. Intriguingly, there was no longer induction in the association of the CPA components (CPSF30, CPSF73, CPSF100, and CPSF160) with Star-PAP in the stress treated cell lysates after RNaseA digestion (Fig. 6B). This indicates stimulation of Star-PAP CPA complex assembly on the target mRNA and not merely protein-protein interactions of the CPA components during stress response. Further, the assembly on target mRNA 3′-UTR was assessed by RNA immunoprecipitation (RIP) of Star-PAP and another key CPA component CPSF73. We found a heightened association of Star-PAP and CPSF73 on mRNAs encoding key antioxidant response protein (*NQO1, HMOX1, PRDX1*) upon oxidative stress induction (Fig. 6C). There was 3- to 5-fold increase in the association of Star-PAP and CPSF73 on the target mRNAs on tBHQ treatment (Supplemental Figs. S6G–H) revealing that oxidative stress induces the affinity of the Star-PAP cleavage complex assembly on mRNA target 3′-UTRs. To confirm this, we prepared a biotinylated *NQO1* 3′-UTR RNA template by in vitro transcription. We then incubated the biotinylated 3′-UTR RNA with active nuclear extract and pulled down the associated proteins on the biotin-*NQO1*-3′-UTR RNA using streptavidin beads followed by analysis of the associated CPA components by Western blotting. We consistently observed an increased association of CPSF160, CPSF100, CPSF73 and Star-PAP with the *NQO1* 3′-UTR RNA on tBHQ treatment (Fig. 6D). Finally, Star-PAP complex assembly on target RNA 3′-UTR was directly assessed by an in vitro RNA-EMSA experiment using FLAG-purified Star-PAP complex and florescent labelled *NQO1* 3′-UTR RNA. We saw a mobility shift of the *NQO1* 3′-UTR RNA in the presence of FLAG-purified Star-PAP complex that was markedly induced with Star-PAP complex from tBHQ treated cells compared to the control untreated cells (Supplemental Fig. S6I). Together, these results demonstrate that Star-PAP-mediated CPA complex assembly is stimulated by oxidative stress treatment of the cell.

Furthermore, to understand the link between the increased affinity of CPA complex assembly with CSH, we then employed the NQO1 reporter construct with the wild-type and CSH mutant (having only the primary CS that does not show CSH) and analysed the association of CPSF73 and Star-PAP by RIP experiment. Consistently, we saw an increased association of Star-PAP and CPSF73 with the *NQO1* with the wild-type 3′-UTR



(caption on next page)

Fig. 6. Increased affinity of cleavage complex assembly on oxidative stress stimulates fidelity of cleavage at the primary site, reducing the CSH (*A*) Immunoprecipitation of Star-PAP from HEK293 cell lysate in the presence and absence of oxidative stress induction by tBHQ treatment followed by Western blot analysis of Star-PAP, CPSF30, CPSF73, CPSF100 and CPSF160 as indicated. Input is equivalent to 10 % of the total protein of the IP lysate used. (*B*) Immunoprecipitation of Star-PAP HEK293 cell lysate as in A but after digestion of the cell lysate with RNaseA. Input is equivalent to 10 % of the total protein of the IP lysate used. (*C*) RNA immunoprecipitation (RIP) analysis of Star-PAP and CPSF73 along with control RNA pol II (PolII) from HEK293 cells in the presence and absence of tBHQ treatment and the analysis of association with oxidative stress response genes (*NQO1, HMOX1, PRDX1*); and also association with reporter WT and mutant *NQO1* RNA as indicated. Input is equivalent to 10 % of the total protein of the IP lysate used. (*D*) Streptavidin pulldown (Pull-Dn) of biotin labelled wild-type *NQO1* from cell extracts followed by Western analysis of Star-PAP, CPSF73, CPSF100, and CPSF160 in the presence and absence of tBHQ induction. Biotinylated and non-biotinylated RNA templates are shown below. Input (In) is equivalent to 10 % of the total protein of the pulldown cell lysate used. CB - control pulldown of the streptavidin beads from the cell lysates. (*E*) Streptavidin pulldown (Pull-Dn) of biotin labelled mutant *NQO1* from cell extracts followed by Western analysis of Star-PAP, CPSF160, CPSF100, and CPSF73 in the presence and absence of tBHQ induction. Biotinylated and non-biotinylated RNA templates are shown below. Input (In) is equivalent to 10 % of the total protein of the pulldown cell lysate used. CB - control pulldown of the streptavidin beads from the cell lysates. (*F–H*) Immunoblotting analysis of oxidative stress response proteins (NQO1, HMOX1, PRDX1, CAT) from HEK293 cells after Star-PAP knockdown and rescue with wild-type (WT) and PRR deletion (ΔPRR) Star-PAP in the presence and absence of tBHQ, H_2O_2, and $NaAsO_2$ treatment as indicated. (*I*) qRT-PCR analysis of *NQO1, HMOX1, PRDXI* and *CAT* from HEK293 cells in the presence and absence of H_2O_2 treatment after Star-PAP knockdown and rescue with wild type (WT) and PRR deletion (ΔPRR) Star-PAP. The error bar represents the standard error mean (SEM) of n3 independent experiments. (p values * <0.05, ** <0.01, *** <0.001, ns - non significant; p-values are relative to control). (*J*) MTT assay in HEK293 cells after Star-PAP knockdown and rescue with wild type and PRR deletion (ΔPRR) Star-PAP in the presence of induction of oxidative stress by H_2O_2 and $NaAsO_2$ as indicated. Cell survival is expressed as fold difference relative to control cells. (*K*) Western blot analysis of reporter FLAG-NQO1 expression (NQO1) from a reporter *NQO1* mini-gene construct with an NRF2-independent promoter and the *NQO1* 3′-UTR (^{ind}P); and another with endogenous NRF2-dependent *NQO1* promoter and *NQO1* 3′-UTR (^{dep}P) in the presence and absence of inhibition of NRF2 (NRF2^{Inh}), Star-PAP depletion (siSPAP), and wild type (WT) and CSH-mutant Star-PAP (ΔPRR) expression in the HEK293 cells as indicated. Control Star-PAP and GAPDH blots are shown below. (L) Western blot analysis of reporter FLAG-NQO1 (NQO1) expression as in K (top 2 panels) from NRF2-independent promoter and the *NQO1* 3′-UTR, and endogenous NRF2-dependent promoter and *NQO1* 3′-UTR (^{dep}P) in the presence and absence of NRF2 inhibition (NRF2^{Inh}), and Star-PAP depletion (siSPAP), and wild type (WT) and CSH-mutant Star-PAP (ΔPRR) expression as indicated. Bottom panels show Western blot analysis of anti-oxidant response proteins (HMOX1, PRDX1, and CAT) from HEK293 cells in the presence and absence of NRF2 inhibition (NRF2^{Inh}), Star-PAP depletion (siSPAP), or wild type (WT) and CSH-mutant Star-PAP (ΔPRR) expression after NRF2 inhibition (NRF2^{Inh}). Control Star-PAP and GAPDH blots are shown. (*M*) qRT-PCR analysis of the *HMOX1, PRDXI* and *CAT* from HEK293 cells under the conditions as in L in the presence and absence of NRF2 inhibition (NRF2^{Inh}), Star-PAP depletion (siSPAP), or wild type (WT) and CSH-mutant Star-PAP (ΔPRR) expression after NRF2 inhibition (NRF2^{Inh}). The error bar represents the standard error mean (SEM) of n = 3 independent experiments. (p values * <0.05, ** <0.01, *** <0.001, ns - non significant; p-values are relative to control). (N) MTT assay and trypan blue staining of HEK293 cells in the presence and absence of NRF2 inhibition (NRF2^{Inh}), Star-PAP depletion (siSPAP), or wild type (WT) and CSH-mutant Star-PAP (ΔPRR) expression after NRF2 inhibition (NRF2^{Inh}). Cell survival is expressed as fold difference relative to control cells.

(that shows CSH) on tBHQ treatment (Fig. 6C, Supplemental Figs. S6G–H). However, in the presence of CSH mutant, there was no further induction in the association of *NQO1* 3′-UTR with CPSF73 and/or Star-PAP in the reporter RNA, indicating that CSH is linked with the increased affinity of the CPA component assembly with the target mRNA. To further evaluate this, we employed a biotinylated *NQO1* 3′-UTR RNA template with the CSH mutant 3′-UTR (having only primary CS, other subsidiary sites were mutated) and analysed the associated CPA factors after the pulldown. Unlike the wild-type RNA, there was no further increase in the association of CPSF73, CPSF100, CPSF30 or Star-PAP with the mutant *NQO1*-UTR on tBHQ treatment (Fig. 6E) confirming a direct link between CPA complex assembly and the CSH. Together, these results suggest that increased affinity of the CPA complex assembly increases the fidelity of cleavage at the primary CS, thus reducing CSH. To test this, we first looked for Star-PAP deletion mutations that will compromise the stimulated association of the CPSF components with Star-PAP. We employed deletion mutations of several Star-PAP domains that were earlier linked with oxidative stress covering most of the Star-PAP CDS region [ZF-RRM deletion (ΔZF-RRM), proline-rich region in the PAP domain deletion (ΔPRR), and C-terminal deletion (ΔCTD)]. Strikingly, we observed a loss of induced association of CPSF73, CPSF30, and CPSF100 with Star-PAP on tBHQ treatment on PRR deletion (Supplemental Fig. S6J). Star-PAP PRR region and its phosphorylation have established roles in the oxidative stress response [26,62]. Our data indicate that Star-PAP ΔPRR is compromised for induced CPA complex assembly under oxidative stress conditions. Assessment of the CSH of reporter *NQO1* 3′-UTR showed no change in the CSH in the presence of ΔPRR mutation. Contrary to this, the wild-type Star-PAP showed reduction in the CSH during oxidative stress response (Supplemental Fig. S6K). There was ~60 % decrease in the heterogeneity in the presence of the wild-type Star-PAP on stress induction, there was no significant reduction in the CSH in the presence of ΔPRR mutant Star-PAP. Moreover, while there is a ~40 % increase in the reporter *NQO1* 3′-UTR primary CS usage in the presence of wild-type Star-PAP, there was no increase in the primary usage in the presence of ΔPRR mutant Star-PAP (Supplemental Fig. S6J). Therefore, Star-PAP ΔPRR mutant acts as the CSH-mutant Star-PAP that is compromised for CSH and CSH-mediated gene control.

We then tested the oxidative stress response gene expression with the wild-type and CSH-mutant Star-PAP (ΔPRR Star-PAP) transfection in HEK293 cells after endogenous depletion of Star-PAP. Strikingly, ΔPRR Star-PAP failed to show induced expression of oxidative stress response proteins (NQO1, HMOX1, PRDX1, and CAT) and the respective mRNA expressions on tBHQ treatment. In contrast, wild-type expression showed induced levels on stress treatment (Fig. 6F, Supplemental Fig. 7A). Together, these results reveal that oxidative stress induction stimulates the affinity of CPA assembly thereby increasing the fidelity of cleavage at the primary CS and reducing the CSH upon oxidative stress treatment. Therefore, we then assessed the effect of CSH-mutant Star-PAP on the anti-oxidant response gene expression and cellular sensitivity to two different stresses (H_2O_2 and $NaAsO_2$). In Western analysis, we observed no induction in the expression levels of anti-oxidant proteins (HMOX1, NQO1, PRDX1, and CAT) upon H_2O_2 and $NaAsO_2$ treatment as opposed to the wild type Star-PAP (Fig. 6G and H). qRT-PCR analysis also showed similar results that CSH-mutant Star-PAP failed to induce the mRNA levels upon stress treatment (Fig. 6I). Concomitantly, in the presence of CSH-mutant Star-PAP, we observed increased cellular sensitivity (>2-fold cell death) compared to wild type under stress conditions (H_2O_2 and $NaAsO_2$) relative to untreated cells in both MTT assay and trypan blue staining experiments (Fig. 6J, Supplemental Fig. 7B). Together, these results establish the involvement of CSH-mediated gene control in the stimulation of anti-oxidant response protein expression under stress.

2.11. CSH-mediated anti-oxidant response mechanism at the 3′-end is distinct and operates in concert with the NRF2-mediated transcriptional control at the 5′-end

Cleavage and polyadenylation at the 3′-end endows mRNA stability, controls turnover, and is essential for efficient translation[2,29]. This processing step at the 3′-end is closely coupled with transcription from the promoter at the 5′-end to prevent a futile cycle of transcriptional

stimulation that is countered by turnover/destabilisation at the 3′-end [41,64]. Stress response genes such as *NQO1*, *HMOX1*, *PRDX1* and *CAT* are transcriptionally regulated at the 5′-end by NRF2, a master transcriptional regulator of anti-oxidant response [65,66]. A coordinated regulation at both 5′-and 3′-ends are required to obtain inductions in these proteins under stress. Therefore, we analysed the interplay between the CSH-mediated mRNA 3′-end processing and NRF2-mediated transcriptional mechanism in the induction of stress response gene expression. First, to assess a potential NRF2-mediated transcriptional effect on the CSH-controlled stimulation of stress response gene expression, we employed a reporter *NQO1* mini-gene construct driven from an NRF2-independent promoter (a constitutive CMV promoter) at the 5′-end along with the *NQO1* 3′-UTR at the 3′-end and compared with expression from wild type endogenous NRF2-dependent *NQO1* promoter and *NQO1* 3′-UTR. We observed an induction in the NQO1 expression from the reporter construct driven from NRF2 independent promoter and the *NQO1* 3′-UTR on stress treatment (Fig. 5A, Supplemental Fig. 5B) indicating a non-transcriptional effect of CSH-mediated mechanism. Further, inhibition of NRF2 did not have any effect on the induced NQO1 reporter expression from the same reporter construct, whereas the induced expression was diminished on Star-PAP knockdown or in the presence of CSH-mutant Star-PAP that compromises CSH (Fig. 6K, Supplemental Fig. 7C). On the other hand, NQO1 expression from the endogenous NRF2-dependent promoter and *NQO1* 3′-UTR was lost on NRF2 inhibition, similar to that of siRNA depletion of Star-PAP or in the presence of CSH-mutant Star-PAP (Fig. 6K, Supplemental Figs. 7C–D). These results reveal that the CSH-mediated induction of antioxidant response gene expression is distinct and is independent of the NRF2-mediated transcriptional pathway. However, in the presence of NRF2 inhibition, there was no significant effect of Star-PAP depletion or of CSH-mutant Star-PAP on the diminished NQO1 expression from endogenous NRF2-dependent promoter (Fig. 6L, Supplemental Figs. 7C–D). Similarly, compromised inductions of stress response gene expression (*HMOX1, CAT, PRDX1*) after NRF2 depletion was not markedly affected by Star-PAP depletion or with CSH-mutant Star-PAP on stress treatment (Fig. 6L and M). Consistently, the increased cellular sensitivity to oxidative stress on NRF2 inhibition was not affected significantly by Star-PAP depletion or CSH-mutant Star-PAP expression (Fig. 6N). The loss of expression of antioxidant genes and cellular stress sensitivity on NRF2 inhibition was further confirmed by siRNA knockdown of NRF2 in the presence and absence of siRNA depletion of Star-PAP, or wild type and CSH-mutant Star-PAP expression (Supplemental Figs. 7E–H). These results reveal that CSH-mediated 3′-UTR regulation is downstream of the NRF2-mediated transcriptional mRNA synthesis. Together, our data establish that the CSH-control at the 3′-end cooperates with the NRF2-transcriptional control at the 5′-end to yield an overall induced anti-oxidant protein expression in the cell.

3. Discussion

3′-end polyadenylation involves two distinct yet coupled steps - endonucleolytic cleavage and polyadenylation [67–69]. Cleavage reaction influences the efficiency of mature mRNA template generation for the downstream polyadenylation. Consequentially, mutation in the CS compromises the overall 3′-end processing reaction, and such mutations are known in human genetic diseases[54,56]. Yet, cleavage reaction is considered stochastic, occurring downstream of the PA-signal, that can result in heterogeneity in CS[50,52,53]. Contrary to this, we demonstrated that cleavage reaction is tightly regulated through CSH to control gene expression and that the heterogeneity is not a result of imprecision or random cleavage events. In the CSH, cleavage occurs predominantly at a primary CS, followed by cleavages at different subsidiary sites. We demonstrated an inverse relationship between the number of CS and the gene expression that regulates cellular antioxidant gene expression. During the antioxidant response, there is a decrease in the cleavage at the subsidiary CS. At the same time, there is an increase in the usage of the primary CS that stimulates the expression of antioxidant response proteins. This mechanism is shown in Fig. 7. This mechanism accounts for almost all mRNAs related to stress response gene expression, inducing cleavage and polyadenylation efficiency. Therefore, it is likely that the imprecision of cleavage reaction occurs typically in the cell and that the precision is improved under oxidative stress, leading to reduced CSH and induced gene expression. Interestingly, while the CSH-mediated stress response mechanism is distinct, it operates in concert with the NRF2-mediated transcriptional induction from the promoter. Controlling 3′-end processing is an efficient way to regulate mRNA stability and translation [1,30]. Stabilising mRNA at the 3′-end will prevent a futile cycle of transcriptional stimulation and degradation [29,64]. In addition, stability of an mRNA determines the steady state level affecting its cellular availability for translation [70,71]. Moreover, processing at the 3′-end has distinct roles in the translation efficiency of an mRNA [30,31]. Thus, CSH-mediated stress response pathway will function downstream but cooperatively with the NRF2-mediated transcriptional pathway to ultimately generate an increase and rapid production of stress response proteins.

Star-PAP is the primary PAP involved in regulating oxidative stress response gene expression[26,62]. The antioxidant response dramatically stimulates Star-PAP polyadenylation activity[26]. Therefore, the increase in the polyadenylation activity will complement the increased cleavage efficiency, thus yielding a stimulated and rapid production of mature mRNAs that encode antioxidant response proteins. Moreover, Star-PAP has a direct role in the cleavage step that recruits the endonuclease CPSF73 in addition to acting as a structural protein to assemble the cleavage complex on target mRNAs [25,63]. Our reporter analysis reveals that the mechanism of CSH is distinct from other transcriptional mechanism of stress response gene expression[72,73]. And, CSH-mediated gene regulation will have ramifications in the disease pathogenesis, such as cardiovascular, cancer, inflammation, pathogenesis, neurodegeneration, aging, or diabetes, where the antioxidant response is critical[74–78].

Next is the mechanism of CSH and its regulation. We have answered this question regarding oxidative stress response gene expression. Antioxidant response involves both a reduction in the CSH and an increase in the primary CS usage. We have shown that the affinity of the CPA complex is induced under oxidative stress, which increases the strength of the CPA complex assembly. We propose that an increase in the CPA complex assembly's strength induces cleavage fidelity at the primary CS, thus reducing the suboptimal heterogenous cleavages. Our study directly links the affinity of CPA complex assembly with CSH. It shows tight regulation of CSH where cleavage occurs mainly at the primary CS, followed by several random suboptimal cleavages at nearby sites. Earlier studies had shown lower cleavage efficiency with compromised *cis*-elements at the 3′-UTR [79,80]. In light of our new finding, it is likely because of the lower affinity/strength of CPA assembly, which results in increased CSH and reduced cleavage efficiency. Moreover, a recent study also suggested a potential relationship between the strength of the CPA complex assembly and CSH, where a decrease in the strength of cleavage on mRNAs with overly distant PA-signal can induce CSH[53]. Additionally, emerging evidence indicates a direct association of APA with CSH and gene expression[53]. Further analysis is required to shed light on the aspect of APA with CSH and its cellular implications.

Furthermore, the signals for regulating CSH during oxidative stress response relay through Star-PAP in the CPA complex. Yet, the mechanism of how PAP regulates cleavage reaction is still unclear. In vitro studies have shown the requirement of PAP in the fractionated CPA components to reconstitute cleavage reaction [81–83]. In the case of the Star-PAP-mediated cleavage complex, Star-PAP directly binds the target mRNA and helps recruit CPSF components, including CPSF73[25,63]. Nevertheless, cleavage efficiency was affected by the addition of PAP in the in vitro cleavage reaction [81]. Moreover, addition of a PAP activity inhibitor (cordycepin) in the in vitro reconstitution experiment reduces

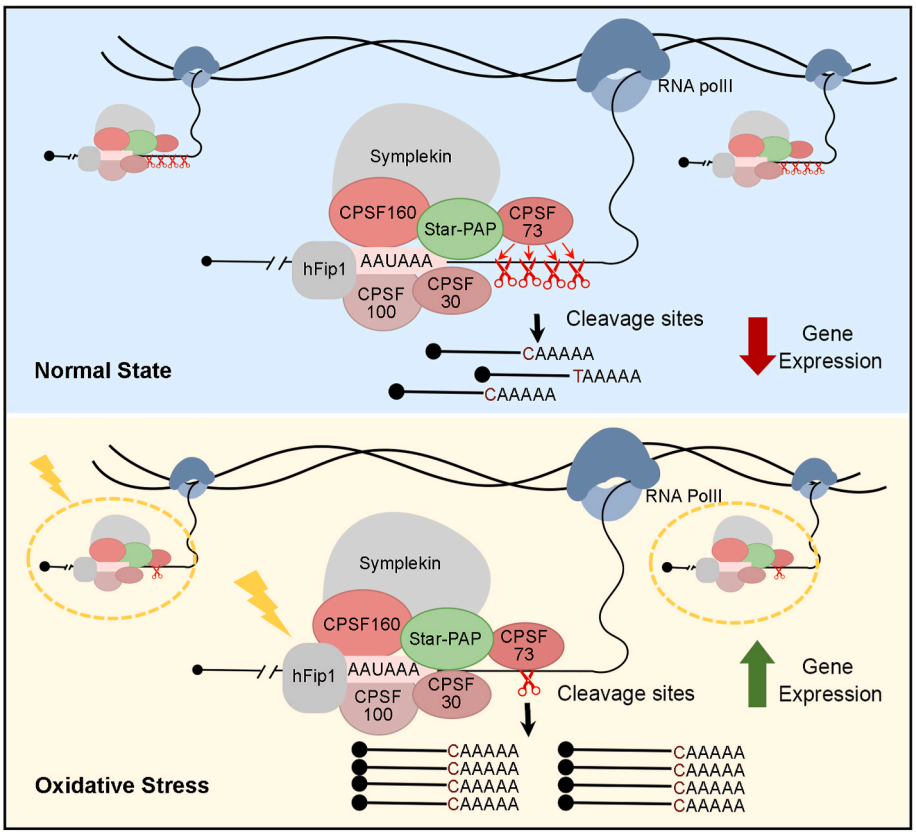


Fig. 7. Schematics of cleavage site heterogeneity and its regulation in oxidative stress response. Model depicts a reduction in the CSH where the number of subsidiary CSs are decreased while increasing the usage of the primary CS. This induces both cleavage and polyadenylation efficiency thereby stimulating stress response protein expression downstream of oxidative stress signal.

cleavage efficiency[83]. However, the exact mechanism of how polyadenylation activity affects cleavage is still unclear. Our study suggests a potential relationship between the cleavage efficiency and the polyadenylation activity through CSH. The PRR region of Star-PAP splits the PAP domain and regulates Star-PAP polyadenylation activity through phosphorylation, particularly during the antioxidant response[62]. Deletion of PRR on Star-PAP will compromise Star-PAP's antioxidant-mediated stimulation of cleavage complex assembly inducing CSH. This induction in the CSH will inhibit the stimulation of cleavage and polyadenylation of target mRNAs. However, further studies are required to understand the role of the PRR region in regulating the fidelity of cleavage and how it transduces the stress signal via CSH of target mRNAs. It is likely mediated through phosphorylation by different kinases, CKIα or PKCδ, that phosphorylates Star-PAP[62,84].

4. Materials and methods

4.1. In silico cleavage site analysis

We used 3′-READS sequencing data (GSE84461) to analyse global CS usage on each PA-site in HEK293 cells [57]. Additionally, RNA-seq datasets were employed to investigate oxidative stress responses: H_2O_2 treatment in Jurkat cells (GSE138290) and sodium arsenate treatment in mouse samples (GSE101851) [60,61]. Raw sequencing data (~7,000,000 reads per replicate) were retrieved from the GEO database. The raw data generated was checked for quality using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). RNA 3′-adapter sequences were identified using FastQC and were further trimmed using the cutadapt tool (https://doi.org/10.14806/ej.17.1.200). This pre-processing step yielded approximately ~3,000,000 trimmed reads. The trimmed reads were aligned to the human genome (hg38) using Bowtie2 in local mode [85]. The sodium arsenate-treated mouse data was aligned against the mouse genome (mm9). The MAPQ score (in Bowtie2) was set at 10, and a single read in the alignment file should map to a single locus in the genome [57]. The total number of reads mapped to human and mouse genomes was determined using samtools [86]. The alignment outputs in SAM/BAM formats were used for downstream analyses [86]. The uniquely mapped reads were then identified using the criteria for unique mapping with a MAPQ score exceeding 10, and at least two unaligned 5′ Ts indicative of the PA-tail as described earlier [57]. There were altogether ~1725768 uniquely mapped reads with a mapping score of >10, with a minimum of two 5′ Ts not aligned to the genome as described earlier [57]. Aligned data were refined by extracting information based on CIGAR strings using custom Python scripts. PA-sites were defined following previously established criteria requiring support from a minimum of two PA-site reads and contributing to over 5 % of a gene's expression [57]. Filtered reads that met these thresholds were localised within a 25-nucleotide range of the cleavage site downstream of the PA-signal hexamer and were clustered for CS analysis. Cleavage events occurring within 25 nucleotides downstream of a PA-signal hexamer were clustered as part of the same PA-site if they accounted for >5 % of the total transcript abundance in at least one replicate [57]. CSs were considered valid if supported by at least five reads converging at the 3′-end of a transcript. Altogether, we obtained >175,000 CSs clustered across >70,000 distinct PA-sites on ~12,000 assigned genes. The final tab-separated text file consisted of gene names, the location of each PA-site detected, the corresponding CS, and the number of reads of the respective CS-specific transcripts. Chromosomal locations of different CSs across mRNAs were visualised using the Integrative Genome Viewer (IGV) [87]. As described below, data formatting and processing were further performed using Python and Perl scripts to generate different plots.

4.2. Plots for data visualisation

Circos Plot: Data formatting and processing for generating a Circos plot was performed using Perl scripts. CS data, including their chromosomal locations and frequency of occurrence, was taken from the final tab-separated text file from the 3′-READS analysis above. Perl script was used to integrate chromosome locations with each CS following the Circos software to visualise the genome-wide CSH distribution from the 3′-READS datasets (Krzywinski et al., 2009). Circos plot integrates chromosomal structure in a circular format and displays the distribution of CSs mapped to their respective PA-sites on corresponding chromosomal locations. The plot consists of five concentric tracks: the outermost track represents the chromosomal ideogram marking the physical arrangement of genes across the genome; the second track represents the location and the number of PA-sites from the PolyA DB database corresponding to respective chromosomal locations[88]; and the remaining three inner tracks display experimental data of CSH plotted against the genomic locations in the outer tracks from three replicate datasets. Each line in the plot represents a PA-site and the radial length of the line in each track corresponds to the number of CS of that PA-site representing heterogeneity. Thus, the circumference (X axis) in the plot shows the location of the PA-site and the length (Y axis) represents the number of CSs on respective PA-sites. The plot compares the distribution of the number of CSs on different PA-sites and their chromosomal locations of three different replicate 3′-READS datasets, highlighting the overlap of the CS and PA-site distribution in the datasets.

CS usage distribution: The SAM files from the 3′-READS analysis of the clustered CS on respective PA-sites and their read counts were processed using Python scripts, and dot plots were generated using Matplotlib to visualise the usage distribution pattern of different CSs. For this purpose, the CSs were ranked from 1st to 20th based on the number of supporting reads of the respective transcripts, and the percentage usage of each CS was calculated relative to the total reads across all CSs of a PA-site. The CS that was used maximally (highest number of supporting reads) was assigned 1st (primary CS) followed by CS numbers based on the rank of CS usage (2nd to 20th) (subsidiary CS). Dot plots were then generated using Matplotlib to display the percentage usage of CSs ranked from 1st (the primary CS) to 20th (in blue dots) alongside the median usage values (in red dots). Here, the X-axis represents the number of CSs ranked from 1st to 20th, plotted against their % usage pattern on different PA-sites in the Y-axis. This plot provides insights into the overall distribution and variability of CS ranked according to their usage, and a predominance of the primary CS site over the subsidiary sites.

Further, stacked column plots were generated to analyse the different CS usage patterns on distinct PA-sites of individual mRNAs. Here, the % of primary CS usage is depicted in black, while the other subsidiary site usage is shown in different shades of grey.

Half Donut Plots: From the filtered 175,000 CSs clustered across 70,000 distinct PA-sites on ~12,000 assigned genes obtained from our 3′-READS analysis, the PA-sites with one CS and those with multiple CSs were separated. The number of PA-sites with only one CS (non-CSH) versus the number of PA-sites with more than one CS (showing CSH) was compared using a half donut plot in Excel 2019. Similarly, the number of PA-sites on mRNAs that show alternative polyadenylation and that do not show alternative polyadenylation were also plotted in half donuts to compare the number of PA-sites that have only one CS versus the number of PA-sites that show CSH in both the cases.

Relative gene expression analysis: For relative gene expression analysis, we used the final tab-separated text file that consists of information on each gene along with respective PA-sites detected for each gene, and the corresponding CSs and the number of associated reads. Processing of these data to analyse relative gene expression was carried out using Python scripts. The analysis involved the assessment of the total read counts per million (CPM) of each CS-specific transcript of a PA-site that was normalised against the CPM values of the internal control gene *GAPDH*. The expression from each PA-site was calculated as the sum of the CPM values from all CSs associated with that PA-site. The relative expression from each PA-site was calculated by dividing the CPM value of a PA-site by the total CPM value of the corresponding mRNA, which accounts for all PA-sites associated with the gene. Then, the relative expression from each PA-site was plotted against the number of CSs associated with each PA site using the Matplotlib library.

For the average expression per PA-site, we determined the mean value of relative expressions from all PA-sites as in Fig. 2A having different numbers of CS (total relative CPMs of all the PA-sites divided by the number of PA-sites in each group) and plotted against the increasing number of CS.

Further, relative expression level from a CS was determined by dividing the CPM value of a CS-specific transcript normalised to *GAPDH* by the total CPM values from all CSs of the corresponding PA-site-specific mRNA isoform. Then, the relative expression values of the primary CSs were plotted against the increasing number of CSs associated with each PA-site.

Jitter Plot: To visualise the distribution of CSH across different chromosomes in the three datasets, we used the final tab-separated text file from the analysis of the 3′-READS datasets above. A jitter plot was generated using the matplotlib library in Python that shows PA-sites associated with an increasing number of CSs on different chromosomes (the X-axis shows the chromosome number while the Y-axis represents the number of CSs detected on each PA-site). CS data from three different replicate datasets are indicated with three different colours (green, yellow, and purple). In this plot, each dot represents PA-site with a distinct number of CS, and the size and intensity of the dot show the number of CSs associated with respective PA-sites on different chromosome.

4.3. Cell lines, transient transfections, and cellular treatments

Human Embryonic Kidney 293 (HEK293) and HeLa cells were obtained from the American Type Culture Collection. Cells were then cultured in Dulbecco's modified Eagle's medium (DMEM) (Himedia, IN) supplemented with high glucose in the presence of 10 % foetal bovine serum and a mix of 50 IU/ml penicillin and streptomycin (Gibco, USA). Cell cultures were maintained in a CO_2 incubator at 37 °C (Thermo Scientific, USA).

For transfection, cells were cultured until they reached approximately 60 % confluency, and transfections of siRNA oligos and plasmid DNA were carried out using a chemical method (calcium phosphate), as described earlier[58]. About 1.5 μg of plasmid DNA per 0.5×10^6 cells or 80 nM of siRNA oligos (Eurogentec, Belgium) were used for transfection. Transfection was repeated 24 h later. Cells were collected 48 h after transfection for downstream analysis of RNAs and proteins.

4.4. Induction of antioxidant response

HEK293 cells were grown to ~70 % confluency. To induce an antioxidant response, cells were treated with 100 μM *tert*-Butyl Hydroquinone (tBHQ), an oxidative stress agonist in DMSO (Sigma, St. Louis, Missouri, USA) for 4 h. DMSO was used as the vehicle control[58]. To induce cells with peroxide stress, 50 μM of H_2O_2 was added to the culture media, then incubated for 2 h, followed by harvesting as per the protocol described earlier[60]. Heavy metal-mediated oxidative stress was induced by treatment with 250 μM $NaAsO_2$ in the cell culture media for 6 h, followed by harvesting the cells for further processing[61].

The antioxidant pathway mediated by NRF2 was blocked using Bexarotene (a potent NRF2 inhibitor). 100 nM of Bexarotene in HEK293 was treated on 60 % confluent HEK293 cells for 24 h [89]. The cells were then harvested and processed for downstream experiments.

4.5. DNA constructs

FLAG-tagged pCMV Tag2a reporter construct used in this study consists of an *NQO1* coding sequence under the control of the pCMV promoter. Additionally, this construct is driven by the full-length *NQO1* 3′-UTR region, approximately 1950 nucleotides long, as described earlier[58]. The distal *NQO1* PA-site was used for the mutational analysis; this PA-site has five major cleavage sites that are consistently observed (namely P, S_1, S_2, S_3, S_4) downstream of which was mutated by site-directed mutagenesis individually and in combination (see Supplemental Table S1). In the first batch of mutants, one CS was mutated at a time: P^m (primary CS), S_1^m to S_4^m (subsidiary CS site S1 to S4). The next batch of mutants have combinations of subsidiary CS mutations. To understand the characteristics of the primary CS sequence and position, corresponding mutations of the CS (CA to GA, TA, AA) and the position of the CS (−10 bp upstream and downstream from the original CS position) were introduced(see Supplemental Table S1). Silent mutations to make the *FLAG-NQO1* construct insensitive to the siRNA sequence employed for the *NQO1* knockdown were introduced using a pair of primers bearing three different mutations on the targeting sequence of the siRNA.

For the synthesis of in vitro transcribed *NQO1* 3′-UTR RNA, a 3′-UTR region (~250 nucleotide) around the distal PA-site region of *NQO1* encompassing the sequences required for both cleavage and polyadenylation were cloned in pTZ vector [58]. Mutations of the CS were introduced in the pTZ19R construct, individually and in combination, as described above. Primer sequences for the site-directed mutagenesis are shown in the Supplemental data.

4.6. Site-directed mutagenesis

Site-directed mutagenesis of the CS or siRNA insensitive mutation was carried out as described earlier [58] using a pair of primers incorporating the changed nucleotides. Briefly, around 50 ng of parent plasmid was taken, and mutations were introduced by inverse PCR reaction using Pfu Ultra II polymerase (Agilent, cat no. 600674) enzyme using primer pairs having the respective mutations. The PCR reaction was digested using Dpn1(NEB, cat no. R0176L) to remove the template plasmid treatment for 1 h at 37 °C, then transforming DH5α competent cells with the in vitro amplified plasmid. Mutated plasmids were then confirmed by Sanger sequencing. Primers used to introduce mutations are listed in the Supplemental data.

4.7. RNA immunoprecipitation

RNA IP was carried out as described earlier [90]. HEK293 cells transfected with wild-type FLAG-tagged NQO1 or various mutant variants were crosslinked with 1 % formaldehyde for 10 min, followed by quenching with 0.125 M glycine for 5 min. Cells were washed in 1 x PBS followed by lysis in 300 μl of cell lysis buffer (10 mM Tris-HCl pH 8.0, 10 mM NaCl, 0.2 % NP-40, 1 × EDTA-free Proteinase Inhibitor, and 1000 U RNase I, Promega, Madison, WI, USA). The crude nuclei were pelleted and resuspended in 400 μl nuclei lysis buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1 % SDS, 1 × EDTA-free Proteinase Inhibitor, and 1000 U RNase I Promega, Madison, WI, USA). The supernatant obtained after sonication and centrifugation steps was digested with DNase I for 20 min at 37 °C. The digestion was stopped by the addition of 20 mM EDTA. Monoclonal antibodies against CPSF73, Star-PAP, and IgG control were added to the supernatant and incubated overnight at 4 °C. Subsequently, Sepharose A beads equilibrated with IP dilution buffer were added, and the mixture was incubated for 2 h at 4 °C. The formed complexes were then pelleted and washed in wash buffer 3 times for 5 min each at 3500 rpm. It was then eluted with elution buffer (1 % SDS and 100 mM Sodium bicarbonate and proteinase K). Crosslinking was reversed by incubating the mixture at 67 °C for 4 h. RNA was isolated using Trizol, and cDNAs were synthesised using random hexamers and MMLV-RT (Invitrogen, Waltham, MA, USA). PCR amplification was carried out with gene-specific primers, and products were visualised on an agarose gel.

4.8. In vitro RNA synthesis

NQO1 RNA substrate was derived from the plasmid pTZ-*NQO1*, which harbours the *NQO1* 3′-UTR under the T7 promoter. This plasmid includes a segment of the *NQO1* 3′-UTR containing the PA-signal, spanning from 120 bp upstream to 122 bp downstream of the cleavage site, inserted into the *Pst*I and *Bam*HI sites of the pTZ19R vector (Fermentas)[58]. Radiolabelled RNA substrates were uniformly prepared through in vitro transcription of linearised plasmid templates using the T7 Transcription Kit (Invitrogen). A typical 20 μl reaction consisted of 1x reaction buffer (supplied in the kit), 20 units of RNase inhibitor, 1 μg of DNA, 500 μM each of ATP and CTP, 50 μM UTP, and 100 μM GTP, supplemented with 50 μCi of α-32P-UTP and 400 μM m7G cap analogue, along with 40 units of T7 RNA polymerase. Following transcription, RNA was purified using the Qiagen Mini RNeasy Kit for subsequent experimentation.

Fluorescent RNA was prepared by a similar in vitro transcription reaction using a pTZ19R plasmid construct expressing pTZ19R *NQO1* 3′-UTR as described above in the presence of Fluorescein-12-UTP (Sigma cat no.11427857910). The reaction mix consisted of 1x reaction buffer, 20 units of RNase inhibitor, 1 μg of DNA, 500 μM each of ATP and CTP, 50 μM UTP, and 100 μM GTP, supplemented with 50 μM Fluorescein-12-UTP, along with 40 units of T7 RNA polymerase as described above.

4.9. In vitro pre-mRNA cleavage assay

Cleavage experiments were carried out as described earlier[25]. Each reaction included a ^{32}P-labelled pre-mRNA substrate in a 25 μl volume of cleavage buffer. The buffer contained 20 mM creatinine phosphate, 0.5 mM $MgCl_2$, 10 % glycerol, 10 mM HEPES (pH 7.9), 50 mM KCl, 0.05 mM EDTA, and 1 % polyvinyl alcohol (PVA), supplemented with 0.8 mM 3′-dATP and 5 μl of HeLa nuclear extract. Reactions were incubated at 30 °C for 2 h and stopped by adding a proteinase K mixture comprising 2 % sarcosyl, 100 mM Tris-Cl (pH 7.5), 20 mM EDTA, and 400 mg/ml proteinase K, followed by an additional incubation at 30 °C for 10 min. RNA extraction was performed using phenol-chloroform, followed by precipitation with absolute alcohol in the presence of 3 M ammonium acetate and 1 μg carrier tRNA. The extracted RNA was then analysed on a 6 % urea-denaturing gel.

4.10. In vitro polyadenylation assay

In vitro polyadenylation assay was carried out as described earlier [58]. Briefly, DNA templates with the forward T7 promoter and reverse primer ending with a CS were amplified from the pTZ19R *NQO1* 3′-UTR vector. The amplified templates were then subjected to in vitro RNA transcription using T7 RNA polymerase generating cleaved RNA templates, ending at different cleavage sites, either the primary or the subsidiary sites of *NQO1*. RNA templates were incubated in an equivalent of ~10 μg of HeLa nuclear extract prepared using the NEPER nuclear extraction kit (ThermoScientific) with α-^{32}P ATP in a PAP assay buffer (250 mM NaCl, 50 mM Tris-HCl, and 10 mM $MgCl_2$, pH 7.9) at 37 °C for 1 h. Then, the polyadenylated RNA was precipitated and washed with 75 % ethanol. RNA products were then analysed on a 6 % urea-denaturing polyacrylamide gel and visualised by phosphor imaging.

4.11. RNA EMSA

EMSA experiments were carried out as described [91] in a 20 μl EMSA-binding buffer (10 mM Tris-HCl, pH-7.5, 1 mM EDTA, 50 mM NaCl, 0.5 mM $MgCl_2$, 1 mM DTT) in the presence of 1 μg/ml bovine

serum albumin, 50 % glycerol with 0.1 nM fluorescent RNA templates. The rection mixture was incubated in the presence of control and tBHQ treated FLAG purified Star-PAP protein complex (30 nM) respectively at room temperature for 30 min. The RNA-Protein complex was analysed on a native polyacrylamide gel and visualised on an iBright FL1500 platform (Invitrogen).

4.12. Immunoblotting experiment

For immunoblotting, cells were lysed in a cell lysis buffer comprising 0.06 M Tris-HCl (pH 7.4), 25 % Glycerol, 2 % SDS, 0.002 % Bromophenol blue, and 1 % β-mercaptoethanol (1 × SDS-PAGE loading buffer) as described earlier [92]. The lysed samples were denatured by heating at 95 °C for 20 min and then analysed on an SDS-PAGE gel in 1 × Tris Glycine Buffer (25 mM Tris pH 8.0, 190 mM Glycine, 0.1 % SDS). Following separation, the proteins were transferred onto a PVDF membrane using a transfer buffer (25 mM Tris-Cl pH 8.0, 190 mM Glycine, 20 % Methanol) at 110 V for 90 min. The PVDF membrane was blocked in 5 % skimmed milk in 1 × TBST (20 mM Tris pH 7.4, 150 mM NaCl, 0.1 % Tween-20) for 45 min at room temperature (RT). Blots were subsequently incubated with primary antibodies at a dilution of 1:3000 in TBST. After three washes in TBST for 10 min each, blots were incubated with HRP-conjugated secondary antibody (Jackson Immuno Research Laboratory, West Grove, PA, USA) at a dilution of 1:5000. Finally, blots were imaged using a chemiluminescent substrate (Bio-Rad, Hercules, CA, USA) on an iBright FL1500 platform (Invitrogen, Waltham, MA, USA). A list of antibodies employed in the study is shown in Supplemental data.

4.13. Immunoprecipitation experiment

For immunoprecipitation, cell lysates were sonicated (28 % amplitude for 5 s on and 5 s off pulse for 5 min) in a cell lysis buffer as described earlier [92] containing 100 mM KCl, 50 mM Tris-HCl (pH 7.4), 0.5 % Nonidet NP-40, 5 mM EDTA, 200 mM $NaVO_4$, 50 mM NaF, 50 mM L-glycerophosphate, and 1 × EDTA-free protease inhibitor, supplemented with 100 μg/ml RNaseA depending on the reaction conditions. Following sonication, lysates were centrifuged at 12,500 rpm for 15 min at 4 °C to collect the cellular lysates. FLAG-tagged Star PAP expressing lysates were incubated in the IP buffer, where FLAG M2 beads were equilibrated overnight at 4 °C. The following day, beads were centrifuged and precipitated, and the supernatant was discarded and washed three times with IP buffer for 10 min each. Finally, an SDS loading buffer was added to the beads, and the mixture was loaded onto an SDS-PAGE gel after denaturation at 95 °C for 5 min. The input loaded on the gel corresponded to 10 % of the total protein utilised in each immunoprecipitation experiment[92]. Antibodies used for immunoblotting are listed in the Supplemental data.

4.14. Real-time qPCR

The quantitative real-time PCR (qRT-PCR) experiments were conducted using the CFX98 multi-color system (Bio-Rad). iTaq universal SYBR Green Supermix was utilised for the PCR reactions with cDNA synthesised by reverse transcription using MMLV-RT from 2 μg total RNA samples as described earlier[92]. Melt curve analysis was performed for each qRT-PCR experiment to verify single-product amplification. To ensure the specificity and accuracy of the PCR reaction, primer efficiency was maintained close to 100 % in all experiments.

For quantification of the real-time experiments, target mRNA levels were normalised to *GAPDH* levels, and changes in expression levels were expressed as fold-change relative to the control samples using the Pfaffl method. Three independent experiments were conducted for each qPCR experiment ($n > 3$). Primers employed in the qRT-PCR are shown in the Supplemental data.

4.15. 3′-RACE and pre-mRNA cleavage assay

3′-RACE assays were conducted using 2 μg of total RNA extracted from HEK293 cells using an RNA isolation kit (Promega) following established protocols [92]. An engineered oligodT reverse primer (comprising 16 oligo dT nucleotides preceded by a 20-nucleotide engineered sequence at the 5′-end) was employed, and a reverse transcription reaction was performed with 100 units of MMLV-RT (Invitrogen, cat. 28,025–013) at 42 °C for 1 h, followed by inactivation at 70 °C for 15 min. Subsequently, 250 ng of cDNA was amplified using a gene-specific forward primer (specific to each gene) and an AUAP reverse primer (specific to the engineered region of the oligo dT primer). The 3′-RACE products were then validated through sequencing.

For the in vivo cleavage assay, cleavage efficacy was evaluated by measuring the uncleaved pre-mRNA level by using a pair of primers across the CS by qRT-PCR and expressed as fraction cleavage relative to the total mRNA level using a pair of primers on an upstream exon. Total mRNA was isolated from HEK293 cells in the presence and absence of WT or different CS mutants of *NQO1*, and qRT-PCR was carried out as described above to measure the uncleaved pre-mRNA levels and the total mRNA levels. The levels of uncleaved pre-mRNAs were normalised to internal control *GAPDH* levels, and then fraction cleavage was analysed relative to total mRNA levels. The relative cleavage was then expressed as fold change relative to respective controls in different plots. Primers used for 3′-RACE and pre-mRNA cleavage assay are listed in the Supplemental data.

4.16. MTT assay

The MTT assay was carried out to assess cell viability as described earlier[90]. Transfected cells were seeded onto a 96-well plate at a density of 1×10^4 cells per well. Following a 48-h incubation period in a humidified incubator at 37 °C with 5 % CO_2, 3-(4, 5-dimethylthiazolyl-2-yl)-2–5 diphenyl tetrazolium bromide (MTT, Sigma, St. Louis, Missouri, USA) was introduced. Subsequently, the plates were further incubated for 3 h, followed by the addition of isopropanol after removing the culture medium. The resultant formazan crystals were resuspended, and absorbance was measured at 570 nm utilising a Varioskan LUX multimode microplate reader (ThermoScientific).

4.17. Trypan blue staining

Trypan blue solution was prepared by mixing equal parts of 0.4 % trypan blue solution and cell suspension, followed by incubation for 3 min at room temperature. Then, a drop of the trypan blue/cell mixture was applied onto a hemocytometer and the unstained (viable) and stained (nonviable) cells were separately counted. To determine the total number of cells in the aliquot, viable and nonviable cells were added together, and the dilution factor was multiplied. Finally, the percentage of viable cells was calculated in each treatment condition [93].

4.18. Catalase activity assays

Catalase activity was determined using Catalase Assay Kit (Cayman chemical cat no. 707002). Briefly, cells were collected by cell scrapper and then homogenised in 1 ml cold lysis buffer (50 mM potassium phosphate, pH 7, containing 1 mM EDTA). Then, the cell suspension was centrifuged at 10,000 g for 15 min at 4 °C, and the supernatant was collected. Further, a series of catalase standards were prepared with known concentrations (from 0 to 10 nmol/ml) using a catalase standard solution. Samples were transferred to a microplate well, and then 20 μl hydrogen peroxide solution was added to all wells, including standards, to initiate the reaction. The microplate was incubated at room temperature for 20 min to allow the catalase enzyme to react with hydrogen

peroxide. Then, 30 μl of potassium hydroxide was added to each well to terminate the reaction, followed by 30 μl of chromogen. The plate was further incubated for 10 min at room temperature in a shaker, and then 10 μl of catalase potassium periodide was added to each well, followed by an additional incubation in the dark for 5 min, and absorbance was measured at 540 nm using a spectrophotometer.

4.19. RNA biotinylation and pulldown

Biotinylated RNA was prepared by in vitro transcription on the pTZ19R plasmid template with Biotin UTP in the in vitro RNA synthesis method as per the manufacturer's instruction (Thermo scientific Cat no. EP0112). For biotinylation pull down, HEK cell lysates were prepared from 90 % confluent 10 cm dish in lysis buffer (10 mM HEPES, pH-7, 200 mM NaCl, 1 % Triton-X 100, 10 mM $MgCl_2$, 1 mM DTT and protease inhibitor) as described earlier[91]. Briefly, 30 μg of biotinylated RNA were incubated with 60 μl of pre-washed avidin-agarose beads at 4 °C for 1 h. Agarose beads were washed three times with RNA binding buffer (10 mM Tris HCl, 0.1 M KCl, and 10 mM $MgCl_2$) and incubated with 5 mg total protein equivalent cell lysates overnight at 4 °C. Agarose beads were collected and washed, and proteins were eluted using an SDS loading buffer after boiling for 10 min at 95 °C. Associated proteins were analysed by Western blotting using specific antibodies.

4.20. Blot quantification

Quantification of blots was performed using Image J software (NIH) [94]. Three replicates of blots were analysed to quantify relative intensities. Bands were quantified in arbitrary units and expressed relative to a reference protein. In the case of immunoprecipitation, bands were normalised to IgG and expressed relative to the Input sample.

4.21. Statistics

Statistical significance was determined using ANOVA to analyse the differences in the means of experimental replicates from each experiment using GraphPad Prism 10. A minimum of three independent experiments were utilised to quantify each experimental dataset. All data are presented as mean ± standard error mean (SEM). Differences were considered statistically significant at a p-value <0.05.

CRediT authorship contribution statement

Feba Shaji: Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **Jamshaid Ali:** Formal analysis, Software, Visualization. **Rakesh S. Laishram:** Conceptualization, Data curation, Funding acquisition, Investigation, Resources, Supervision, Writing – original draft, Writing – review & editing.

Data availability statement

The article and supplemental data contain all the data generated during this study.

Funding

This work was supported by Department of Science and Technology, Govt. of India, Swarnajayanti Fellowship (SB/SJF/2019-20/09) and Anusandhan National Research Foundation, Govt. of India (CRG/2023/003578) to RSL, and SRF fellowship from Council of Scientific and Industrial Research to FS.

Declaration of competing interest

The authors declare no competing interests.

Acknowledgement

We thank Fiona P. Ukken (University of Maryland, USA) and RSL lab members for carefully reading the manuscript. We thank Dr Ani V Das, RGCB-Trivandrum for providing the NRF2 inhibitor.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.redox.2025.103565.

References

[1] V. Boreikaite, L.A. Passmore, 3'-End processing of eukaryotic mRNA: machinery, regulation, and Impact on gene expression, Annu. Rev. Biochem. 92 (2023) 199–225.
[2] J. Neve, R. Patel, Z. Wang, A. Louey, A.M. Furger, Cleavage and polyadenylation: ending the message expands gene regulation, RNA Biol. 14 (2017) 865–890.
[3] C.R. Mandel, Y. Bai, L. Tong, Protein factors in pre-mRNA 3'-end processing, Cell. Mol. Life Sci. 65 (2008) 1099–1122.
[4] K. Xiang, L. Tong, J.L. Manley, Delineating the structural blueprint of the pre-mRNA 3'-end processing machinery, Mol. Cell Biol. 34 (2014) 1894–1910.
[5] S. Bienroth, W. Keller, E. Wahle, Assembly of a processive messenger RNA polyadenylation complex, EMBO J. 12 (1993) 585–594.
[6] G.M. Gilmartin, J.R. Nevins, Molecular analyses of two poly(A) site-processing factors that determine the recognition and efficiency of cleavage of the pre-mRNA, Mol. Cell Biol. 11 (1991) 2432–2438.
[7] U. Kuhn, M. Gundel, A. Knoth, Y. Kerwitz, S. Rudel, E. Wahle, Poly(A) tail length is controlled by the nuclear poly(A)-binding protein regulating the interaction between poly(A) polymerase and the cleavage and polyadenylation specificity factor, J. Biol. Chem. 284 (2009) 22803–22814.
[8] Y. Takagaki, L.C. Ryner, J.L. Manley, Four factors are required for 3'-end cleavage of pre-mRNAs, Genes Dev. 3 (1989) 1711–1724.
[9] S.L. Chan, I. Huppertz, C. Yao, L. Weng, J.J. Moresco, J.R. Yates 3rd, J. Ule, J. L. Manley, Y. Shi, CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing, Genes Dev. 28 (2014) 2370–2380.
[10] L. Schönemann, U. Kühn, G. Martin, P. Schäfer, A.R. Gruber, W. Keller, M. Zavolan, E. Wahle, Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33, Genes & development: gad. (2014) 250985, 250114.
[11] Y. Sun, Y. Zhang, K. Hamilton, J.L. Manley, Y. Shi, T. Walz, L. Tong, Molecular basis for the recognition of the human AAUAAA polyadenylation signal, Proc. Natl. Acad. Sci. U. S. A. 115 (2018) E1419–E1428.
[12] C.C. MacDonald, J. Wilusz, T. Shenk, The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location, Mol. Cell Biol. 14 (1994) 6647–6654.
[13] K.G. Murthy, J.L. Manley, The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation, Genes Dev. 9 (1995) 2672–2683.
[14] K.M. Brown, G.M. Gilmartin, A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im, Mol Cell 12 (2003) 1467–1476.
[15] H. de Vries, U. Ruegsegger, W. Hubner, A. Friedlein, H. Langen, W. Keller, Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors, EMBO J. 19 (2000) 5895–5904.
[16] U. Ruegsegger, K. Beyer, W. Keller, Purification and characterization of human cleavage factor Im involved in the 3' end processing of messenger RNA precursors, J. Biol. Chem. 271 (1996) 6107–6113.
[17] R.C. Deo, J.B. Bonanno, N. Sonenberg, S.K. Burley, Recognition of polyadenylate RNA by the poly(A)-binding protein, Cell 98 (1999) 835–845.
[18] C.R. Mandel, S. Kaneko, H. Zhang, D. Gebauer, V. Vethantham, J.L. Manley, L. Tong, Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease, Nature 444 (2006) 953–956.
[19] T. Raabe, K.G. Murthy, J.L. Manley, Poly(A) polymerase contains multiple functional domains, Mol. Cell Biol. 14 (1994) 2946–2957.
[20] K. Ryan, O. Calvo, J.L. Manley, Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease, RNA 10 (2004) 565–573.
[21] C.B. Kyriakopoulou, H. Nordvarg, A. Virtanen, A novel nuclear human poly(A) polymerase (PAP), PAP gamma, J. Biol. Chem. 276 (2001) 33504–33511.
[22] T. Raabe, F.J. Bollum, J.L. Manley, Primary structure and expression of bovine poly (A) polymerase, Nature 353 (1991) 229–234.
[23] S.L. Topalian, S. Kaneko, M.I. Gonzales, G.L. Bond, Y. Ward, J.L. Manley, Identification and functional characterization of neo-poly(A) polymerase, an RNA processing enzyme overexpressed in human tumors, Mol. Cell Biol. 21 (2001) 5614–5623.
[24] E. Wahle, G. Martin, E. Schiltz, W. Keller, Isolation and expression of cDNA clones encoding mammalian poly(A) polymerase, EMBO J. 10 (1991) 4251–4257.
[25] R.S. Laishram, R.A. Anderson, The poly A polymerase Star-PAP controls 3'-end cleavage by promoting CPSF interaction and specificity toward the pre-mRNA, EMBO J. 29 (2010) 4132–4145.
[26] D.L. Mellman, M.L. Gonzales, C. Song, C.A. Barlow, P. Wang, C. Kendziorski, R. A. Anderson, A PtdIns4,5P2-regulated nuclear poly(A) polymerase controls expression of select mRNAs, Nature 451 (2008) 1013–1017.

[27] R. Elkon, A.P. Ugalde, R. Agami, Alternative cleavage and polyadenylation: extent, regulation and function, Nat. Rev. Genet. 14 (2013) 496–506.
[28] S. Komili, P.A. Silver, Coupling and coordination in gene expression processes: a systems biology view, Nat. Rev. Genet. 9 (2008) 38–48.
[29] A. Kumar, M. Clerici, L.M. Muckenfuss, L.A. Passmore, M. Jinek, Mechanistic insights into mRNA 3'-end processing, Curr. Opin. Struct. Biol. 59 (2019) 143–150.
[30] M.J. Moore, N.J. Proudfoot, Pre-mRNA processing reaches back to transcription and ahead to translation, Cell 136 (2009) 688–700.
[31] L.A. Passmore, J. Coller, Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression, Nat. Rev. Mol. Cell Biol. 23 (2022) 93–106.
[32] P. Mitchell, D. Tollervey, mRNA stability in eukaryotes, Curr. Opin. Genet. Dev. 10 (2000) 193–198.
[33] L. Weill, E. Belloc, F.A. Bava, R. Mendez, Translational control by changes in poly (A) tail length: recycling mRNAs, Nat. Struct. Mol. Biol. 19 (2012) 577–585.
[34] H. Imataka, A. Gradi, N. Sonenberg, A newly identified N-terminal amino acid sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent translation, EMBO J. 17 (1998) 7480–7489.
[35] A. Kahvejian, Y.V. Svitkin, R. Sukarieh, M.N. M'Boutchou, N. Sonenberg, Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms, Genes Dev. 19 (2005) 104–113.
[36] X. Qu, S. Lykke-Andersen, T. Nasser, C. Saguez, E. Bertrand, T.H. Jensen, C. Moore, Assembly of an export-competent mRNP is needed for efficient release of the 3'-end processing complex after polyadenylation, Mol. Cell Biol. 29 (2009) 5327–5338.
[37] M. Shi, H. Zhang, X. Wu, Z. He, L. Wang, S. Yin, B. Tian, G. Li, H. Cheng, ALYREF mainly binds to the 5' and the 3' regions of the mRNA in vivo, Nucleic Acids Res. 45 (2017) 9640–9653.
[38] J.C. Dantonel, K.G. Murthy, J.L. Manley, L. Tora, Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA, Nature 389 (1997) 399–402.
[39] S.M. Flaherty, P. Fortes, E. Izaurralde, I.W. Mattaj, G.M. Gilmartin, Participation of the nuclear cap binding complex in pre-mRNA 3' processing, Proc. Natl. Acad. Sci. U. S. A. 94 (1997) 11893–11898.
[40] X. Jiao, J.H. Chang, T. Kilic, L. Tong, M. Kiledjian, A mammalian pre-mRNA 5' end capping quality control mechanism and an unexpected link of capping to pre-mRNA processing, Mol Cell 50 (2013) 104–115.
[41] C.K. Mapendano, S. Lykke-Andersen, J. Kjems, E. Bertrand, T.H. Jensen, Crosstalk between mRNA 3' end processing and transcription initiation, Mol Cell 40 (2010) 410–422.
[42] S. Awasthi, J.C. Alwine, Association of polyadenylation cleavage factor I with U1 snRNP, RNA 9 (2003) 1400–1409.
[43] L. Huang, G. Li, C. Du, Y. Jia, J. Yang, W. Fan, Y.Z. Xu, H. Cheng, Y. Zhou, The polyA tail facilitates splicing of last introns with weak 3' splice sites via PABPN1, EMBO Rep. 24 (2023) e57128.
[44] A. Kyburz, A. Friedlein, H. Langen, W. Keller, Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3' end processing and splicing, Mol Cell 23 (2006) 195–205.
[45] S. Millevoi, C. Loulergue, S. Dettwiler, S.Z. Karaa, W. Keller, M. Antoniou, S. Vagner, An interaction between U2AF 65 and CF I(m) links the splicing and 3' end processing machineries, EMBO J. 25 (2006) 4854–4864.
[46] S. Vagner, C. Vagner, I.W. Mattaj, The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing, Genes Dev. 14 (2000) 403–413.
[47] D.C. Di Giammartino, K. Nishida, J.L. Manley, Mechanisms and consequences of alternative polyadenylation, Molecular cell 43 (2011) 853–866.
[48] N.K. Mohanan, F. Shaji, G.R. Koshre, R.S. Laishram, Alternative polyadenylation: an enigma of transcript length variation in health and disease, Wiley Interdiscip Rev RNA 13 (2022) e1692.
[49] B. Tian, J.L. Manley, Alternative polyadenylation of mRNA precursors, Nat. Rev. Mol. Cell Biol. 18 (2017) 18–30.
[50] F. Chen, C.C. MacDonald, J. Wilusz, Cleavage site determinants in the mammalian polyadenylation signal, Nucleic Acids Res. 23 (1995) 2614–2620.
[51] E.A. Weiss, G.M. Gilmartin, J.R. Nevins, Poly(A) site efficiency reflects the stability of complex formation involving the downstream element, Embo J 10 (1991) 215–219.
[52] E. Pauws, A.H. van Kampen, S.A. van de Graaf, J.J. de Vijlder, C. Ris-Stalpers, Heterogeneity in polyadenylation cleavage sites in mammalian mRNA sequences: implications for SAGE analysis, Nucleic Acids Res. 29 (2001) 1690–1694.
[53] E.K. Stroup, Z. Ji, Deep learning of human polyadenylation sites at nucleotide resolution reveals molecular determinants of site usage and relevance in disease, Nat. Commun. 14 (2023) 7378.
[54] D.F. Bishop, R. Kornreich, R.J. Desnick, Structural organization of the human alpha-galactosidase A gene: further evidence for the absence of a 3' untranslated region, Proc. Natl. Acad. Sci. U. S. A. 85 (1988) 3903–3907.
[55] N.H. Gehring, U. Frede, G. Neu-Yilik, P. Hundsdoerfer, B. Vetter, M.W. Hentze, A. E. Kulozik, Increased efficiency of mRNA 3' end formation: a new genetic mechanism contributing to hereditary thrombophilia, Nat. Genet. 28 (2001) 389–392.
[56] M.D. Sheets, S.C. Ogg, M.P. Wickens, Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro, Nucleic Acids Res. 18 (1990) 5799–5805.
[57] W. Li, W. Li, R.S. Laishram, M. Hoque, Z. Ji, B. Tian, R.A. Anderson, Distinct regulation of alternative polyadenylation and gene expression by nuclear poly(A) polymerases, Nucleic Acids Res. 45 (2017) 8930–8942.
[58] A. Sudheesh, N. Mohan, N. Francis, R.S. Laishram, R.A. Anderson, Star-PAP controlled alternative polyadenylation coupled poly (A) tail length regulates protein expression in hypertrophic heart, Nucleic acids research 47 (2019) 10771–10787.
[59] A. Fabregat, S. Jupe, L. Matthews, K. Sidiropoulos, M. Gillespie, P. Garapati, R. Haw, B. Jassal, F. Korninger, B. May, et al., The reactome pathway knowledgebase, Nucleic Acids Res. 46 (2018) D649–D655.
[60] N.P.Y. Mullani, A. Mangelinck, C. Rachez, M. Costallat, E. Batsché, M. Goodhardt, G. Cenci, C. Mann, C. Muchardt, Reduced RNA turnover as a driver of cellular senescence, Life Sci. Alliance 4 (2021) e202000809.
[61] D. Zheng, R. Wang, Q. Ding, T. Wang, B. Xie, L. Wei, Z. Zhong, B. Tian, Cellular stress alters 3'UTR landscape through alternative polyadenylation and isoform-specific degradation, Nat. Commun. 9 (2018) 2268.
[62] M.L. Gonzales, D.L. Mellman, R.A. Anderson, CKIalpha is associated with and phosphorylates star-PAP and is also required for expression of select star-PAP target messenger RNAs, J. Biol. Chem. 283 (2008) 12665–12673.
[63] D.T. Kandala, N. Mohan, V. A, S. Ap, R. G, R.S. Laishram, CstF-64 and 3'-UTR cis-element determine Star-PAP specificity for target mRNA selection by excluding PAPalpha, Nucleic Acids Res. 44 (2016) 811–823.
[64] T.E. Adamson, D.C. Shutt, D.H. Price, Functional coupling of cleavage and polyadenylation with transcription of mRNA, J. Biol. Chem. 280 (2005) 32262–32271.
[65] F. He, X. Ru, T. Wen, NRF2, a transcription factor for stress response and beyond, Int. J. Mol. Sci. 21 (2020).
[66] Q. Ma, Role of nrf2 in oxidative stress and toxicity, Annu. Rev. Pharmacol. Toxicol. 53 (2013) 401–426.
[67] S. Chan, E.A. Choi, Y. Shi, Pre-mRNA 3'-end processing complex assembly and function, Wiley Interdiscip Rev RNA 2 (2011) 321–335.
[68] Y. Sun, K. Hamilton, L. Tong, Recent molecular insights into canonical pre-mRNA 3'-end processing, Transcription 11 (2020) 83–96.
[69] E. Wahle, W. Keller, The biochemistry of polyadenylation, Trends Biochem. Sci. 21 (1996) 247–250.
[70] M. Dreyfus, P. Regnier, The poly(A) tail of mRNAs: bodyguard in eukaryotes, scavenger in bacteria, Cell 111 (2002) 611–613.
[71] A. Radhakrishnan, R. Green, Connections underlying translation and mRNA stability, Journal of molecular biology 428 (18) (2016) 3558–3564.
[72] E. de Nadal, G. Ammerer, F. Posas, Controlling gene expression in response to stress, Nat. Rev. Genet. 12 (2011) 833–845.
[73] C. Tonelli, I.I.C. Chio, D.A. Tuveson, Transcriptional regulation by Nrf2, Antioxid Redox Signal 29 (2018) 1727–1745.
[74] X.M. Cheng, Y.Y. Hu, T. Yang, N. Wu, X.N. Wang, Reactive oxygen species and oxidative stress in vascular-related diseases, Oxid. Med. Cell. Longev. 2022 (2022) 7906091.
[75] Y.A. Hajam, R. Rani, S.Y. Ganie, T.A. Sheikh, D. Javaid, S.S. Qadri, S. Pramodh, A. Alsulimani, M.F. Alkhanani, S. Harakeh, et al., Oxidative stress in human pathology and aging: molecular mechanisms and perspectives, Cells 11 (2022).
[76] M.D. Jelic, A.D. Mandic, S.M. Maricic, B.U. Srdjenovic, Oxidative stress and its role in cancer, J Cancer Res Ther 17 (2021) 22–28.
[77] S.K. Raut, M. Khullar, Oxidative stress in metabolic diseases: current scenario and therapeutic relevance, Mol. Cell. Biochem. 478 (2023) 185–196.
[78] U. Saleem, S. Sabir, S.G. Niazi, M. Naeem, B. Ahmad, Role of oxidative stress and antioxidant defense biomarkers in neurodegenerative diseases, Crit. Rev. Eukaryot. Gene Expr. 30 (2020) 311–322.
[79] S.A. Abbass, S.L. Asa, S. Ezzat, Altered expression of fibroblast growth factor receptors in human pituitary adenomas, J. Clin. Endocrinol. Metab. 82 (1997) 1160–1166.
[80] C.F.E. Montell, M.H. Caruthers, A.J. Berk, Inhibition of RNA cleavage but not polyadenylation by a point mutation in mRNA 3' consensus sequence AAUAAA, Nature 19 (1983) 5935.
[81] G. Christofori, W. Keller, Poly(A) polymerase purified from HeLa cell nuclear extract is required for both cleavage and polyadenylation of pre-mRNA in vitro, Mol. Cell Biol. 9 (1989) 193–203.
[82] Y. Takagaki, L.C. Ryner, J.L. Manley, Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation, Cell 52 (1988) 731–742.
[83] M.P. Terns, S.T. Jacob, Role of poly(A) polymerase in the cleavage and polyadenylation of mRNA precursor, Mol. Cell Biol. 9 (1989) 1435–1444.
[84] W. Li, R.S. Laishram, Z. Ji, C.A. Barlow, B. Tian, R.A. Anderson, Star-PAP control of BIK expression and apoptosis is regulated by nuclear PIPKIalpha and PKCdelta signaling, Mol Cell 45 (2012) 25–37.
[85] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods 9 (2012) 357–359.
[86] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, S. Genome Project Data Processing, The sequence alignment/map format and SAMtools, Bioinformatics 25 (2009) 2078–2079.
[87] J.T. Robinson, H. Thorvaldsdottir, W. Winckler, M. Guttman, E.S. Lander, G. Getz, J.P. Mesirov, Integrative genomics viewer, Nat. Biotechnol. 29 (2011) 24–26.
[88] R. Wang, R. Nambiar, D. Zheng, B. Tian, PolyA_DB 3 catalogs cleavage and polyadenylation sites identified by deep sequencing in multiple genomes, Nucleic acids research 46 (D1) (2018).
[89] A. Warda, L.J.P. Staniszewski, Z. Sabir, S. Livingston, M. Sausedo, S. Reshi, E. Ron, M.T. Applegate, D. Haddad, M. Khamisi, P.A. Marshall, C.E. Wagner, P.W. Jurutka, Development of Bexarotene analogs for treating cutaneous T-cell lymphomas, Cells 12 (21) (2023) 2575.
[90] N.K. Mohanan, F. Shaji, A.P. Sudheesh, A. Bangalore Prabhashankar, N. R. Sundaresan, R.S. Laishram, Star-PAP controls oncogene expression through primary miRNA 3'-end formation to regulate cellular proliferation and tumour formation, Biochim. Biophys. Acta, Mol. Basis Dis. 1870 (2024) 167080.

[91] N. Francis, R.S. Laishram, Transgenesis of mammalian PABP reveals mRNA polyadenylation as a general stress response mechanism in bacteria, iScience 24 (2021) 103119.
[92] F.M.N.K. Shaji, S. Shahzad, G. V P, A. Bangalore Prabhashankar, N.R. Sundaresan, R.S. Laishram, Proto-oncogene cSrc-mediated RBM10 phosphorylation arbitrates anti-hypertrophy gene program in the heart and controls cardiac hypertrophy, Life Sci. 341 (2024).
[93] W. Strober, Trypan blue exclusion test of cell viability, Curr Protoc Immunol 111 (2015) A3.B.1–A3.B.3.
[94] C.A. Schneider, W.S. Rasband, K.W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis, Nat. Methods 9 (2012) 671–675.