**Characterizing heat shock-induced mRNA produced past the 3’ end of genes**

**Ana Mirita**

**Biochemistry Departmental Honors Thesis**

**University of Colorado Boulder**

Defense Date: October 28, 2024

**Defense Committee**

Dr. Jennifer Kugel, Department of Biochemistry

Dr. Joseph Falke, Department of Biochemistry

Stephanie Renfrow, Department of Ecology and Evolutionary Biology

**Table of Contents**

|  |  |
| --- | --- |
| **Abstract** | 1 |
| **Introduction** |  |
|  The central dogma | 2 |
|  Transcription  | 3 |
|  Transcriptional termination | 5 |
|  Cellular stress results in termination defects | 6 |
| **Results** |  |
|  Heat shock induces termination defects | 8 |
|  A loss in 3’ end cleavage occurs at many genes with heat shock-induced t termination defects | 10 |
| Readthrough transcripts show nuclear enrichment | 11 |
| Readthrough transcripts are less stable | 15 |
| **Discussion** | 19 |
| **Materials and Methods** | 22 |
|  Cell culture | 22 |
|  Heat shock protocol | 22 |
|  RNA extraction and RT-qPCR analysis | 22 |
|  Identifying heat shock induced termination defects | 23 |
|  Cellular localization | 25 |

|  |  |
| --- | --- |
|  RNA stability Table of sequences of primers used in RT-qPCR Acknowledgements References | 26273031 |
|  |  |

|  |  |
| --- | --- |
| Schematic of the central dogma | 2 |
| Schematic of a gene | 3 |
| Schematic of pre-mRNA processing to yield mRNA | 4 |
| Schematic of pre-mRNA cleavage and Pol II displacement | 5 |
| Heat shock induces RNA downstream of the 3’ end of a gene | 7 |
| Many genes with heat shock induced termination defects show a loss in 3’ cleavage | 9 |
| Cells can be efficiently fractionated | 11 |
| RNA produced from heat shock-induced termination defects show nuclear retention | 13 |
| Readthrough transcripts are less stableSchematic of 3’ cleaved and uncleaved readthrough transcript | 1520 |

**Figures**

**Abstract**

Mammalian transcription is a tightly controlled process for mRNA production. Cellular stress has been shown to impact several steps in transcription. For example, heat shock induces termination defects, causing dysfunctional termination of mRNA synthesis that may interfere with downstream genes and alter protein synthesis. It is unclear how malfunctioning termination machinery permits Pol II to continue transcribing past the normal 3’ end of mRNA genes, or how the RNA produced after failed termination impacts the cell. This is a pervasive biological phenomenon with unknown implications and is likely a fundamental biological process that will advance our understanding of cellular biology. This project aimed to investigate RNA produced during stress induced-termination defects through characterizing the 3’ cleavage that releases mRNA from RNA polymerase II (Pol II), the cellular localization of mis-terminated RNAs, and the stability of these RNA molecules. We found that RNA produced upon stress-induced termination defects can be characterized by having a general loss of 3’ cleavage, nuclear retention, and decreased stability compared to mRNA in unstressed cells. These findings provide insight into the mechanisms that give rise to termination defects, and the cellular fate of the transcripts that result.

**Introduction**

**The central dogma of cellular biology**

The central dogma is the process by which genetic information is transferred between three biomolecules: from DNA to RNA to proteins1 (Figure 1). Proteins can then go on to carry out cellular functions2. DNA is composed of 4 nitrogenous bases that are polymerized to form two paired DNA strands each with a phosphodiester backbone3. These bases encode genes that specify the sequences of RNA and proteins, and regulatory sequences that are important for directing DNA replication and transcription using recognition sites for DNA-binding proteins3. In a process called transcription, the DNA template strand is transcribed into messenger RNA (mRNA)3. In a separate process, this RNA gets spliced, polyadenylated, and exported from the nucleus to the cytoplasm. Once in the cytoplasm, the information encoded in mRNA is used in protein synthesis by ribosomes in a process called translation. The bases of RNA code for amino acids that are building blocks for proteins. Proteins largely influence the cell through a wide variety of functions. It is crucial for each step to be highly regulated and controlled to ensure the cell remains in a homeostatic state4 and responds to changes or stresses in the cell and from the environment.

**Figure 1 Schematic of the central dogma.** DNA can undergo replication for cellular division. DNA can also undergo transcription to encode an RNA copy of one strand of DNA. RNA undergoes translation, where the RNA code is used to synthesize proteins that will serve cellular functions. Figure created by BioRender.

A schematic of a protein-coding gene is shown in Figure 2. The core promoter of a gene is a DNA sequence that is used to initiate transcription (green bar). It is found directly upstream of the gene. The body of the gene contains exons (dark blue bars) and introns (light blue bars). Exons are coding sequences while introns are non-coding sequences, which are removed in a process called splicing5. Nearing the end of the gene exists the sequences that signal for transcription termination, a process in which the transcription machinery and mRNA ultimately detach from the DNA. The end of a gene is marked 3’ end in the figure, and all DNA past the 3’ end is termed to be downstream of the gene (yellow bar).

**Figure 2 Schematic of a gene.** A gene consists of the gene body containing introns (light blue) and exons (dark blue), and the 3’ end of the gene which defines the downstream region of the gene (yellow).

**Transcription and pre-mRNA processing**

Controlling transcription is the key process that regulates most gene expression. Misregulation of transcription and gene expression can lead to human disease and cancer6. Transcription results in the synthesis of mRNA from one strand of DNA, during which ribonucleotide analogs to deoxyribonucleotides are added to a growing chain of nascent mRNA. The primary enzyme that executes this function is RNA polymerase II (Pol II). However, Pol II cannot synthesize mRNA alone. It requires many general transcription factors (GTFs) that act on most genes and regulatory transcription factors (TFs) that are gene-specific7. Accessory factors help recruit Pol II and assist in forming the initiation complex6,8, aid processive elongation7, and control termination9–11.

Transcription occurs in the nucleus in multiple steps, including initiation, elongation, and termination8. Beginning with initiation, a set of GTFs facilitate promoter recognition and the unwinding of promoter DNA and initiation6,8,12. Recruitment of regulatory TFs allows for the recruitment and binding of additional TFs as well as Pol II to the DNA13. Elongation is the step where Pol II, assisted by elongation factors, transcribes along the body of the gene. Transcription ends with termination, where the nascent mRNA is cleaved to be released from Pol II, polyadenylated, and exported to the cytoplasm. During termination, Pol II is also displaced from the DNA strand8.

In addition to the role of TF proteins, a crucial method of transcription regulation is chromatin condensation6,13. DNA is wrapped around histone octamers in nucleosomes that come together to form higher-order structures called chromatin. The organization of chromatin determines which genes are exposed for transcription4,6–8,12–14. Chromatin organization plays a crucial role in regulating transcription during cellular differentiation through facilitating unique gene expression programs of each cell type2-4,8,10-12.

mRNA prior to being processed is referred to as pre-mRNA. This RNA undergoes multiple processing events before it can be exported into the cytoplasm, including 5’ end capping and splicing, which occur co-transcriptionally, and the addition of a poly adenosine monophosphate (poly(A)) tail, which occurs after 3’ cleavage5,7,14,15 (Fig 3). Processing is essential because it protects RNA from degradation while creating a functional RNA molecule that can be translated into a protein. The first processing event is 5’ end capping, a reaction that is highly coupled to transcription14,15. This process entails the addition of a modified guanine nucleotide to the 5’ end of the RNA molecule as it is being transcribed, thereby minimizing mRNA degradation and permitting its interaction with the ribosome in the cytoplasm. Another processing step is the removal of introns through splicing by a complex called the spliceosome14. This occurs because only exons contain the genetic information to encode proteins. Splicing also occurs co-transcriptionally. The last processing event is the addition of a poly(A) tail4,15,16, which is done by the poly(A) polymerase after the cleavage step in termination. This entails the addition of dozens or hundreds of adenine nucleotides to the 3’ end of the RNA. The length of the poly(A) tail varies per mRNA, however, it is on average 200 nucleotides16. This contributes to the increased stability of the molecule and improved translation15. Once the mRNA has been fully processed it can be exported into the cytoplasm for translation.

**Figure 3 Schematic of pre-mRNA processing to yield mRNA.** Pre-mRNA contains exons (purple and pink blocks) and introns (gray lines) Upon pre-mRNA processing, a 5’ cap is added, the introns are spliced out to leave only the exons, and a poly(A) tail is added.

**Transcription termination**

To initiate termination, Pol II and the transcribed nascent RNA must undergo many steps for proper termination. As Pol II transcribes through the end of the gene, known as the 3’ end, it passes the polyadenylation signal (PAS). The PAS is a T-rich DNA sequence that is transcribed to produce an A-rich RNA sequence, termed the polyadenylation signal. This is an RNA recognition element for binding termination machinery for post-transcriptional processing. One subunit of the termination machinery is an endonuclease called cleavage and polyadenylation specificity factor (CPSF73). As seen in Figure 4 CPSF73 cleaves the nascent RNA, resulting in a 3’ end of mRNA that undergoes polyadenylation and a 5’ end of RNA that is still associated with Pol II4,15. To remove this remaining nascent RNA, Xrn2, an exonuclease, digests the attached RNA until it reaches Pol II and then assists in the displacement of Pol II from the DNA template strand8,16. Therefore, the mRNA has a defined 3’ end sequence dictated by the site of cleavage, but Pol II is displaced from the genome at random locations downstream of the gene. Although the exact mechanism is unknown, during termination there are factors that cause kinetic pausing of Pol II and allow for the termination machinery to attach itself to the Pol II and nascent mRNA to aid in termination4,10.

**Figure 4 Schematic of pre-mRNA cleavage and Pol II displacement.** Nascent RNA (red) is cleavage by CPSF73 (scissors). Pol II is displaced from the DNA by Xrn2 (green). The purple oval represents termination machinery.

**Cellular stress induces termination defects**

Transcriptional regulation is a highly dynamic process that adapts to cellular abnormalities. Cellular stress can arise from many external factors, such as heat shock10,17,18, oxidative stress19,20, viral stress21–23, and osmotic stress24,25. Under stress, the cell generally shows global repression of transcription of thousands of genes while hundreds of genes show activation10. More recent literature has shown that in response to cellular stress, Pol II occupancy and active mRNA transcription are observed in kilobases downstream of the 3’ end of genes7,10. This observed phenomenon is termed a termination defect. The mechanism is unknown. It is likely the defect occurs when the termination machinery loses its ability to execute all its functions, for example, kinetic pausing of Pol II, mRNA cleavage, and Pol II displacement from the genome, resulting in aberrant mRNA transcripts.

Termination defects are a phenomenon that is observed only in cells that are subjected to stress or some cancerous cells. Having a termination defect impacts the characteristics of the RNA produced, as well as the protein products of translation. Several studies have been made to characterize the mRNA transcribed during the defect. These RNAs show strong nuclear retention10,11,21,24,26, which is thought to be primarily due to dysfunctional pre-mRNA processing. A severe termination defect and/or close neighboring genes can result in Pol II reading into neighboring downstream genes10 and producing chimeric RNA transcripts with more than one coding sequence and intergenic RNA. These chimeric RNAs have been observed to produce aberrant intergenic splicing events21,24,27, which have been identified in cancerous cells27,28. These aberrations often prevent mRNA translation and protein synthesis10,11. Regarding the stability of these mRNAs, one study did not observe a change in half-life upon osmotic stress29.

Another aspect of mRNA produced during stress-induced termination defects is a change to 3’ cleavage9–11. One study reported a portion of these RNA transcripts do undergo correct 3’ end formation (i.e. cleavage) and are translatable. The same study speculated a loss of 3’ cleavage due to lower levels of CPSF73 recruitment10. The effects of termination defects on a larger scale show chromatin accessibility downstream of the gene matches readthrough regions9,16,21,26,27. In other words, the chromatin downstream of genes with defects is open and accessible, showing this region is transcribed. A primary speculation of the biological significance of termination defects has been supporting chromatin reorganization downstream of genes10,21,24,27,29.

It is unclear how the malfunctioning termination machinery permits Pol II to continue transcribing downstream of genes in response to stress, or the cellular consequences of transcription past the 3’ end. There is a gap in the knowledge of what mechanistically fails during termination, such as the displacement of Pol II from the DNA. Moreover, determining the fate of RNA produced during termination defects will shed light on its consequences on cellular function. The metabolism of these RNA molecules is unclear, such as their 3’ cleavage, where they are located in the cell, and how the stability of these molecules differs from RNA produced under normal termination. This project investigated termination defects using heat shock as the representative cellular stress. The data provide insight into the mechanism of termination defects, and characteristics of the RNA produced by termination defects, such as the 3’ cleavage of the molecule, the cellular localization, and stability of these molecules.

**Results**

**Heat shock induces termination defects**

To investigate mechanisms of stress-induced termination defects, I established an experimental system to detect and quantify read-through transcription at representative genes after heat shock. To assess transcription past the end of a gene, RT-qPCR was used to detect how much RNA was being transcribed from downstream regions for each gene and was compared before and after heat shock (Fig 5A, see yellow region of the gene schematic). Heat shock changes the overall level of transcription at thousands of genes10. Therefore, to control for changes in downstream transcription being influenced by changes in overall transcription, the levels of RNA from the downstream regions were normalized to the levels in the gene body (blue regions in the gene schematic) to isolate just the change in the downstream RNA after heat shock. To establish this model, HEK293 (human embryonic kidney) cells were used as a model system. Cell culture was used to plate these cells in flasks and subjected to heat shock at 44℃ for 1 hour. The control condition was cells that were not subjected to heat shock, but rather left at normal growth conditions. (Fig 5B). The RNA from both heat shock and no heat shock conditions were extracted from the cells and used in RT-qPCR.

**Figure 5 Heat shock induces RNA downstream of the 3’ end of a gene. (A)** Schematic of a gene, including the gene body containing introns and exons, and the 3’ end of the gene which defines the downstream region of the gene. **(B)** Schematic of experimental setup shows cells were heat shocked and their RNA was extracted and used in RT-qPCR. **(C)** The graph shows the fold change in downstream RNA upon heat shock, normalized to a non-heat shocked control sample and the respective gene body to isolate the transcription change only in the downstream region. The line at 1.0 represents no change in RNA levels after heat shock. \* indicates a p-value < 0.05, as compared to 1.0 using a one-sided t-test.

A majority of genes displayed a termination defect, as seen by a fold change greater than one (Fig 4C). A fold change greater than one shows there was more RNA being produced downstream of genes after heat shock, which defined readthrough transcription. Readthrough transcription suggests an inability to properly terminate transcription upon cellular stress. These representative genes were chosen because they all showed heat shock-induced termination defects in a 4sU-seq experiment performed in the same cell line under the same heat shock conditions (K. Walsh, unpublished data). In the data shown here, STK35, ZFAND5, and SRSF3 showed the most severe defects. Their higher fold change values signify the greatest increase in the quantity of readthrough transcripts after heat shock. ANXA5 and ICE1 showed more moderate defects. NCL strayed from this pattern, as it did not present readthrough transcription. This gene had a fold change of one, signifying that the quantity of RNA detected after heat shock was equal to no heat shock, which does not indicate readthrough transcription. It is unclear why NCL did not display readthrough transcription, despite having shown a termination defect in prior 4sU-seq experiments. All in all, these data show that multiple genes have an increase in downstream RNA after heat shock and establish the experimental system I used to study termination defects after stress.

**A loss in 3’ end cleavage occurs at many genes with heat shock-induced termination defects**

Under normal cellular conditions, actively transcribed RNA is cleaved by CPSF73 at the cleavage site, which is at the 3’ end of the gene. This creates the 3’ end of the genes and is one of the first signals to terminate transcription. To investigate if this cleavage mechanism is impacted by heat shock, uncleaved RNA was quantified by RT-qPCR across the cleavage site (Fig 6A). The forward primer was in the gene body and the reverse primer was downstream of the gene site, therefore the RT-qPCR product is only produced from uncleaved RNA. Most genes tested presented a loss of cleavage upon heat shock, as seen by a fold change of uncleaved RNA above 1.0 (Fig 6B). Specifically, STK35, SRSF3, and NCL, showed a defect in RNA cleavage. The average fold change of ZFAND5 and ICE1 showed an increase in uncleaved RNA, however, there was no statistical significance across replicates. However, different genes had different fold changes of uncleaved RNA, which suggests that the severity in the loss of cleavage varied per gene. An exception to this pattern of readthrough transcripts having a decrease in cleavage was ANXA5. Despite displaying a termination defect (Fig 5C), this gene did not show a change in RNA cleavage upon heat shock. These data suggest that mechanistically there are different categories of termination defects. Termination defects can be defined as having both readthrough transcription and a cleavage defect, or readthrough transcription and normal cleavage. ANXA5 contained readthrough transcription without a loss in cleavage.

**Figure 6 Many genes with heat shock induced termination defects show a loss in 3’ cleavage**. **(A)** shows a schematic of an RT-qPCR product (gray box) spanning the cleavage site that was used to quantify uncleaved RNA. **(B)** The fold change upon heat shock of uncleaved RNA is plotted. The signal was normalized to a non-heat shocked control sample and the gene body to isolate the change only in uncleaved RNA. The line at 1.0 represents no change in RNA levels after heat shock. \* indicates a p-value < 0.05, as compared to 1.0 using a one-sided t-test.

SLC23A2 and SMIM13 were tested as no-defect controls, genes that did not display termination defects upon heat shock in prior 4sU-seq experiments (K. Walsh, unpublished data). A gene with normal termination would be expected to show normal cleavage, and thus show a fold change of 1.0 in uncleaved RNA after heat shock. SLC23A2 expressed a fold change value close to 0.9. SMIM13 possessed a fold change that hovered close to 0.5. This implies that there was half as much uncleaved RNA at SMIM13 upon heat shock, or cleavage of RNA occurs twice as often upon heat shock in this gene. It is not known why there would be enhanced cleavage of SMIM13 after heat shock, but it is consistent with having no termination defect. Further testing would be required to explain the observation of enhanced cleavage.

The overall trend of this data supports the hypothesis that heat shock results in a loss of cleavage at the majority of genes with readthrough transcripts. This is seen by an increased fold change in RNA product across the cleavage site upon heat shock. However, this does not hold true for every gene, which suggests multiple mechanisms for termination defects.

**Readthrough transcripts show nuclear enrichment**

To investigate if 3’ end readthrough after heat shock impacts the ability to process and export transcripts to the cytoplasm, I tested the cellular localization of these RNAs. Readthrough RNA enriched in the cytoplasm would show that the mechanism of processing and exporting RNA from the nucleus to the cytoplasm is maintained even when termination is defective. An accumulation of readthrough RNA in the nucleus would suggest that export mechanisms are lost. Hence, cellular localization offers insight into which mechanistic aspects of RNA metabolism are defective after heat shock-induced termination defects.

**Figure 7 Cells can be efficiently fractionated.**Panel **(A)** describes the experimental setup where cells were heat shocked and fractionated into the nuclei and cytoplasm. RNA from each fraction was extracted and used in RT-qPCR. **(B)** Plotted is the relative enrichment of marker RNAs in either fraction in the absence of heat shock, where a positive log2 value denotes nuclear enrichment and a negative log2 value denotes cytoplasmic enrichment. \* indicates a p-value < 0.05, as compared to 0 using a one-sided t-test.

Readthrough transcripts in nuclear and cytoplasmic fractions of the cell were quantified to investigate cellular localization. The experiment is illustrated in Figure 7A. Cells underwent heat shock at 44℃ for 1 hour and were fractionated into the nuclei and cytoplasm. The RNA from each fraction was extracted to yield RNA found in the nuclei and cytoplasm, which was used in RT-qPCR. To verify the efficacy of the protocol, nuclear and cytoplasmic RNA markers were tested by RT-qPCR. These markers were selected because they are genes that are known to produce RNAs NCL intron. The gene XIST is a historic nuclear marker. NCL intron is an intronic set of primers on the NCL gene. The placement of this primer is crucial because the processing of pre-mRNA includes the splicing out of introns to leave the mRNA only with exons before it gets exported to the cytoplasm. Introns should always be retained in the nucleus, which can be used as a nuclear marker. The fractionation protocol was mostly successful, as seen by the expected localization of the nuclear markers XIST and NCL intron in the nuclear fraction (Fig 7B). In the plot, nuclear enrichment is displayed by a positive value on a log2 scale, while cytoplasmic enrichment is a negative value on a log2 scale. The cytoplasmic markers used were 7SL, 18S rRNA (18S), and RPS19 ex/ex. 7SL and 18S are noncoding RNAs that are involved in translation and highly abundant in the cytoplasm, making them good candidates for cytoplasmic markers. The primers used for the RPS19 gene amplified exonic regions. Upon splicing, mature mRNA containing only exons gets exported to the cytoplasm. The RPS19 ex/ex PCR product was amplified using primers spanning two exons in the RPS19 gene. Unlike 7SL and 18S, the purpose of testing RPS19 was related to the primer placement rather than the function of the gene. Contrary to its expected cytoplasmic enrichment, 7SL showed enrichment in the nucleus across all three replicates. 18S displayed cytoplasmic enrichment as expected (a negative average log2); however, the upper error bar shows a positive log2 value. One of the three replicates showed enrichment of RNA in the nuclear fraction, while the other two showed enrichment in the cytoplasmic fraction. Yet RPS19 ex/ex showed the predicted pattern of enrichment of RNA in the cytoplasm. All together the nuclear markers showed clean enrichment in the nucleus, and the cytoplasmic markers showed total (7SL), partial (18S), or no (RPS19 ex/ex) enrichment in the nucleus across three replicates. This suggests there was some level of cytoplasmic RNA contamination of the nuclear RNA. Hence, the fractionation has limitations that need to be kept in mind when assessing the rest of this data.

 

**Figure 8 RNA produced from heat shock-induced termination defects show nuclear retention (A)** Similar to Fig 6B, plotted is the enrichment of RNA produced downstream (ds) of the 3’ end of the indicated genes. The signal was normalized to a non-heat shock control. \* indicates a p-value < 0.05, as compared to 0 using a one sample t-test. T-test on ICE1 ds samples were not run because of an n of 2. **(B)** An alternative analysis on the same data in panel A. Plotted is the fold change upon heat shock in the nuclear and cytoplasmic fractions independently. The line at 1.0 represents no change in RNA levels after heat shock. \* indicates a p-value < 0.05, as compared to 1.0 using a one-sided t-test. T-test on ICE1 ds samples were not run because of an n of 2.

The protocol yielded sufficient cellular fractionation to evaluate the cellular localization of RNA produced from heat shock induced termination defects. Readthrough transcripts showed strong nuclear localization upon heat shock (Fig 8A). All tested genes displayed a positive log2 value, providing evidence to support a model that these RNAs are retained in the nucleus and are not transported into the cytoplasm. Although the data in Fig. 7B suggests there was cytoplasmic RNA contamination of the nuclear RNA, the levels of downstream RNA found in the nucleus upon heat shock showed high levels of enrichment, suggesting this is unlikely to arise from contamination. This is consistent across all genes and contributes to the reliability of the data.

As an additional means of analysis, I also quantified the fold change in levels of downstream RNA upon heat shock in the nucleus and the cytoplasm (Fig. 8B). The heat shock-induced increase in downstream RNA was located almost entirely in the nucleus compared to the cytoplasm at all genes tested (Fig 8B). This means readthrough transcripts are significantly more nuclear upon heat shock. This conclusion is important because it supports the hypothesis that upon heat shock there is a mechanistic abnormality in termination that prevents proper processing and export of these RNAs. Downstream RNA located in the cytoplasm did not change upon heat shock (Fig 8B). A majority of genes have a fold change nearing a value of 1.0, arguing that the amount of downstream RNA found in the cytoplasm did not change after heat shock. The exception to this trend is ICE1, however, this data is not statistically significant, which allows us to draw the same conclusions. Overall, readthrough transcripts lack cytoplasmic enrichment and show strong nuclear enrichment, suggesting these RNAs are retained in the nucleus as a result of not being processed and exported.

**Readthrough transcripts are less stable**

The abundance of cellular RNAs is often controlled at the level of RNA stability. Accordingly, I investigated the stability of the downstream RNA produced from heat shock-induced termination defects, and how that compared to the stability of the coding mRNA produced from the upstream

 **Figure 9 Readthrough transcripts are less stable. (A)** Experimental setup where cells were heat shocked and treated with Actinomycin D (Act D). A non-heat shock condition was also treated with Act D. RNA was extracted from each condition and used in RT-qPCR. Signal was normalized to the non-Act D treatment controls. **(B)** The plot represents the fraction of RNA remaining after 2 hours of Act D treatment. Tested was the gene body of the indicated genes in non-heat shocked (NHS) and the downstream region of genes in heat shocked cells (HS ds). All data were normalized to the non-Act D control samples and 18S rRNA. T-tests were not run due to an n of 2 for all samples.

gene. To investigate the stability of readthrough transcripts, the experiment diagrammed in Fig. 9A was performed. Cells were heat shocked to produce termination defects (indicated by the first set of brackets) and control cells were left at a normal growth temperature. Half the heat shocked and non-heat shocked samples were treated with Actinomycin D (Act D) for 2 hours (purple flasks), while the remainder were left untreated (orange flasks). RNA was extracted from all cells to give rise to the four experimental conditions shown in Fig. 9A. Act D is a transcription inhibitor that blocks all new RNA synthesis, thus the level of a transcript remaining after Act D treatment is a reflection of its stability/decay rate. It should be noted that the protocol was done in triplicate, but RT-qPCR analysis showed that one replicate did not display a termination defect at the tested genes, suggesting these cells did not respond to the heat shock. Since the biological phenomenon that was being investigated was not observed, this replicate was not included in the analysis. The following analysis was done with the other two biological replicates.

To analyze RNA stability, I determined the fraction of RNA that remained after 2 hours of Act D treatment before and after heat shock, using different sets of primers. Theoretically, the highest fold change should be a value of 1.0. This signifies that the fraction of RNA remaining is 100%, or that no RNA decay was observed over 2 hours. As the RNA is being degraded the fold change should decrease. Fig. 8B shows the fraction of RNA remaining for 2 sets of conditions/primer pairs: 1) non-heat shock, gene body primers (gray bars) 2) heat shock, downstream primers (pink).

Most of the non-heat shock gene body bars (gray) hover around 1.0, indicating these mRNAs were stable over 2 hours. For some genes, such as ANXA5, SRSF3, INPPL1, the fold change after Act D treatment is slightly greater than one. This indicates there was more RNA after 2 hours of Act D treatment. Because it is unlikely that Act D failed at specific genes, it is assumed that the increased fold change reflects technical variability more than a biological response. The no heat shock gene bar of STK35 and ZFAND5 have a fold change nearing a value of 0.85. This implies after two hours of Act D treatment, 85% of the RNA for these genes remains, suggesting these genes are relatively stable.

We found that the RNA detected by downstream primers after heat shock was less stable, as seen by the decrease in the fraction of RNA remaining compared to the no heat shock condition (Fig 9B). However, these readthrough RNAs do remain relatively stable. The fraction of RNA remaining was always over 0.5. This means that after 2 hours, more than half of these RNA were not yet degraded. However, the fraction varies per gene. Most readthrough RNAs showed a smaller change in stability upon heat shock compared to the proper mRNA present in non-heat shocked cells. The exception to this is ANXA5 and INPPL1, where the reduction in stability was significantly larger. What can be concluded from this data is that upon heat shock, RNA produced past the 3’ end of a gene is less stable than in the mRNA produced in the absence of heat shock, however, the amount of degradation is possibly gene specific. Models that could give rise to the variability are described in the Discussion section.

**Discussion**

This project aimed to investigate mRNA produced during heat shock-induced termination defects by characterizing nascent RNA cleavage to separate the mRNA from Pol II, cellular localization, and stability. This was primarily done through total RNA extraction from heat shocked and non-heat shocked cells, and quantification of specific RNAs by RT-qPCR. This data supports that RNA produced upon stress-induced termination defects can be characterized by having a general loss of 3’ cleavage, nuclear retention, and decreased stability. These findings provide insight into the mechanisms that give rise to termination defects, and the cellular fate of the transcripts that result.

In investigating the cleavage that releases pre-mRNA from Pol II, we found a loss of 3’ cleavage in many, but not all, genes (see Fig 5C). Our findings are consistent with related literature10,11,29 and contribute to a growing understanding of how stress-induced termination defects occur. In the absence of stress, functional termination involves cleavage of the nascent RNA to generate the 3’ end of pre-mRNA, and limited transcription downstream of genes as Pol II is displaced from the DNA by Xrn2 (see Fig 4). Our data suggests that mechanistically there are two categories of termination defects that differentially impact cleavage and give rise to readthrough transcription. The first category is the loss of 3’ cleavage and the presence of readthrough transcription. The second category is the retention of 3’ cleavage with the presence of readthrough transcription. Fig 5C showed that most tested genes fall into the first category, a loss in cleavage and the presence of readthrough transcription after heat shock. This could manifest from CPSF73 failing to cleave nascent RNA either through a failure of recruitment or maintained recruitment but a loss in function. Without cleavage, the Xrn2 exonuclease has nothing to grab onto, and cannot displace Pol II from the genome, thereby allowing transcription to continue for kilobases downstream. ANXA5, a gene that strays from this pattern, fits into the second category. The 3’ cleavage was retained after heat shock while readthrough transcription was observed. The retention of cleavage suggests that CPSF73 is functioning, but the inability of Pol II to stop transcribing suggests a lack of displacement from the genome. It is possible that Xrn2, the enzyme responsible for this function, cannot perform its function at some genes either due to a loss of recruitment or loss of exonuclease activity in response to heat shock. It can be concluded from this data that not all genes undergo the same mechanism of termination defects. However, this uncovers many more questions. It is unclear if there is a pattern between types of genes and the mechanisms they partake in. Perhaps the proximity of neighboring genes and the architecture of chromatin plays a role in a potential pattern10,21,24,27,29.

In investigating cellular localization, we found that mRNAs produced during heat shock-induced termination defects are almost always located in the nucleus. Related literature also observed nuclear retention of these mRNAs10,11,21,26. It is possible that lack of 3’ cleavage leads to nuclear retention. Before being exported to the cytoplasm, mRNA must be cleaved from Pol II and undergo the addition of a poly(A) tail. A loss of cleavage at the normal site, as seen in our data for many genes, would prevent normal polyadenylation and could result in nuclear retention. However, the population of mRNA that undergoes correct cleavage during stress is thought to be translatable9. The downstream region stays associated with Pol II in the nucleus, but it is unclear whether these transcripts are polyadenylated.

 In investigating the stability of RNA produced downstream of the 3’ end of the gene, we found that readthrough transcripts are generally less stable than RNA produced in the absence of heat shock. This aspect of termination defects has not been well established. One study showed no difference in stability upon osmotic stress29. Genes intrinsically differ in their mRNA stability and different types of cellular stress will have varying impacts on mRNA stability. However, using this method it is difficult to distinguish whether the instability is a result of heat shock, or the RNA produced downstream of the gene. The simplest model is that the RNA produced downstream of the gene is one population consisting of uncleaved RNA due to our observations of a loss of 3’ cleavage at almost all genes with readthrough after heat shock (Fig. 6B). However, it is not guaranteed that all readthrough transcripts are uncleaved. We could be detecting a mixed population of cleaved and uncleaved RNA (Fig 10). This poses an issue for clear interpretation because the stability of 3’ cleaved and uncleaved RNA produced downstream of the 3’ peak could vary.

**Figure 10 Schematic of 3’ cleaved and uncleaved readthrough transcript.** The top shows 3’ uncleaved transcripts. The bottom shows a properly 3’ cleaved transcript, yielding a separate RNA molecule, the body of the gene (light pink) and the downstream of the gene (dark pink).

Further research could continue to shed light onto the mechanism of stress-induced termination defects. It is unknown whether the RNA produced downstream of the gene eventually undergoes cleavage and polyadenylation. This could provide insight into the stability and the extent of processing these RNAs undergo. This could be done by isolating an oligo(dT) column that binds the poly(A) tail and probing for readthrough RNA30. To investigate a change in recruitment of termination factors under cellular stress, techniques such as Crosslinking Immunoprecipitation sequencing31 and Chromatin Immunoprecipitation sequencing32 can be used to monitor proteins such as CPSF73 and Xrn2. In determining the mechanism of termination defects and the metabolism of the resulting RNA, insight could be offered onto the biological significance of stress induced-termination defects.

**Methods**

**Cell culture**

Human Embryonic Kidney (HEK) 293 cells were grown in Dulbecco's Modified Eagle Medium containing 10% fetal bovine serum and 1% penicillin/streptomycin. The flasks were kept in an incubator that maintained a temperature of 37℃ and at 5% CO2 concentration.

**Heat shock protocol**

Approximately 24 hours after plating cells, the confluency of the flasks were verified and recorded. The flasks were placed in a water bath set to 44℃ for 1 hour. During this time the control conditions remained in the cell culture incubator.

**RNA extraction and RT-qPCR analysis**

TRIzol was added directly to cells after aspirating off media, 1 mL per T25 or 3 mL per T75. Alternatively, cells that were trypsinized first were pelleted via centrifugation and 1 mL of TRIzol was added per 3 million cells. The flasks containing TRIzol were left to incubate for 5 minutes rocking on a bioshaker at room temperature. Samples were transferred to tubes and 200 μL of chloroform was added per 1 mL of TRIzol. To ensure thorough mixing, tubes were vigorously shaken by hand for 15 seconds before being left to incubate for 2 minutes at room temperature. At the end of this incubation, the tubes were placed in a centrifuge at 4℃ and spun down at 12,000 x g for 15 minutes. Once the tubes were removed from the centrifuge, the aqueous upper phase, containing the RNA, was combined to a new 15 mL centrifuge tube. To precipitate the RNA, 500 μL of isopropanol was added per 1 mL of the starting volume of TRIzol and transferred to 1.5 mL Eppendorf tubes. The tubes were inverted to mix and left to incubate at room temperature for 10 minutes, then spun down at 4℃ at 12,000 x g for 10 minutes. Pellets were washed with 1 mL of 75% ethanol per tube and spun down at 7,500 x g for 5 minutes. The supernatant was removed, and the tubes were centrifuged again at 7,500 x g for 1 minute and any remaining supernatant was discarded. The RNA pellet was resuspended in 15 μL of nuclease free water and concentration was determined by nanodrop.

After the RNA extraction, an RT-qPCR was run to quantify the amount of RNA detected for transcripts arising from specific regions of the genome using the primers in Table 1. All RNA samples were diluted to 5 ng/μL, with the exception of RNA used for the 18S primer. This RNA was diluted to 0.01 ng/μL. RT-qPCR was run using the Luna Universal One-Step RT-qPCR kit and a real-time PCR machine.

**Identifying heat shock induced termination defects**

In establishing termination defects and determining changes to 3’ end cleavage, the two conditions tested were heat shock and no heat shock. Each condition consisted of 5 million cells plated in a T75 flask. The heat shock flask underwent heat shock to induce termination defects. The media of both conditions was aspirated and TRIzol was added in the flask for RNA extraction. This process was done in biological triplicate. The same RNA that was used to establish termination defects was also used to determine changes to 3’ end cleavage. The total cellular RNA was then assessed by RT-qPCR.

RT-qPCR analysis for establishing termination defects

Six representative genes that have previously been found to show termination defects after heat shock were tested with RT-qPCR. Each gene was tested in the heat shock and no heat shock conditions to establish the presence of termination defects. RNA transcribed from the gene body and downstream (~5-10 kb) of each gene was tested using primers designed for these regions. To analyze the data, the first step was to average the CT values of the two technical replicate wells in the RT-qPCR plate. For both heat shock and no heat shock condition, the signal downstream of the gene was normalized to the signal in the gene body by taking a delta CT (dCT). This was done by subtracting the two CT values (downstream minus gene body). The next step was to normalize the heat shock condition to the no heat shock condition by taking a delta delta CT (ddCT, heat shock minus no heat shock). The ddCT was then converted to a fold change using the equation: 2^-(ddCT). The average was plotted, and the standard deviation was determined for plotting error bars.

RT-qPCR analysis for investigating cleavage

The same six representative genes that were used in establishing termination defects were also used for investigating 3’ end cleavage. However, the regions of the genes amplified by RT-qPCR were the gene body and the cleavage site. An average of the technical replicates for each sample was taken. The cleavage site signal was normalized to the gene body signal independently in the heat shock condition and no heat shock condition to obtain dCT. Then, the heat shock dCT was normalized to the no heat shock dCT by taking a ddCT (heat shock dCT minus no heat shock dCT). The ddCT was then converted to a fold change using the equation: 2^-(ddCT). The average was plotted, and the standard deviation was determined for plotting error bars.

**Cellular localization**

To test cellular localization, a heat shock and a no heat shock condition was used. For each condition 5 million cells in T75 flasks were plated. At the end of the hour-long heat shock, cells were trypsinized, and pelleted. To separate nuclei from cytoplasm, the cells were resuspended in Nuclear Run-On buffer. This buffer contained 10 mM TRIS pH 7.9, 10 mM NaCl, 4 mM MgCl2, 0.5% NP-40, 0.4 mM PMSF, and 1x Protease Inhibitor. The cells were incubated with the NRO buffer for 5 minutes on ice, spun down at 0.8 RCF for 10 minutes. The pellet contained nuclear RNA, and the supernatant contained cytoplasmic RNA. This wash was repeated, and RNA was extracted with theTRIzol protocol. Once nuclei and cytoplasm were isolated, the RNA from each was extracted using the TRIzol protocol. This cellular localization protocol was done in triplicate.

RT-qPCR Analysis for cellular localization

Two sets of genes were tested in the RT-qPCR, the representative genes used to test termination defects, as well as nuclear and cytoplasmic markers to validate the efficacy of the cellular localization protocol.

An average of the technical replicates for each sample was taken. Two different methods of analysis were done on the raw data. One method depicted the relative enrichment of RNA detected in the nuclear versus cytoplasmic fraction on a log2 scale. A dCT was taken between the nuclear fraction and the cytoplasmic fraction for each primer for heat shock and no heat shock independently (nuclear minus cytoplasmic). The fold change was taken of each dCT value using the equation: 2^(dCT). Then, the log2 of the fold change was determined. This allowed positive numbers to reflect nuclear enrichment and negative numbers to reflect cytoplasmic enrichment with equal magnitude scales. Marker genes and representative defect genes were plotted separately and the standard deviation of the log2 values were taken for error bars.

The second method of analysis was to plot the fold change of the nuclear and cytoplasmic RNA levels upon heat shock normalized to no heat shock. This analysis was performed only on the 6 representative defect genes. A dCT was taken between the downstream region and the gene body of the nuclear and cytoplasmic RNA independently. The ddCTs were subtracted to normalize heat shock to no heat shock independently for each fraction. The fold change of the ddCT values were taken using the equation: 2^-(ddCT). The average of the biological replicates was plotted, and the standard deviation of all fold changes was calculated for error bars.

**RNA stability**

To investigate RNA stability four conditions were used, + and - Actinomycin D (Act D) in both heat shock and no heat shock conditions. All conditions were plated in T25 flasks with 1.5 million cells per flask. Both heat shock conditions, + and -Act D, were placed in the water bath for heat shock. At the end of the heat shock, the +Act D heat shock and +Act D no heat shock flasks were treated with 5μg/mL Act D. The flasks were placed in the cell culture incubator at 37℃ for 2 hours. During this time, the growth media of the -Act D treatment flasks, which were treated as the time point 0 conditions, was aspirated and treated with TRIzol. This was repeated for the +Act D conditions at the end of the 2-hour incubation.

RT-qPCR analysis of RNA stability

The gene body and downstream region of six representative genes were tested with RT-qPCR. In addition to these genes, 18S rRNA, a stable housekeeping RNA, was also tested. Both the gene body and downstream regions of the gene signals were normalized to 18S rRNA signal by taking a dCT for heat shock and no heat shock independently (gene minus 18S). A ddCT of each region was taken to normalize heat shock to no heat shock (dCT heat shock minus dCT no heat shock). The ddCT was then converted to a fold change using the equation: 2^-(ddCT). An average of the biological replicates was plotted, and the standard deviation of all fold changes was calculated for error bars.

**Table 1**. Sequences of primers used in RT-qPCR.

|  |  |
| --- | --- |
| Primer | Sequence (5' → 3') |
| STK35 gene body Forward | GGTGGAGACCTGAATCAGTATG |
| STK35 gene body Reverse | TTTGTGCAGGAAGGCAATG |
| STK35 downstream Forward | GATGTGGGAAGAGGCAGTATC |
| STK35 downstream Reverse | TCTCAGGATCACACAGAAAGT |
| STK35 cleavage site Forward | TTAAGGGACGTTTGTAATAAAG |
| STK35 cleavage site Reverse | TTGCCACCCTCCCAAAC |
| ANXA5 gene body Forward | GCCTTCCTTCAGCACCTTTA |
| ANXA5 gene body Reverse | TCTATGACGTGTATGTGTTGGTC |
| ANXA5 downstream Forward | TATGCAGCCAACAGACCTATG |
| ANXA5 downstream Reverse | CTGCGTCAGATGGTATCTCATT |
| ANXA5 cleavage site Forward | CCCTGAATTATGTGTACATGTGTG |
| ANXA5 cleavage site Reverse | TCTCAACTAACACCACTGTCAA |
| ZFAND5 gene body Forward | GACTTCTTAGTGGTATCCAGCAC |
| ZFAND5 gene body Reverse | TGCAGCCCATGCAATACA |
| ZFAND5 downstream Forward | CCTGGTTTACATCTGTGGGAAG |
| ZFAND5 downstream Reverse | GGAAGATGCCCAGTGTCAAATA |
| ZFAND5 cleavage site Forward | CTGTTACAGCCATAGAAGTAAAGTT |
| ZFAND5 cleavage site Reverse | CACTTAGCAGTTGCAGACAAA |
| SRSF3 gene body Forward | ACATGGCTGTTCGTGACATTC |
| SRSF3 gene body Reverse | GTACCCTTAAACTGGCAGGAC |
| SRSF3 downstream Forward | ACTGATCCTTGAGCTTTATGTATCAG |
| SRSF3 downstream Reverse | GCCTCCCTCAAGTGGTTAG |
| SRSF3 cleavage site Forward | TCAAATCGACAGAGAAAGCAGGA |
| SRSF3 cleavage site Reverse | TGTCCATGCAACTTTCCCCT |
| NCL exon Forward | GGGCTGGGCCAAACCTAA |
| NCL exon Reverse | TCCTCGGCCTCCTCTAC |
| NCL intron Forward | ACCCATTTCCCTTGGTTTCCT |
| NCL intron Reverse | CACTGTAGCTGATTTAATTCCCACA |
| NCL downstream Forward | AATTGACAAGATGGGCCGG |
| NCL downstream Reverse | CTCGAACTCCAGACCTTGTGAT |
| NCL cleavage site Forward | GAGGCAAGCCTAAGGACAAA |
| NCL cleavage site Reverse | AGCTAACAGACCTAATCCACAAC |
| ICE1 gene body Forward | GACACAGGATACCTCCCAAAG |
| ICE1 gene body Reverse | GCTAAATGCCTCACTCTCTACC |
| ICE1 downstream Forward | CCCGAGTGAGCAAAGAGAAA |
| ICE1 downstream Reverse | CCTGGGAAGCCAGGCATAG |
| ICE1 cleavage site Forward | AGCCATGGATATTACTGTTCTGAT |
| ICE1 cleavage site Reverse | GGTTAGATGTCTTCTTCACCCTT |
| INPPL1 gene body Forward | GTACCCAAATGGACCCCCAC |
| INPPL1 gene body Reverse | TTGCCATGGTTACCGAGGAC |
| INPPL1 downstream Forward | CCTGAACTGATCACTGGGAAG |
| INPPL1 downstream Reverse | AACACCTGCAGTCACATTCT |
| INPPL1 cleavage site Forward | ACAGAGGGATTCTCTTTCCCTAA |
| INPPL1 cleavage site Reverse | TTGCGGCGTGATGTCTTCAA |
| RPS19 ex/ex Forward | CTGACACCTCAGGGACAAAG |
| RPS19 ex/ex Reverse | CCAGACCAGGATTACGAATGA |
| XIST Forward | GTTGGCACTTTCTCTGCATTC |
| XIST Reverse | CCTGGCATTTGGCACTTTAC |

**Acknowledgements**

I would like to express my gratitude to Dr. James Goodrich and Dr. Jennifer Kugel for allowing me to be a part of their lab and for their kindness throughout the years. I would also like to thank the members of the Goodrich/Kugel Lab, Calvin Voong, Stephen Archuleta, Shamira Gonzalez, Joseline Aquino, Lisa Brady, and Ethan Johnson for creating a supportive and enjoyable environment. I truly believe I could not have been a part of a better lab. Most importantly, thank you to Katy Walsh, who guided me throughout this project. I would like to thank CU Boulder’s Biological Sciences Initiative, Undergraduate Research Opportunity, and Biochemistry Undergraduate Research Scholar Award for funding my research. None of this could have been possible without the support of everyone listed here.

**References**

1. Camacho, M. P. Beyond descriptive accuracy: The central dogma of molecular biology in scientific practice. *Stud. Hist. Philos. Sci.* **86**, 20–26 (2021).

2. Eisenberg, D., Marcotte, E. M., Xenarios, I. & Yeates, T. O. Protein function in the post-genomic era. *Nature* **405**, 823–826 (2000).

3. DNA structure and function - Travers - 2015 - The FEBS Journal - Wiley Online Library. https://febs.onlinelibrary.wiley.com/doi/full/10.1111/febs.13307.

4. Kuehner, J. N., Pearson, E. L. & Moore, C. Unravelling the means to an end: RNA polymerase II transcription termination. *Nature reviews. Molecular cell biology* **12**, 283 (2011).

5. Zhang, Y., Sun, Y., Shi, Y., Walz, T. & Tong, L. Structural insights into the human pre-mRNA 3′-end processing machinery. *Molecular cell* **77**, 800 (2019).

6. Casamassimi, A., Ciccodicola, A. & Rienzo, M. Transcriptional Regulation and Its Misregulation in Human Diseases. *Int. J. Mol. Sci.* **24**, 8640 (2023).

7. Evolution of Transcriptional Regulation in Eukaryotes | Molecular Biology and Evolution | Oxford Academic. https://academic.oup.com/mbe/article/20/9/1377/976747.

8. Boeger, H. *et al.* Structural basis of eukaryotic gene transcription. *FEBS Letters* **579**, 899–903 (2005).

9. Protein function in the post-genomic era | Nature. https://www.nature.com/articles/35015694.

10. Cardiello, J. F., Goodrich, J. A. & Kugel, J. F. Heat Shock Causes a Reversible Increase in RNA Polymerase II Occupancy Downstream of mRNA Genes, Consistent with a Global Loss in Transcriptional Termination. *Mol Cell Biol* **38**, e00181-18 (2018).

11. Genotoxic stress impacts pre‐mRNA 3′‐end processing - Biswas - 2024 - BioEssays - Wiley Online Library. https://onlinelibrary.wiley.com/doi/full/10.1002/bies.202400037.

12. Kornberg, R. D. The molecular basis of eukaryotic transcription. *Proceedings of the National Academy of Sciences* **104**, 12955–12961 (2007).

13. Nucleosome-Driven Transcription Factor Binding and Gene Regulation: Molecular Cell. https://www.cell.com/molecular-cell/fulltext/S1097-2765(12)00896-9.

14. Moore, M. J. & Proudfoot, N. J. Pre-mRNA Processing Reaches Back toTranscription and Ahead to Translation. *Cell* **136**, 688–700 (2009).

15. Cramer, P. *et al.* Coordination between transcription and pre-mRNA processing. *FEBS Letters* **498**, 179–182 (2001).

16. Passmore, L. A. & Coller, J. Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression. *Nature reviews. Molecular cell biology* **23**, 93 (2021).

17. Richter, K., Haslbeck, M. & Buchner, J. The Heat Shock Response: Life on the Verge of Death. *Molecular Cell* **40**, 253–266 (2010).

18. Velichko, A. K., Markova, E. N., Petrova, N. V., Razin, S. V. & Kantidze, O. L. Mechanisms of heat shock response in mammals. *Cell. Mol. Life Sci.* **70**, 4229–4241 (2013).

19. Sies, H. Oxidative stress: oxidants and antioxidants. *Experimental Physiology* **82**, 291–295 (1997).

20. Vilborg, A. *et al.* Comparative analysis reveals genomic features of stress-induced transcriptional readthrough. *Proc Natl Acad Sci U S A* **114**, E8362–E8371 (2017).

21. HSV-1-induced disruption of transcription termination resembles a cellular stress response but selectively increases chromatin accessibility downstream of genes - PMC. https://pmc.ncbi.nlm.nih.gov/articles/PMC5886697/.

22. Widespread disruption of host transcription termination in HSV-1 infection | Nature Communications. https://www.nature.com/articles/ncomms8126.

23. Wang, X. *et al.* Herpes simplex virus blocks host transcription termination via the bimodal activities of ICP27. *Nat Commun* **11**, 293 (2020).

24. Alpert, T., Straube, K., Oesterreich, F. C. & Neugebauer, K. M. Widespread Transcriptional Readthrough Caused by Nab2 Depletion Leads to Chimeric Transcripts with Retained Introns. *Cell reports* **33**, 108324 (2020).

25. Rosa-Mercado, N. A. *et al.* Hyperosmotic stress alters the RNA polymerase II interactome and induces readthrough transcription despite widespread transcriptional repression. *Mol Cell* **81**, 502-513.e4 (2021).

26. Rodríguez-Molina, J. B., West, S. & Passmore, L. A. Knowing when to stop: Transcription termination on protein-coding genes by eukaryotic RNAPII. *Molecular cell* **83**, 404 (2023).

27. It’s a DoG-eat-DoG world—altered transcriptional mechanisms drive downstream-of-gene (DoG) transcript production - ScienceDirect. https://www.sciencedirect.com/science/article/pii/S1097276522003197?via%3Dihub.

28. Grosso, A. R. *et al.* Pervasive transcription read-through promotes aberrant expression of oncogenes and RNA chimeras in renal carcinoma. *eLife* **4**, e09214 (2015).

29. Vilborg, A., Passarelli, M. C., Yario, T. A., Tycowski, K. T. & Steitz, J. A. Widespread Inducible Transcription Downstream of Human Genes. *Molecular cell* **59**, 449 (2015).

30. Jalkanen, A. L., Coleman, S. J. & Wilusz, J. Determinants and Implications of mRNA Poly(A) Tail Size - Does this Protein Make My Tail Look Big? *Seminars in cell & developmental biology* **0**, 24 (2014).

31. Chen, X., Castro, S. A., Liu, Q., Hu, W. & Zhang, S. Practical considerations on performing and analyzing CLIP-seq experiments to identify transcriptomic-wide RNA-Protein interactions. *Methods (San Diego, Calif.)* **155**, 49 (2018).

32. Park, P. J. ChIP-Seq: advantages and challenges of a maturing technology. *Nature reviews. Genetics* **10**, 669 (2009).