Molecular Cell
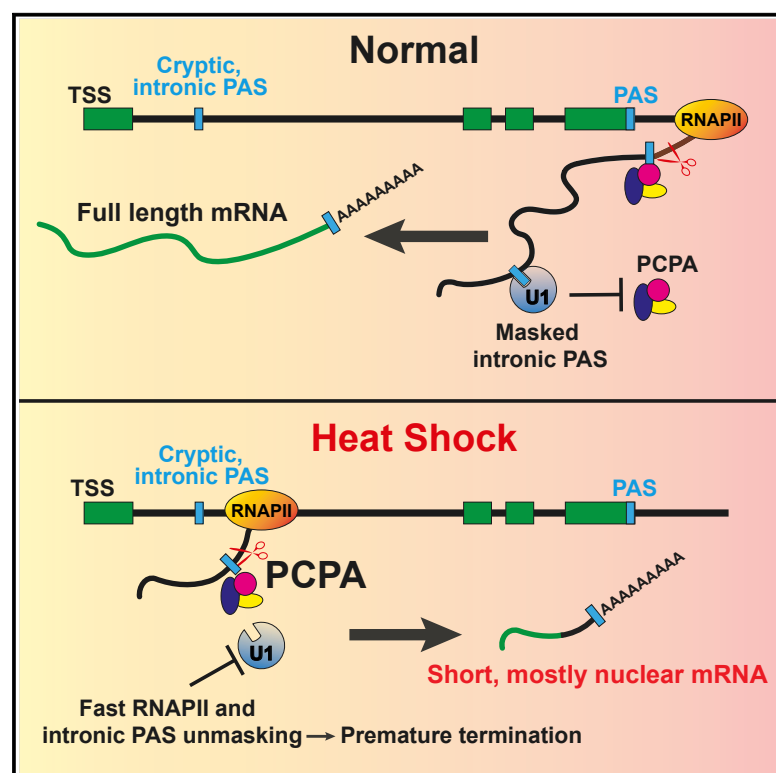
Heat shock induces premature transcript termination and reconfigures the human transcriptome

Graphical abstract

Highlights

- Transcription downregulation during HS is caused by premature termination

- Transcription elongation rate increases markedly during HS

- HS disrupts telescripting and activates cryptic poly(A) sites in introns

- HS results in widespread synthesis of new, short mRNAs species

Authors

Simona Cugusi, Richard Mitter,
Gavin P. Kelly, ..., Michael Wierer,
Aengus Stewart, Jesper Q. Svejstrup

Correspondence

jsvejstrup@sund.ku.dk

In brief

Cugusi et al. show that the elongation rate is dramatically increased during heat shock, while the global downregulation of transcription observed under these conditions is caused by premature termination at cryptic PAS in introns, which results in the synthesis of a large number of new, short mRNA species.

CellPress

Article

Heat shock induces premature transcript termination and reconfigures the human transcriptome

Simona Cugusi,[1] Richard Mitter,[2] Gavin P. Kelly,[2] Jane Walker,[1] Zhong Han,[1,3] Paola Pisano,[4] Michael Wierer,[4] Aengus Stewart,[2] and Jesper Q. Svejstrup[1,3,5,*]

[1]Mechanisms of Transcription Laboratory, The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK
[2]Bioinformatics and Biostatistics, The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK
[3]Department of Cellular and Molecular Medicine, Panum Institute, Blegdamsvej 3B, University of Copenhagen, 2200 Copenhagen N, Denmark
[4]Proteomics Research Infrastructure, Panum Institute, Blegdamsvej 3B, University of Copenhagen, 2200 Copenhagen N, Denmark
[5]Lead contact
*Correspondence: jsvejstrup@sund.ku.dk
https://doi.org/10.1016/j.molcel.2022.01.007

SUMMARY

The heat shock (HS) response involves rapid induction of HS genes, whereas transcriptional repression is established more slowly at most other genes. Previous data suggested that such repression results from inhibition of RNA polymerase II (RNAPII) pause release, but here, we show that HS strongly affects other phases of the transcription cycle. Intriguingly, while elongation rates increase upon HS, processivity markedly decreases, so that RNAPII frequently fails to reach the end of genes. Indeed, HS results in widespread premature transcript termination at cryptic, intronic polyadenylation (IPA) sites near gene 5′-ends, likely via inhibition of U1 telescripting. This results in dramatic reconfiguration of the human transcriptome with production of new, previously unannotated, short mRNAs that accumulate in the nucleus. Together, these results shed new light on the basic transcription mechanisms induced by growth at elevated temperature and show that a genome-wide shift toward usage of IPA sites can occur under physiological conditions.

INTRODUCTION

The ability to respond rapidly to stress is crucial for cellular, tissue, and organismal survival. The heat shock response (HSR) is a highly conserved cellular program induced by the exposure to a variety of environmental stressors, among which heat shock (HS) was the first observed and the most studied (Lindquist, 1986; Morimoto, 1998). Following HS, a subset of genes, known as the HS genes, are rapidly induced by heat shock factor (HSF) to help maintain protein homeostasis and ensure cell survival (Vihervaara et al., 2018), not least through the production of protein chaperones to negate heat-induced protein unfolding (Lindquist and Craig, 1988; Schopf et al., 2017). However, several lines of evidence suggest that the HSR has a broad range of physiological roles (Akerfelt et al., 2010), and a growing body of research indicates that it is co-opted by cancer cells to support malignancy and that it also plays a role in the pathogenesis of neurodegenerative diseases (Gomez-Pastor et al., 2018).

The mechanism underlying HS-induced gene activation has been extensively studied, and the HSR has generally proven to be a helpful model to more generally understand the dynamics of transcription (Wu et al., 2003; Ardehali et al., 2009; Guertin et al., 2010; Chen et al., 2017; Liang et al., 2018; Vihervaara et al., 2018; Gressel et al., 2019). Interestingly, while the upregulation of HS genes has been recognized and studied for decades, a fascinating and still poorly understood aspect of the HSR is the parallel transcriptional repression of a large portion of the genome (Jamrich et al., 1977; Gasch et al., 2000; Mahat et al., 2016). HS-associated, global gene repression has been documented in different species, and in recent years efforts have been made to reveal the underlying mechanism (Allen et al., 2004; Mariner et al., 2008; Mahat et al., 2016; Vihervaara et al., 2017; Aprile-Garcia et al., 2019; Rawat et al., 2021). Specific short interspersed nuclear elements (SINEs), which are upregulated during HS, are capable of binding directly to RNA polymerase II (RNAPII) and preventing the formation of the pre-initiation complex (Allen et al., 2004; Kettenberger et al., 2006; Mariner et al., 2008). This led to the suggestion that repression by HS might occur at initiation. However, this mechanism does not appear to be a widespread mode of action. In fact, subsequent genome-wide studies indicate that at most HS-repressed genes, transcription starts normally but progression through the gene body is halted by an unknown mechanism (Hieda et al., 2005; Teves and Henikoff, 2011; Mahat et al., 2016; Vihervaara

et al., 2017). The prevalent current models resulting from this work suggest that HS somehow prevents RNAPII from transitioning into productive elongation, resulting in increased promoter-proximal pausing. Interestingly, however, the approaches typically used so far have arguably not been ideal for analyzing the entire transcription cycle, and the results obtained have yet to present a compelling molecular model for HS-induced repression of gene expression.

While regulation of transcription in general was initially thought to occur almost exclusively at initiation, research over the last two decades has shown that regulation of other phases of the transcription cycle can affect gene expression as well (Bentley, 2014; Kamieniarz-Gdula and Proudfoot, 2019). For example, most RNAPII genes are subject to a low level of premature 3′-end cleavage and polyadenylation (PCPA) at cryptic, intronic polyadenylation (IPA) sites. It was first thought that only polyadenylation sites (PASs) in terminal exons or 3′ untranslated regions (UTRs) are utilized, but more recent work shows PCPA occurring across genes and particularly at IPA sites, depending on cell type, growth conditions, and sometimes disease state (Wang et al., 2008, 2018; Gruber and Zavolan, 2019). Intriguingly, interference with U1 snRNA (U1) base pairing to 5′-splice sites, necessary for U1 function in splicing, was found to cause widespread PCPA and frequent premature transcription termination (Kaida et al., 2010; Berg et al., 2012). Indeed, U1's suppression of PCPA activity and transcription termination is crucial for allowing the production of long (and full-length) transcripts from thousands of vertebrate genes, a process termed telescripting (Venters et al., 2019). Whether inhibition of telescripting also controls the level of gene expression under physiological conditions has remained unknown.

To better understand the mechanisms employed by cells to repress gene expression after HS, we investigated whether other phases of the transcription cycle such as elongation and termination are affected as well. Here, we present evidence that transcription downregulation during HS is mainly achieved by premature transcript termination through disruption of telescripting. This results in dramatic reconfiguration of the human transcriptome with thousands of new, short mRNAs being produced.

RESULTS

HS results in defective transcription elongation

To obtain an expansive view of the transcription changes occurring during HS in human cells, we analyzed nascent RNA synthesis using transient transcriptome sequencing (TT$_{chem}$-seq) (Schwalb et al., 2016; Gregersen et al., 2020). We initially used colony survival assays to identify the longest time of continuous HS that minimally affects cell viability (Figure S1). These experiments were carried out in parental and HSF1 knockdown MRC5-VA cell lines, using mild HS conditions for up to 4 h. The results led us to perform the TT$_{chem}$-seq experiment after exposing cells to 43°C for 2 h.

Single gene profiles identified by manual inspection of the data using genome browser nicely illustrate the results of such analysis: at the *DAP* and *FLI1* genes, for example, HS treatment appeared to have little, if any, effect on transcription in the 5′ region,

while activity progressively decreased in the rest of the gene (Figure 1A). These and other similar observations (Figure S2A) were confirmed by qPCR analysis (Figure S2B). In these follow-up experiments, we also noticed that transcription typically recovered between 1 and 3 h after cessation of HS, depending on gene length.

Metagene analysis confirmed that this effect on transcription was general; while transcription near the 5′-end of genes was largely unperturbed in response to HS, a progressive decrease in activity was observed toward the 3′-end of genes (Figure 1B). This pattern was most easily observed in longer genes. Furthermore, coverage plots of the same data, obtained by aligning genes at their transcription start sites (TSSs), indicated that HS and control profiles typically begin to diverge only 10–30 kb downstream of the TSS (Figure 1C), well beyond the promotor-proximal pause site. This was also observed in the individual gene examples (Figures 1A and S2A).

To study the behavior of the genes downregulated by HS, we also performed differential gene expression analysis, which resulted in the identification of 2,324 downregulated genes and 1,622 upregulated genes (2-fold change; FDR 0.01) (Table S1). The two subsets were then used for travel ratio analysis (Figure 1D), calculated by dividing the reads in the promoter-proximal region by the reads in the rest of the gene (Reppas et al., 2006). Downregulated genes showed a clear shift toward higher ratios, reflecting the reduction of activity in the gene body. By contrast, upregulated genes exhibited a shift in the opposite direction, consistent with the increased transcription. As suggested by the data in Figure 1, longer genes were relatively more affected, with ∼55% of genes longer than 90 kb downregulated this way, while only ∼32% of genes shorter than 15 kb were similarly affected (Figure S2C).

To investigate the generality of the phenomenon of decreased elongation after HS, we also performed a metagene analysis of previously published PRO-seq data from human embryonic kidney 293 (HEK293) cells exposed to mild HS for 90 min (Aprile-Garcia et al., 2019). To focus on differences in nascent RNA production across the entire gene body, we excluded from the analysis the first 500-bp downstream of the TSS, which are dominant in PRO-seq experiments. Interestingly, also in this cell line, there are dramatic HS-induced reductions in transcription activity that occur well beyond the pause release site (Figure S3A). This was also illustrated by single-gene examples, which showed a similar overall transcription pattern in MRC5-VA and HEK293, despite the different techniques used for measuring transcription activity (Figure S3B).

HS impairs RNAPII processivity

Due to the approach used, we could not initially rule out that we were observing a transcription recovery phase from an earlier initiation- or pause-release inhibition, rather than repression of transcription. We therefore performed a time course experiment. Here, nascent transcription was analyzed in samples exposed either to 30, 60, or 180 min of HS.

The *DAP* gene illustrates the results obtained (Figure 2A; see also Figure S4A for more examples). As time progresses, less and less transcription is observed in the downstream regions of the gene, while substantial activity is detected in the 5′-end
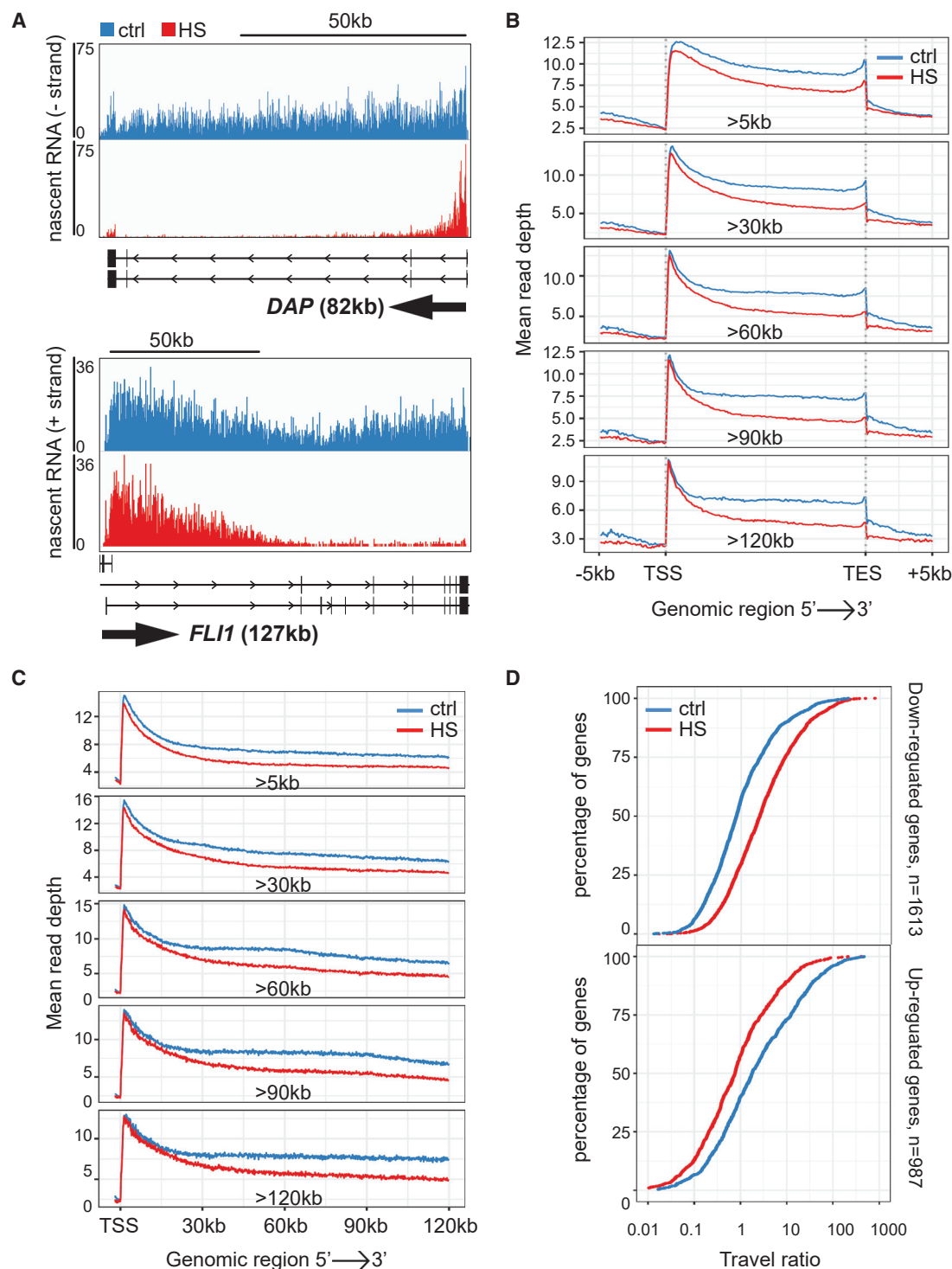
Figure 1. Heat shock results in defective transcript elongation

(A) IGV genome browser views of TT$_{chem}$-seq for the *DAP* and *FLI1* genes. Arrows indicate the direction of transcription.

(B) Metagene profiles of TT$_{chem}$-seq stratified by gene length.

(C) Coverage plots of TT$_{chem}$-seq stratified by gene length. TES, 3′ transcript end site (poly(A) site).

(D) Travel ratios of genes downregulated by HS (upper panel) and genes upregulated by HS (lower panel). See also Figures S1–S3. Genome-wide visualizations in the figures of this report generally represent at least two merged biological replicates.

A

B
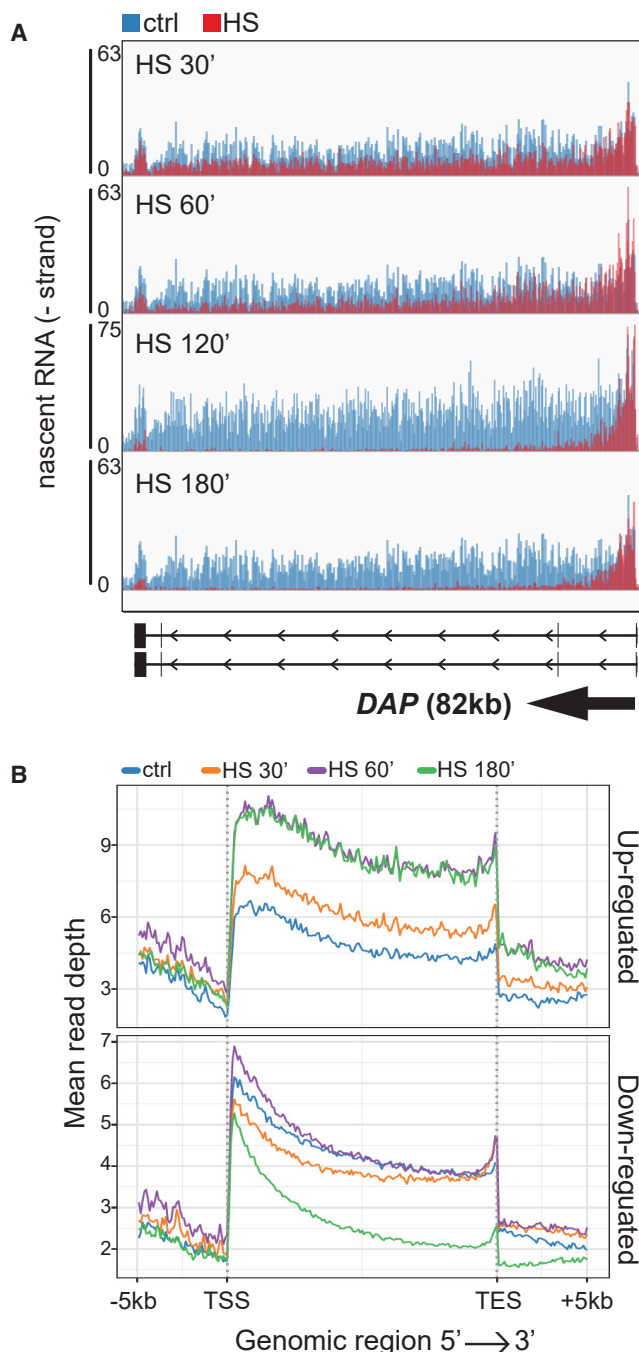
Figure 2. Heat shock impairs transcription processivity
(A) IGV genome browser views of TT_chem-seq for the *DAP* gene in cells exposed to HS for different times.
(B) Metagene TT_chem-seq profiles over time of genes upregulated (upper panel) or downregulated (lower panel) by HS. See also Figure S4.

throughout. Metagene analysis was performed on upregulated and downregulated genes separately. Upregulated genes showed the expected increase in reads throughout the transcription unit already after 30 min (Figure 2B, upper panel). More

importantly, downregulated genes were generally only modestly affected with shorter treatment (30 and 60 min) while at the longest treatment (180 min) there was a clear reduction in reads toward the 3′-end (Figure 2B, lower panel), like the results obtained at the 2-h treatment in the previous experiments (Figure 1). It is, however, worth noting that although we do not observe a striking pattern of downregulation over the entire gene unit at earlier time points in the metagene profiles, the single gene examples clearly show that downregulation has already started at 30 and 60 min, but that it has not reached complete shutdown of the gene, or the end of the gene unit. In general, it is only the 3′ end of the downregulated genes that is affected, with activity in this region progressively reduced to a level that then remains largely unchanged between 2- and 3-h treatments.

Interestingly, we also uncovered a small group of long genes, such as *STK39* and *ATXN10*, in which a marked drop of reads is observed with 180-min treatment, and which appear to be the result of a wave of decreased activity originating in the beginning of the gene which then spread over time to the rest of the gene (Figures S4B and S4C). Intriguingly, the region in which the decrease in reads first occurs is again downstream of the promotor-proximal pause site (see Figure S4C), suggesting that this phenomenon is also not explained by inhibition of pause release.
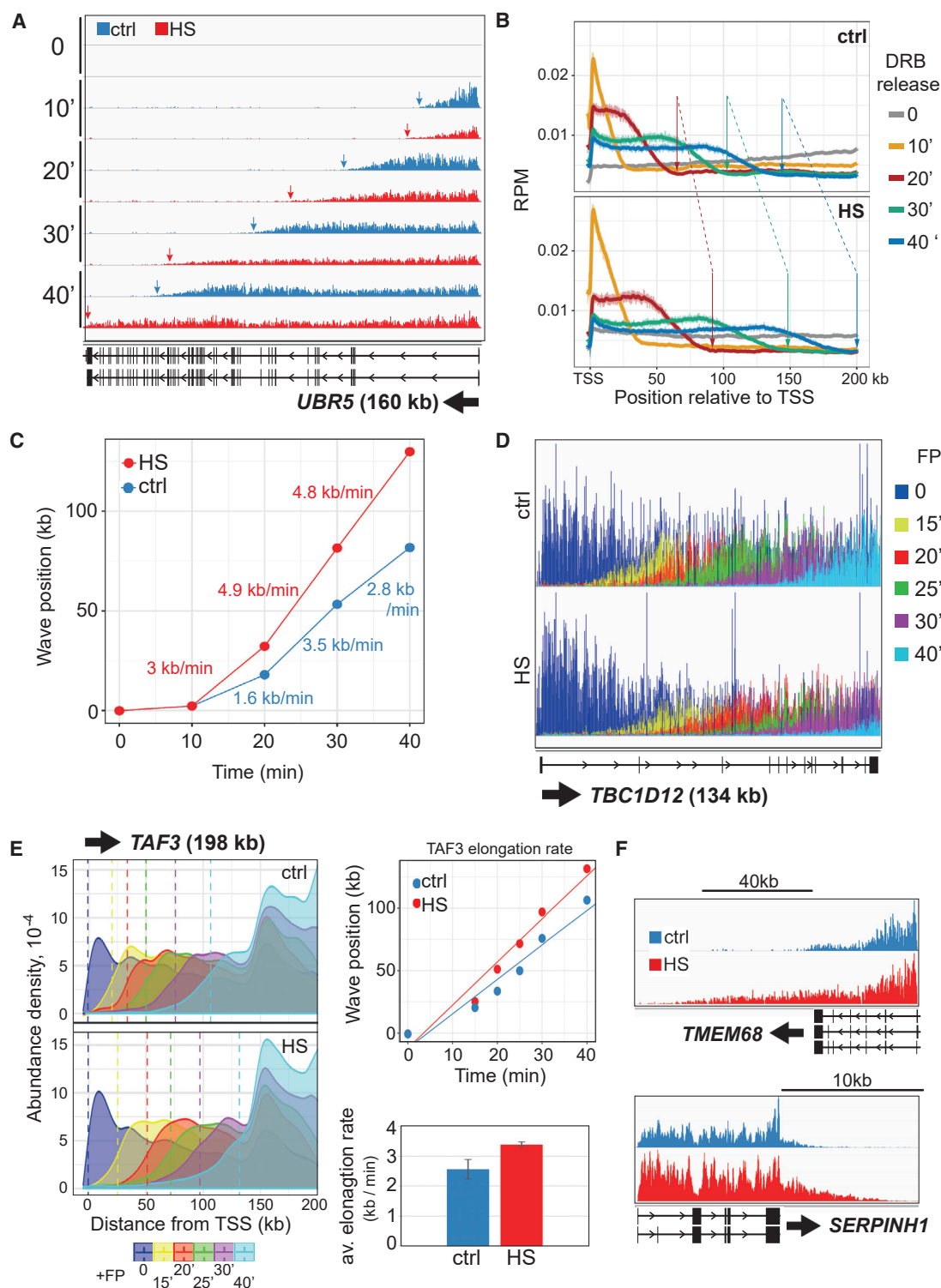
HS increases the speed of elongation

Previous studies had suggested that HS might not alter elongation speed (Ardehali and Lis, 2009; Mahat et al., 2016). However, considering our results, we now sought to directly measure the elongation rate during HS. For this purpose, we first utilized the CDK9 inhibitor 5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole (DRB) with TT_chem-seq, in DRB/TT_chem-seq (Gregersen et al., 2019, 2020). Here, the position of RNAPII in the body of genes is analyzed by TT_chem-seq at different times after removing DRB to release RNAPII from promoter-proximal areas. The labeling and collection steps were designed to obtain a transcription time course from 0 to 40 min after DRB removal under HS conditions.

Surprisingly, single gene profiles and wavefront peak calling analysis clearly showed that RNAPII advances further into genes in HS conditions (Figures 3A, 3B, and S5A). The average transcript elongation rate thus increased from ~2.1 kb/min to ~3.4 kb/min after HS, a 60% increase. This was not the result of faster release from pausing as the elongation rates between the different time points were consistently higher and increasing with time (i.e., further into the gene) (Figure 3C).

The increase in elongation rates was unexpected, especially given that it was also observed in genes that are downregulated by HS. Interestingly, however, alongside the primary observations on elongation rate, we noticed that the downregulated genes appeared to be transcribed at an overall level similar to that of untreated cells for the first 20 min following DRB release and only started showing signs of repression at the 30 min time point (see examples in Figure S5B). Given that these cells had already been subjected to HS for 2 h prior to DRB release, this shows the establishment of repression is transcription dependent, e.g., that a factor(s) required for RNAPII to reach the end of genes is being increasingly "depleted" by transcription activity in HS under normal conditions, but not in DRB where there is

Figure 3. HS increases the speed of transcription elongation

(A) IGV genome browser view of DRB/TT$_{chem}$-seq for the *UBR5* gene. Arrows indicate the front of the transcription wave.

(B) DRB/TT$_{chem}$-seq metagene profiles. Arrows indicate the front of the transcription wave.

(C) Plot of transcription as a function of time after DRB release to calculate elongation rate. The rates between time points are indicated for the two conditions.

(D) IGV genome browser view of FP/TT$_{chem}$-seq for *TBC1D12* gene.

(legend continued on next page)

little or no transcription. Importantly, this likely explains how evidence of transcription not proceeding very far into genes after HS could be obtained by TT$_{chem}$-seq (Figure 1), only for RNAPII apparently running "further" into genes after HS in the DRB/TT$_{chem}$-seq experiments (Figure 3). As discussed later, this observation also provides important evidence for the underlying mechanism.

Given the unexpected nature of these findings, we also investigated the effect of HS by using the kinase inhibitor flavopiridol (FP) combined with TT$_{chem}$-seq. In this approach (Jonkers et al., 2014), addition of FP prevents new transcription complexes from being released from promoter-proximal areas but allows already elongating polymerases to proceed. Over time, this leads to regions of increasing size without transcription activity as elongating polymerases vacate them. Briefly, after exposing cells to HS for 2 h so that repression was established, FP was added and TT$_{chem}$-seq performed at five consecutive time points in HS conditions. For obvious reasons, it was not possible to follow the dynamics of RNAPII clearance in the genes that were most severely downregulated by HS because the region remaining actively transcribed in these genes was typically too short for the earliest time point analyzed (see example in Figure S6A, left). However, genes that were only mildly affected by HS and genes insensitive to HS showed markedly faster RNAPII clearance in HS conditions (Figures 3D and S6A, right). Measuring average elongation rates across the genome is very challenging using FP/TT$_{chem}$-seq, but we calculated the rate on three individual genes: *MCU* (2.09 kb/min without HS; 3.45 kb/min during HS), *TAF3* (2.76/3.45) and *TBC1D12* (2.8/3.26), suggesting an increase in average rate from ~2.6 kb/min to ~3.4 kb/min during HS (Figure 3E), akin to what was observed by DRB/TT$_{chem}$-seq.

Moreover, to obtain a genome-wide view, we aligned all long measurable genes at their TSS and generated an average coverage plot of the first 120 kb (Figure S6B). In this approach, the slope on the posterior side of the transcriptional wave is indicative of the extent of clearance by the elongating RNAPII and therefore attests to its speed. Progressively shallower slopes were observed for increasing time points, indicative of RNAPIIs having increasingly vacated these areas of genes. Tellingly, the posterior side of the transcription wave was markedly shallower for the HS samples at every single time point, indicating more clearance and thus faster elongation speed, across the genome (Figure S6B).

Fast elongation often results in increased readthrough beyond the normal transcription end site (TES) (McDowell et al., 1994; Fong et al., 2015; Gregersen et al., 2019). Interestingly, terminator readthrough had actually already been reported in mouse cells exposed to HS (Vilborg et al., 2017; Cardiello et al., 2018). Given its relevance as an indicator of increased elongation rates, we investigated readthrough by measuring the relative transcription levels in regions 20-kb downstream of the TES. HS resulted in a clear increase in readthrough, which was more pronounced for upregulated genes (Figures 3F and S6C). Actually, the presence of frequent readthrough almost certainly led to an overestimate of the number of upregulated genes in our initial expression analysis: some of the genes in this category thus appear to represent false positives due to readthrough transcription from an upstream HS gene (Figure S6D).

Evidence that HS results in widespread usage of cryptic IPA sites

Whether the increase in elongation rate plays a role in the downregulation of genes during HS was unclear. However, if it does, it would appear to not be through the established positive correlation between elongation rate and processivity (Mason and Struhl, 2005; Fong et al., 2015; Liang et al., 2018; Geisberg et al., 2020). One explanation for low processivity in HS would be premature termination brought about by the canonical termination factors acting at cryptic IPA sites. To investigate this possibility, we generated a cell line in which cleavage and polyadenylation specificity factor 73 KDa (CPSF73, also known as CPSF3) can be rapidly depleted using the dTAG system (Nabet et al., 2018; Figures 4A, S7A, and S7B). CPSF73 is required for transcript cleavage at both IPA and 3′-end poly-adenylation sites, and thus for transcriptional termination (Kamieniarz-Gdula and Proudfoot, 2019). We tested the effect of CPSF73 depletion on nascent transcription after HS, initially by RT-qPCR using primers targeting introns at the 5′end and the 3′end of the analyzed genes, respectively. Consistent with other data (Eaton et al., 2020), we observed that CPSF73 depletion resulted in general downregulation of transcription as can be seen from the generally lower signal at the 5′ end of genes. Importantly, however, despite the overall reduction of transcription, we observed a substantial recovery of transcription levels at the 3′-end of downregulated genes after HS (Figure 4B).

Encouraged by these results, we expanded the analysis to the genome-wide level, generating TT$_{chem}$-seq maps upon CPSF73 depletion. Remarkably, depleting this poly(A)-site-specific RNA endonuclease largely reversed the effect of HS on transcription, so that RNAPII could now transcribe to the end of genes even in HS (single gene examples in Figures 4C, S7C, and S7D; genome-wide analysis in Figures 4D). Together, these data strongly indicate an important role for transcriptional termination downstream of premature/cryptic IPA sites in the mechanism underlying transcription downregulation after HS.

HS transcription and U1 telescripting

Besides CPSF73, several other factors are known to affect PCPA, including PCF11 and CPSF6 (Li et al., 2015; Kamieniarz-Gdula et al., 2019; So et al., 2019; Wang et al., 2019), RNAPII kinases such as CDK12 and CDK13 (Dubbury et al., 2018; Krajewska et al., 2019), SCAF4/SCAF8 (Gregersen et al., 2019), and U1 snRNP in the context of U1 telescripting (Di et al., 2019). Interestingly, we found that U1 expression is markedly downregulated by HS in MRC5-VA cells (Figure S8A). U1,

(E) Left, *TAF3* gene coverage profile as smoothed density from FP/TT$_{chem}$-seq; dashed lines indicate the intersect with the x axis of the midpoint at the rear end of the receding waves. Right top, plot of receding wave position as a function of time after FP addition in the *TAF3* gene. Right bottom, average elongation rate calculated using the *TAF3, MCU*, and *TBC1D12* genes. Error bars indicate ±SD.

(F) Readthrough examples; IGV genome browser view of TT$_{chem}$-seq for *TMEM68* and *SERPINH1*. See also Figures S5, S6, and S12.
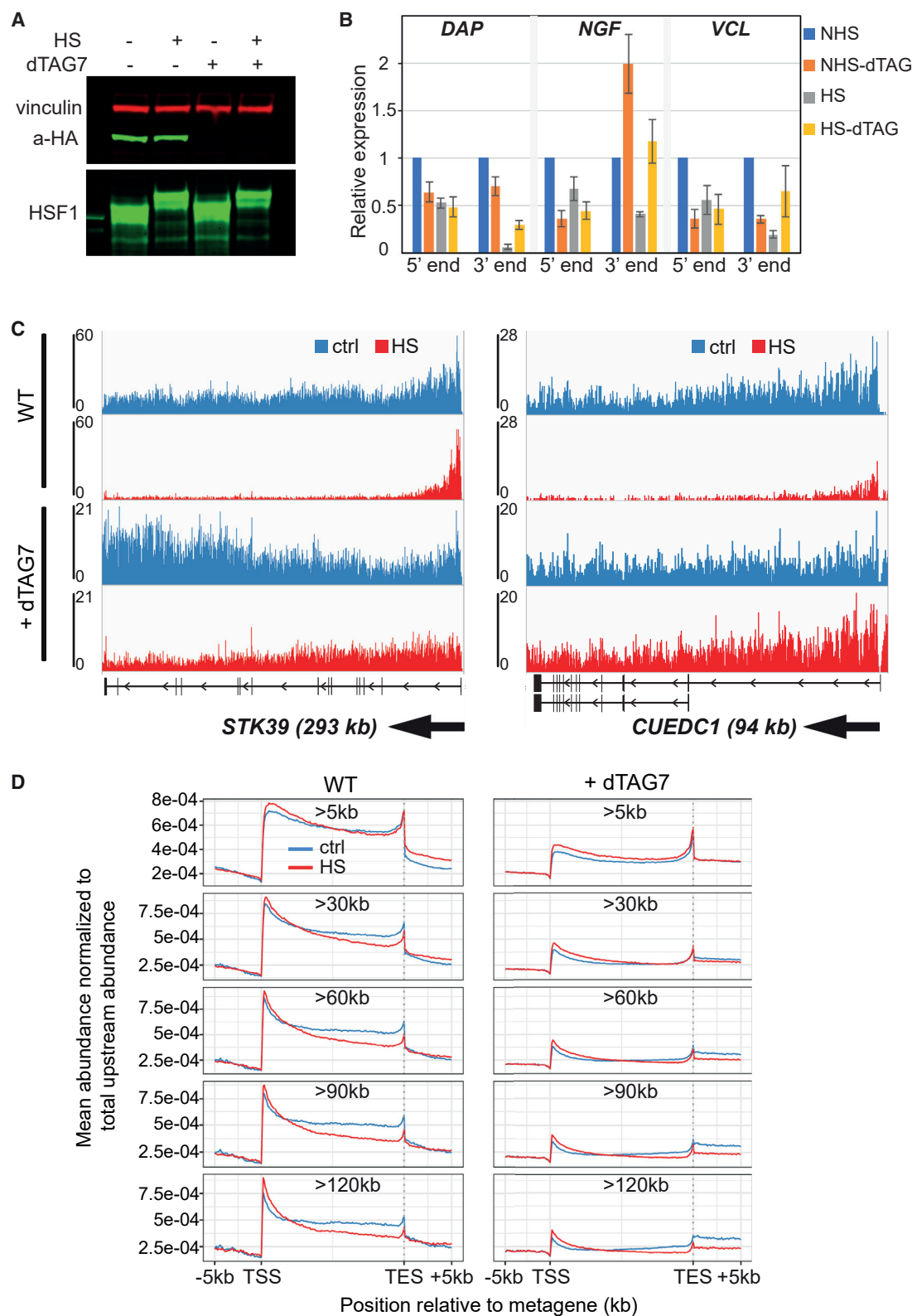
Figure 4. CPSF73 depletion rescues HS downregulation

(A) Western blot analysis of CPSF73(dTAG)-HA cells treated with dTAG7 degradation-inducer and HS. Vinculin is used as a loading control and HSF1 as a control for HS treatment.

(legend continued on next page)

otherwise best known as a component of the spliceosome (Lerner et al., 1980; Papasaikas and Valcarcel, 2016), serves the additional role of preventing the recognition of cryptic IPA sites by the transcription termination factors, effectively repressing PCPA (Kaida et al., 2010; Di et al., 2019). Its depletion by siRNA, or interference with its binding to the nascent RNA by morpholinos, thus leads to suppression of telescripting, via increased usage of IPA sites and increased short isoform production (Berg et al., 2012; Oh et al., 2017; So et al., 2019). Given that our initial results were in effect already clear evidence for inhibition of telescripting (i.e., inhibition of the ability to transcribe "far"), we therefore investigated whether there was a connection between HS-induced termination and this U1 mechanism.

Several genomic features correlate with sensitivity to the disruption of U1-telescripting and more broadly to the likelihood of PCPA occurrence, including gene length (Oh et al., 2017; Krajewska et al., 2019; Wang et al., 2019). We first found that HS-downregulated genes were indeed generally longer (Figure S8B, left). Actually, after HS, the shorter the gene, the more likely it was that its transcription level increased, and conversely, the longer the gene, the more likely its transcription level decreased (Figure S8B, right). Importantly, other features were also consistent with the idea that HS downregulation is due to the inhibition of U1 telescripting: HS-downregulated genes have longer introns, high frequency of IPA sites, low GC content, and low U1/PAS ratio (Figure S8C; see STAR Methods for details) (Oh et al., 2017; Krajewska et al., 2019; Wang et al., 2019). By contrast, HS-induced genes and unchanged genes do not show these features, lowering their sensitivity to disruption of telescripting. In fact, the classical HS genes are typically short and intron-less, which may explain how they escape this repression mechanism.

Most significantly, however, we observed that the pattern and manner of gene downregulation in MRC5-VA during HS is strikingly similar to that previously described in HeLa cells in which U1 function was artificially disrupted using U1 morpholinos (U1 AMO, U1 antisense morpholino oligonucleotide) (So et al., 2019), with high levels of transcription in the 5′ region of the genes, but similar, dramatic reductions further into the gene (Figure 5A). The correlation between HS downregulation and disruption of U1 telescripting was confirmed at the global level by gene set enrichment analysis (GSEA) (Subramanian et al., 2005) using a list of genes resulting in PCPA following U1 AMO treatment (Oh et al., 2017; Figure 5B, top panel). Here, genes from the HS dataset were ranked by their level of differential gene expression and compared with genes showing PCPA after exposure to U1 AMO, with the enrichment score showing a strong correlation between the downregulated gene sets (ES −0.38, padj < 1.5e−10). A similar overlap was observed between upregulated genes (Figure 5B, bottom panel) (ES 0.45, padj < 1.5e−10), further supporting U1 involvement in the HS response.

The genes downregulated by HS express new, shorter mRNAs

The data presented so far were on nascent RNAPII transcription. Now, to gain insight into the fate of the RNA produced during HS, and to further investigate the involvement of telescripting in determining isoform production, we examined the generation of new mature mRNAs, using neo-mRNA-seq. In this approach, after exposing cells to elevated temperature to establish downregulation of genes, 4SU was added for an additional period of RNA labeling under continued HS conditions. By subsequently isolating the polyadenylated fraction of the 4SU-labeled RNA, we could enrich for new, mature mRNAs produced during a time in which HS-mediated gene downregulation was being established.

Innumerable genes showed a pattern indicative of shorter mRNA isoform production during HS, correlating well with the profile of the nascent RNA and position of the cryptic IPA sites (Figures 6 and S9A). This shows that the nascent transcripts detected by TT_chem-seq generate stable mRNAs. Tellingly, while purification of poly(A)-tagged mRNA normally results in overrepresentation of 3′, poly(A)-proximal exons because of random RNA breaks occurring during sample preparation, we instead observed a clear shift toward enrichment of 5′-end polyadenylated gene fragments after HS. Although these shorter mRNA isoforms are typically not annotated, their bona fide nature was demonstrated by a dramatic increase of reads in the 5′ proximal exons (as well as in portions of the adjacent introns, marked by green background) and, more importantly, by a relative lack of splicing to exons downstream of these enriched regions. This correlated with a decrease in signal from canonical exons further downstream (Figures 6 and S9A, marked by purple background). Importantly, the putative 3′-ends of these shorter isoforms align nicely with previously mapped IPA sites (see PASs, Figures 6 and S9A) from the PolyA_DB database (Wang et al., 2018), as expected from inhibition of U1 telescripting. Together, these features point to widespread premature transcript termination being triggered through the activation of cryptic IPA sites after HS, and thus the pervasive production of new mRNA species with the potential to encode new protein/peptide species. Key results were confirmed by qPCR analysis of newly synthesized mRNA, quantifying the relative abundance of the putative proximal terminal exon relative to the canonical distal terminal exon during HS (Figure S9B). We note that qPCR analysis performed on total mRNA, without selecting for the mRNAs produced specifically during the HS treatment, led to similar conclusions (Figure S9C).

If premature termination generally occurs across the genome during HS to generate shorter stable mRNAs, a relative increase in reads of the first annotated exon compared with the other exons in the gene should be detected, across the neo-mRNA-seq dataset. Indeed, in a "first-exon enrichment analysis" no less than 70% of all transcripts showed a relative enrichment of the first exon after HS (Figure 7A). Tellingly, a similar pattern

(B) qPCR quantification of nascent RNA near the 5′-end or the 3′-end of the indicated genes, relative to *GAPDH*, normalized to the control. Average of three biological replicates; error bars indicate ±SD.
(C) IGV genome browser views of TT_chem-seq for the *STK39* and *CUEDC1* genes in cells exposed to HS (or not), without (WT) or with CPSF73 depletion (+dTAG7).
(D) Metagene TT_chem-seq profiles. See also Figure S7.

A

B
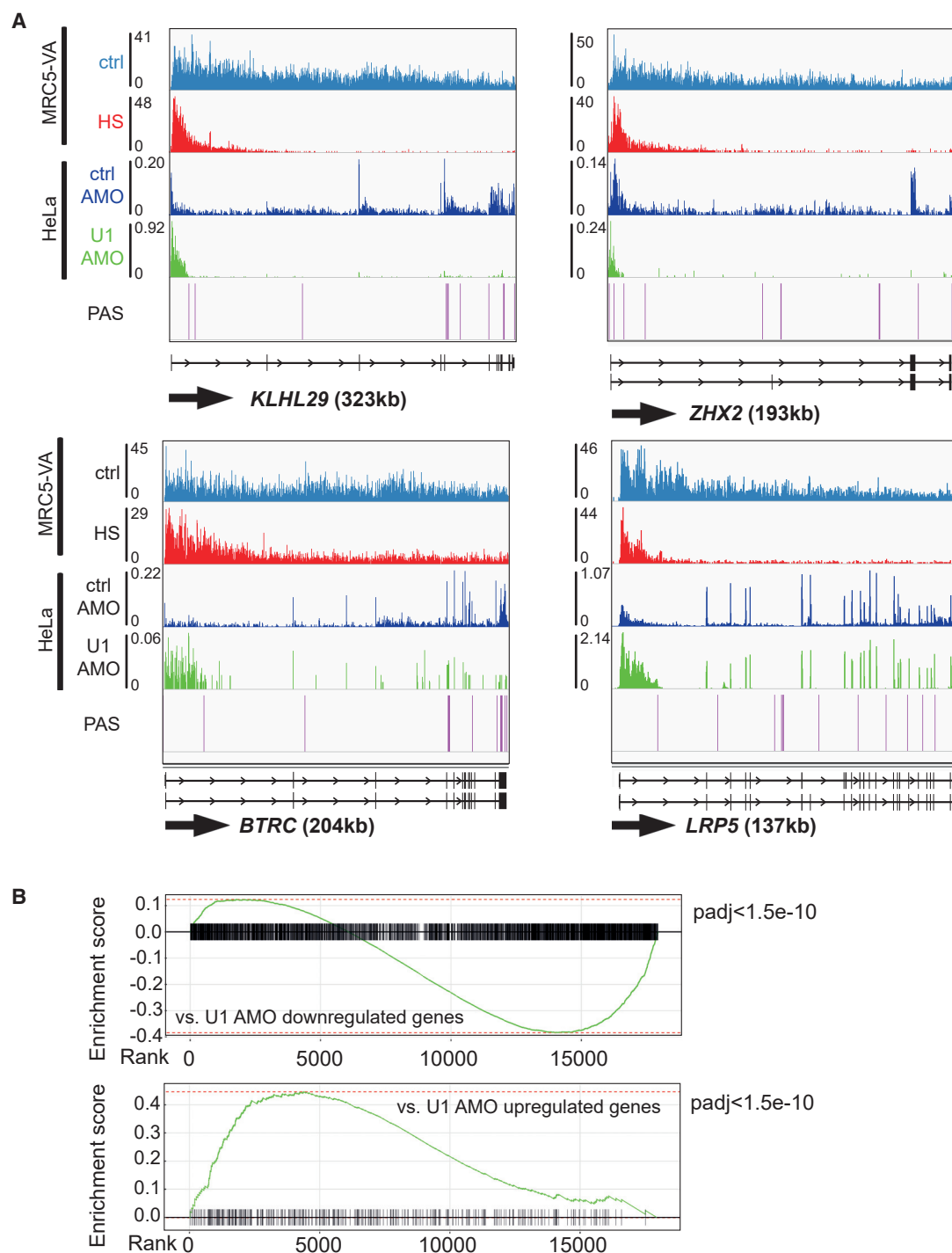
Figure 5. HS mimics telescripting disruption

(A) IGV genome browser views of TT_{chem}-seq for the *KLHL29*, *ZHX2*, *BTRC*, and *LRP5* genes in MRC5-VA cells, aligned with nascent RNA from HeLa cells treated either with ctrl- or U1-AMO (So et al., 2019) and polyadenylation sites (PASs) from the PolyA_DB database (Wang et al., 2018).

(B) Top, GSEA analysis showing the degree of overlap between the HS treatment in MRC5-VA cells, with genes ranked from most upregulated (far left) to most downregulated (far right), and genes resulting in PCPA after U1-AMO treatment in HeLa cells. Bottom, as above but compared with genes upregulated after treatment with U1-AMO in HeLa cells. The green lines represent a running enrichment score. The p values adjusted (padj) are shown. See also Figure S8.
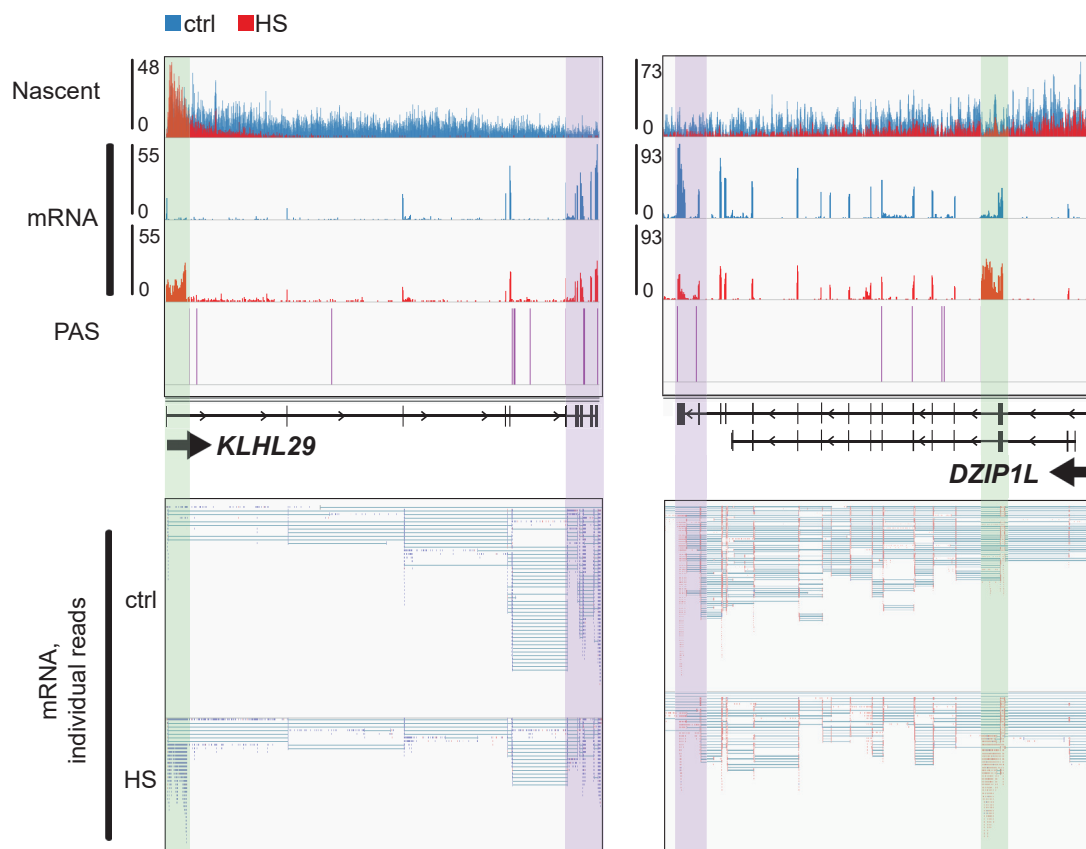
Figure 6. HS induces widespread premature termination
IGV genome browser view of TT$_{chem}$-seq, *neo*-mRNA-seq, and polyadenylation sites (PASs) (Wang et al., 2018) for the *KLHL29* and *DZIP1L* genes. For the individual mRNA reads (bottom panels), blue indicates the (+) strand and red the (−) strand. See also Figure S9.

of expression, characterized by a general and strong enrichment in expression of 5′ proximal regions, can also be observed without selecting for newly made transcripts, in a previously published RNA-seq dataset of U2OS cells exposed to HS (Seifert et al., 2015; Figures 7B and S10).

We note that although thousands of stable mRNAs are generated, a fraction of the premature termination events caused by HS will potentially produce unstable RNAs that are degraded by the exosome, which (although a consequence of premature termination) would result in an even reduction of all the exons of certain genes. Examples of genes that may be representative of this category are shown in Figure S9D.

mRNA translation?
To investigate whether the new, unannotated mRNA transcripts might be translated into novel proteins, we performed whole-proteome mass spectrometry analysis. For this purpose, we generated a virtual library of open reading frames designed to detect potential protein products (Table S2; Figure S11; see STAR Methods for details). The theoretical tryptic peptides produced from this library were then used to search data from mass spec analysis of HS-treated cells and their controls. Such analysis uncovered few, if any, novel polypeptides from this library

and those detected were present at similar level in HS and control conditions (Table S3). This suggests that the dramatic transcriptome-wide restructuring observed after HS does not translate into a new proteome to a substantial degree. One possible explanation for this would be if the novel mRNA transcripts remain in the nucleus. Indeed, short HS-induced transcripts such as those from *KLHL29*, *DZIP1L*, and *NXN* were markedly enriched in the nucleus, in contrast to the normal, "long" forms expressed from the same genes and other mature mRNAs tested (Figure 7C, left) but akin to the non-coding RNA *MALAT1* and nascent RNA expressed from *HSPH1* and *OPA1* (Figure 7, right).

Together, these results indicate that HS induces a dramatic reconfiguration of the human transcriptome, with new, short, typically unannotated transcripts being produced. These do not appear to be significantly translated but may instead remain predominantly nuclear.

DISCUSSION

HS affects the expression of a surprisingly large fraction of human genes. A limited number of these are dramatically and rapidly upregulated by HSF1-mediated gene activation. However, most
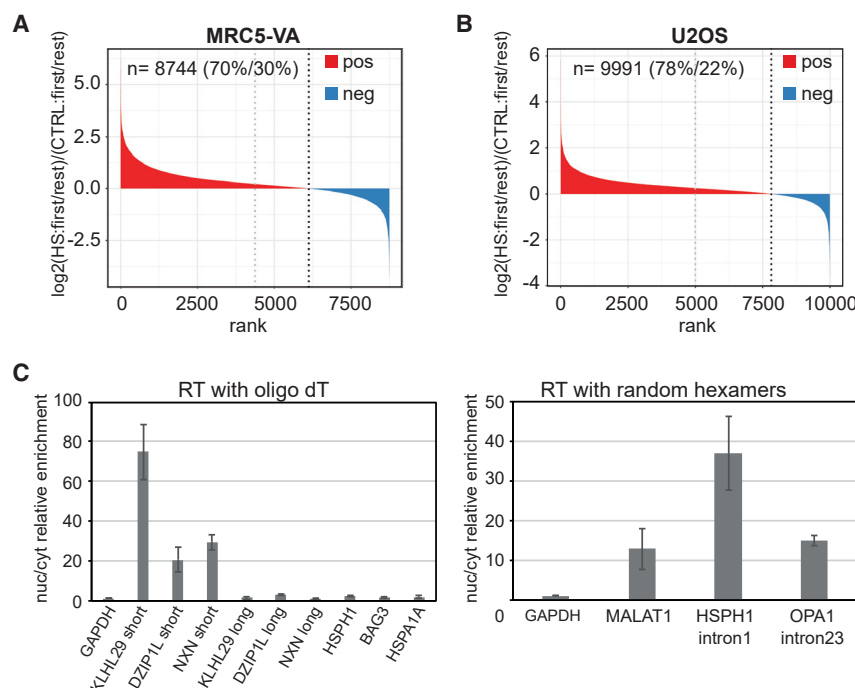
Figure 7. HS induces widespread premature termination

(A) First exon enrichment analysis plot of *neo*-mRNA-seq data from MRC5-VA cells.

(B) First exon enrichment analysis plot of RNA-seq data from U2OS cells.

(C) Left, RT-qPCR quantification in the nuclear fraction relative to the cytoplasm, of the short and long isoforms of *KLHL29*, *DIZIP1L*, and *NXN* mRNA and other canonical mRNAs (RT with oligoT), normalized to *GAPDH* (set to 1). Right, similar data for the stable non-coding RNA, MALAT1, and the nascent transcripts from *HSPH1* and *OPA1* (RT with random hexamers) for comparison. Average of three biological replicates, error bars indicate ± SD. See also Figures S10 and S11.

neo-mRNA-seq analysis of mRNA synthesized during HS showed that this leads to the production of short, typically unannotated mRNAs, resulting from the activation of cryptic IPA sites. This provides an unexpected, but compelling explanation for HS-induced downregulation of genes.

In conjunction with the previous data on pause release, our results support a model

genes are downregulated, and the underlying mechanism has remained unclear. In this report, we show that HS results in a general increase in transcript elongation rates, but also a profound failure of RNAPII to transcribe the entire gene, so that much fewer full-length transcripts are produced. Transcripts across the genome are prematurely terminated at cryptic IPA sites, likely through inhibition of U1 telescripting, providing the first example of widespread activation/usage of cryptic IPA sites in response to a physiological stimulus.

Premature transcription termination is an integral part of heat-shock-mediated gene repression

Global transcription downregulation is a conserved but poorly understood feature of the HS response (Jamrich et al., 1977; Gasch et al., 2000; Mahat et al., 2016). Somewhat surprisingly, this dramatic change in transcription does not correlate with a reorganization of chromatin structure (Teves and Henikoff, 2011; Ray et al., 2019). Moreover, it is clear that at the vast majority of genes, transcription starts normally but progression through the gene body is prevented (Allen et al., 2004; Hieda et al., 2005; Mariner et al., 2008; Teves and Henikoff, 2011; Mahat et al., 2016; Vihervaara et al., 2017; Aprile-Garcia et al., 2019; Rawat et al., 2021). The prevalent current model describes this phenomenon as an effect of pause-release inhibition. However, while previous studies were highly effective in analyzing the initial phases of the transcriptional cycle, the genome-wide approaches used (typically PRO-seq and ChIP-seq) are arguably less well suited for investigating events occurring during later stages such as elongation and termination. Using TT$_{Chem}$-seq analysis in CPSF73-depleted cells, we found that the impairment observed in RNAPII processivity is caused by premature transcript termination by the canonical termination machinery.

in which multiple mechanisms concur to achieve transcription repression during HS, and that among them, defective processivity and premature termination play a principal role, certainly during prolonged HS where the substantial downregulation of gene expression is observed.

HS elicits an increase in transcription elongation rate

Previous attempts to determine the effect of HS on elongation rate led to conflicting conclusions in different organisms (Ardehali and Lis, 2009; Miguel et al., 2013; Mahat et al., 2016). Through direct measurement of elongation speed using two complementary approaches, we now demonstrate that elongation speed increases significantly during HS in human cells. Such an increase might either be a direct consequence of faster biochemical reaction rates, the result of regulation by co-factors, or both. Given that RNA polymerases move by Brownian motion (Bar-Nahum et al., 2005), it might be expected to transcribe faster at higher temperature unless critical protein domains required for nucleotide addition and other fundamental polymerase functions begin to unfold and loose optimal function. We note that transcription elongation experiments with highly purified mammalian RNAPII alone failed to show a noteworthy difference in elongation rates between 37°C and 43°C *in vitro* (Figure S12). Conversely, factors such as EF1A1 and B2 RNA affect elongation by RNAPII during HS (Vera et al., 2014; Zovoilis et al., 2016), suggesting that RNAPII elongation rates might be elevated through the activation of such co-factors. The basis for the marked increase in cellular elongation rates after HS requires further study.

In general, the elongation phase of transcription can be described in terms of speed and processivity. These features are tightly linked, with higher elongation rates correlating with

higher processivity (Mason and Struhl, 2005; Liang et al., 2018; Geisberg et al., 2020) rather than decreased processivity as observed here. At the same time, however, slow elongation also typically results in increased exon inclusion during co-transcriptional splicing (de la Mata et al., 2003), whereas fast elongation generally results in exon exclusion (Saldi et al., 2016). The kinetic coupling model (Naftelberg et al., 2015) thus posits that slow elongation enlarges (while fast elongation reduces) the time window for recognition of upstream splice sites, thereby increasing (or decreasing) inclusion of alternative exons during co-transcriptional mRNA splicing. Intriguingly, as illustrated by the U1 morpholino experiments (Venters et al., 2019), telescripting likewise requires efficient recognition of the 5′ splice site, by U1. Akin to what is proposed in the kinetic coupling model, it is thus possible that the very high elongation rate observed during HS results in less efficient recognition of the (fast-moving) splice sites or the cryptic IPA sites themselves by the RNAPII-associated U1 snRNP, so that premature termination at cryptic IPA sites becomes more prevalent. During HS, increased elongation rates thus correlate with "decreased" processivity, but arguably in an explicable manner.

Interestingly, it seems possible that, during HS, increased elongation rates might at the same time be responsible for premature termination at some genes and transcription readthrough of canonical terminators at others. Such readthrough was previously described in mouse cells (Vilborg et al., 2017; Cardiello et al., 2018) but is detectable in our system as well; it is particularly evident in genes that are upregulated by HS. Intriguingly, the dramatic increase in elongation rate observed during HS raises the possibility that the readthrough "downstream of genes (DOGs)" transcripts reported by Steitz and co-workers during osmotic stress (Vilborg et al., 2015; Rosa-Mercado et al., 2021) might also, at least partly, be a result of increased elongation rates under these conditions. This should be investigated.

Premature termination regulation and U1-telescripting
Most human genes contain cryptic PASs in introns and these elements represent a constant, potential threat to normal gene expression (Tian and Manley, 2017; Auboeuf, 2018). Their recognition by the termination machinery is prevented by telescripting through a splicing-independent function of U1 snRNP (Di et al., 2019), or by the SCAF4/8 anti-terminator proteins (Gregersen et al., 2019). Other factors have been shown to participate in cryptic IPA site selection, either by stimulating (PCF11) or suppressing PCPA events (CPSF6, CDK12, and CDK13), but it is not yet clear whether they function independently or in the context of U1 telescripting or SCAF proteins (Li et al., 2015; Dubbury et al., 2018; Kamieniarz-Gdula et al., 2019; Krajewska et al., 2019; So et al., 2019; Wang et al., 2019).

The first ground-breaking evidence for the existence of telescripting was obtained by using U1-AMOs to compete for U1's binding to the 5′ splicing site in nascent RNA (Kaida et al., 2010). Such treatment not only compromised mRNA splicing but also resulted in widespread premature termination. Although it has long been clear that a low level of IPA normally occur in most genes, and that a handful of specific gene expression events are regulated via alternative polyadenylation (Vorlová et al., 2011; Oh et al., 2017; Kamieniarz-Gdula et al., 2019;

Wang et al., 2019), it has not previously been clear if IPA, on a genome-wide scale, can be triggered in response to a physiological stimulus. It has thus remained unclear, for example, why cells have retained IPA sites despite their obvious harmful potential. Because long genes tend to rely more on U1-telescripting, it was speculated that a general, transient switch to PCPA isoforms might be favorable in acute responses, to decrease competition for RNA processing factors (Oh et al., 2017). In line with a key role for inhibition of telescripting in determining HS-induced transcription shutdown, we find that long genes are indeed particularly affected. Several other features that characterize genes inhibited by telescripting are observed in HS-repressed genes as well.

It is important to stress that several pathways might cooperate to inhibit telescripting during HS. We initially pursued the line of investigation described here partly based on the discovery that the U1 gene is strongly repressed during HS. Frustratingly, our attempts to reverse the effect of HS by artificially overexpressing U1 were hampered for technical reasons. However, it has also been reported, for example, that SRSF10 is dephosphorylated during HS, and that this can prevent U1 snRNP binding to nascent RNA, resulting in splicing defects (Shin et al., 2004; Shi and Manley, 2007). It seems likely that a reduction of available U1 snRNPs activity this way might affect PCPA events as well. Even without reducing the total amount of U1 snRNPs, a transient imbalance between overall transcription levels and available U1 snRNPs can be sufficient to trigger PCPA events (Berg et al., 2012). This situation would likely apply to HS where induced genes reach very high level of transcription and thus might "deplete" U1 levels. Indeed, an U1 snRNP imbalance being generated by inducing transcription would be in agreement with the delay in establishing repression observed when transcription restarts after DRB release even though cells were already in HS (Figure S5B). Finally, as described above, the elevated rate of elongation during HS might itself impede recognition by U1 of 5′ splice sites, and the cryptic IPA sites themselves, and thus result in an inability to suppress the recognition of cryptic IPA sites, further helping to trigger premature termination.

The HS-induced genes escape disruption of telescripting for several reasons. First, most of the classical HS genes are short and intron-less, and more generally, genes induced by HS present genomic features that confer higher resistance to U1 depletion such as fewer and shorter introns. It is, of course, also possible that U1 is recruited to HS genes to prioritize their correct transcription and processing, akin to what happens to mRNA splicing which changes during HS as well (Biamonti and Caceres, 2009; Shalgi et al., 2014; Hussong et al., 2017).

Why terminate genes prematurely?
While the physiological importance of the mechanism outlined here is difficult to establish due to the inordinately large number of genes affected, it is an obvious possibility that a global shift to expressing short mRNA transcripts, in response to an ancient physiological stress, might help explain why cryptic PASs in introns have been retained during evolution. So, why might such a mechanism have evolved?

First, premature termination at the 5′-ends of genes that are not required for the stress response will release more polymerases

into the free RNAPII pool as less time is spent transcribing them, which in turn would enable elevated levels of transcription of the typically very short HS genes that need to be highly active for supporting survival during heat stress. Indeed, simple computational modeling of transcription based on our previous work (Ehrensberger et al., 2013; Tufegdžić Vidaković et al., 2020) indicates that the enlarged RNAPII pool resulting from premature termination at thousands of genes across the genome would "in itself" (i.e., even without HSF-activated transcription) allow unaffected (short) genes to be upregulated (see Video S1). As an aside, we note that the upregulation of some genes after U1 morpholino treatment (Venters et al., 2019) might be governed by this mechanism as well; i.e., these genes may be upregulated due to the increased RNAPII pool. In support of this idea, genes upregulated by HS and morpholino treatment significantly overlap (padj < 1.5e−10), despite their (at first glance) very different manner of induction.

Second, selectively inhibiting the expression of long mRNAs, responsible for the synthesis of large proteins, might ensure that the protein chaperone system is not overwhelmed during HS. Finally, given that the new, short mRNAs appear to accumulate in the nucleus and not support productive protein synthesis, it is tempting to speculate that some of them might have evolved to support HS survival, for example, by acting as lncRNAs or other stable RNAs, allowing more rapid transcription recovery. The fact that genome-wide use of cryptic PASs in introns can support such a multipronged stress response might even explain why such "deleterious" sites have evolved and/or been retained during evolution.

Limitations of the study

This study provides evidence that the global transcription shutdown induced in the HS response is caused by widespread activation of cryptic IPA sites in introns, resulting in premature termination. This per definition constitutes inhibition of telescripting. However, although we detected repression of U1 snRNA transcription during HS and all features of the HS-induced process also resemble inhibition of U1 telescripting, we do not provide direct evidence that the disruption of telescripting we observed is mediated by regulation of U1 activity. To investigate the importance of U1 for the mechanism, we attempted to either overexpress U1 snRNA by transient transfection of plasmids carrying the gene, or by expressing it from an inducible promoter. However, these approaches failed to yield meaningful results. First, transient transfection of plasmids carrying the U1 snRNA gene is cell lethal, and second, U1 appears to require its own promoter to start transcription at the correct start site and produce a functional snRNA. U1 is generally very abundant and stable, so it is also possible that its activity after HS is regulated by other means than by abundance.

Finally, the precise biological function of the global transcription shutdown described here remains unresolved. As described in the discussion, we can only speculate on its overall effect on, and importance for, the HS response. A reversal of transcription shutdown, of literally thousands of genes, would be required to fully characterize and understand this feature of the HSR. Such reversal is not currently possible.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell lines and culture conditions
- METHOD DETAILS
  - Generation of stable RNAi cell line
  - Generation of CPSF73 (dTAG) cell line
  - Clonogenic survival assay
  - Western blot
  - RT-qPCR
  - Nucleus-cytoplasm fractionation
  - TT_{chem}-seq
  - DRB/TT_{chem}-seq
  - FP/TT_{chem}-seq
  - 4SU/mRNA-seq
  - Library preparation
  - Mass spectrometry analysis
  - Purification of bovine RNAP II from calf thymus
  - In vitro transcription assay
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - TT_{chem}-seq alignment and quantification
  - Differential expression
  - BAM files merging
  - BigWig files
  - Read depth profiles
  - Travel ratios
  - DRB/TT_{chem}-seq: wave-front and elongation rate
  - FP/TT_{chem}-seq: wave-front and elongation rates
  - FP/TT_{chem}-seq: read-through-ratio
  - TTchem-seq: defective transcript elongation after heat shock
  - mRNA-seq first exon enrichment analysis
  - Genomic features analysis
  - mRNA-seq first exon enrichment analysis (U2OS cells)
  - HEK293 Pro-seq data analysis
  - Correlation of HS expression and U1 induced PCPA sites
  - Computer modeling of transcription
  - Data analysis, mass spectrometry

AUTHOR CONTRIBUTIONS

DECLARATION OF INTERESTS

REFERENCES

Akerfelt, M., Morimoto, R.I., and Sistonen, L. (2010). Heat shock factors: integrators of cell stress, development and lifespan. Nat. Rev. Mol. Cell Biol. *11*, 545–555. https://doi.org/10.1038/nrm2938.

Allen, T.A., Von Kaenel, S., Goodrich, J.A., and Kugel, J.F. (2004). The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. Nat. Struct. Mol. Biol. *11*, 816–821. https://doi.org/10.1038/nsmb813.

Aprile-Garcia, F., Tomar, P., Hummel, B., Khavaran, A., and Sawarkar, R. (2019). Nascent-protein ubiquitination is required for heat shock-induced gene downregulation in human cells. Nat. Struct. Mol. Biol. *26*, 137–146. https://doi.org/10.1038/s41594-018-0182-x.

Ardehali, M.B., and Lis, J.T. (2009). Tracking rates of transcription and splicing *in vivo*. Nat. Struct. Mol. Biol. *16*, 1123–1124. https://doi.org/10.1038/nsmb1109-1123.

Ardehali, M.B., Yao, J., Adelman, K., Fuda, N.J., Petesch, S.J., Webb, W.W., and Lis, J.T. (2009). Spt6 enhances the elongation rate of RNA polymerase II *in vivo*. EMBO J. *28*, 1067–1077. https://doi.org/10.1038/emboj.2009.56.

Auboeuf, D. (2018). Alternative mRNA processing sites decrease genetic variability while increasing functional diversity. Transcription *9*, 75–87. https://doi.org/10.1080/21541264.2017.1373891.

Bar-Nahum, G., Epshtein, V., Ruckenstein, A.E., Rafikov, R., Mustaev, A., and Nudler, E. (2005). A ratchet mechanism of transcription elongation and its control. Cell *120*, 183–193. https://doi.org/10.1016/j.cell.2004.11.045.

Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. Nat. Rev. Genet. *15*, 163–175. https://doi.org/10.1038/nrg3662.

Berg, M.G., Singh, L.N., Younis, I., Liu, Q., Pinto, A.M., Kaida, D., Zhang, Z., Cho, S., Sherrill-Mix, S., Wan, L., and Dreyfuss, G. (2012). U1 snRNP determines mRNA length and regulates isoform expression. Cell *150*, 53–64. https://doi.org/10.1016/j.cell.2012.05.029.

Biamonti, G., and Caceres, J.F. (2009). Cellular stress and RNA splicing. Trends Biochem. Sci. *34*, 146–153. https://doi.org/10.1016/j.tibs.2008.11.004.

Cardiello, J.F., Goodrich, J.A., and Kugel, J.F. (2018). 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–e00118. https://doi.org/10.1128/MCB.00181-18.

Chen, F.X., Xie, P., Collings, C.K., Cao, K., Aoi, Y., Marshall, S.A., Rendleman, E.J., Ugarenko, M., Ozark, P.A., Zhang, A., et al. (2017). PAF1 regulation of

promoter-proximal pause release via enhancer activation. Science *357*, 1294–1298. https://doi.org/10.1126/science.aan3269.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372. https://doi.org/10.1038/nbt.1511.

de la Mata, M., Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2003). A slow RNA polymerase II affects alternative splicing *in vivo*. Mol. Cell *12*, 525–532.

Di, C., So, B.R., Cai, Z., Arai, C., Duan, J., and Dreyfuss, G. (2019). U1 snRNP telescripting roles in transcription and its mechanism. Cold Spring Harb. Symp. Quant. Biol. *84*, 115–122. https://doi.org/10.1101/sqb.2019.84.040451.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

Dubbury, S.J., Boutz, P.L., and Sharp, P.A. (2018). CDK12 regulates DNA repair genes by suppressing intronic polyadenylation. Nature *564*, 141–145. https://doi.org/10.1038/s41586-018-0758-y.

Eaton, J.D., Francis, L., Davidson, L., and West, S. (2020). A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes. Genes Dev. *34*, 132–145. https://doi.org/10.1101/gad.332833.119.

Ehrensberger, A.H., Kelly, G.P., and Svejstrup, J.Q. (2013). Mechanistic interpretation of promoter-proximal peaks and RNAPII density maps. Cell *154*, 713–715. https://doi.org/10.1016/j.cell.2013.07.032.

Fong, N., Brannan, K., Erickson, B., Kim, H., Cortazar, M.A., Sheridan, R.M., Nguyen, T., Karp, S., and Bentley, D.L. (2015). Effects of transcription elongation rate and Xrn2 exonuclease activity on RNA polymerase II termination suggest widespread kinetic competition. Mol. Cell *60*, 256–267. https://doi.org/10.1016/j.molcel.2015.09.026.

Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D., and Brown, P.O. (2000). Genomic expression programs in the response of yeast cells to environmental changes. Mol. Biol. Cell *11*, 4241–4257. https://doi.org/10.1091/mbc.11.12.4241.

Geisberg, J.V., Moqtaderi, Z., and Struhl, K. (2020). The transcriptional elongation rate regulates alternative polyadenylation in yeast. Elife *9*, e59810. https://doi.org/10.7554/eLife.59810.

Gomez-Pastor, R., Burchfiel, E.T., and Thiele, D.J. (2018). Regulation of heat shock transcription factors and their roles in physiology and disease. Nat. Rev. Mol. Cell Biol. *19*, 4–19. https://doi.org/10.1038/nrm.2017.73.

Gregersen, L.H., Mitter, R., and Svejstrup, J.Q. (2020). Using TTchem-seq for profiling nascent transcription and measuring transcript elongation. Nat. Protoc. *15*, 604–627. https://doi.org/10.1038/s41596-019-0262-3.

Gregersen, L.H., Mitter, R., Ugalde, A.P., Nojima, T., Proudfoot, N.J., Agami, R., Stewart, A., and Svejstrup, J.Q. (2019). SCAF4 and SCAF8, mRNA anti-terminator proteins. Cell *177*, 1797–1813.e18. https://doi.org/10.1016/j.cell.2019.04.038.

Gressel, S., Schwalb, B., and Cramer, P. (2019). The pause-initiation limit restricts transcription activation in human cells. Nat. Commun. *10*, 3603. https://doi.org/10.1038/s41467-019-11536-8.

Gruber, A.J., and Zavolan, M. (2019). Alternative cleavage and polyadenylation in health and disease. Nat. Rev. Genet. *20*, 599–614. https://doi.org/10.1038/s41576-019-0145-z.

Guertin, M.J., Petesch, S.J., Zobeck, K.L., Min, I.M., and Lis, J.T. (2010). Drosophila heat shock system as a general model to investigate transcriptional regulation. Cold Spring Harb. Symp. Quant. Biol. *75*, 1–9. https://doi.org/10.1101/sqb.2010.75.039.

Han, Z., Libri, D., and Porrua, O. (2017). Biochemical characterization of the helicase Sen1 provides new insights into the mechanisms of non-coding transcription termination. Nucleic Acids Res. *45*, 1355–1370. https://doi.org/10.1093/nar/gkw1230.

Hieda, M., Winstanley, H., Maini, P., Iborra, F.J., and Cook, P.R. (2005). Different populations of RNA polymerase II in living mammalian cells. Chromosome Res. *13*, 135–144. https://doi.org/10.1007/s10577-005-7720-1.

Hussong, M., Kaehler, C., Kerick, M., Grimm, C., Franz, A., Timmermann, B., Welzel, F., Isensee, J., Hucho, T., Krobitsch, S., and Schweiger, M.R. (2017). The bromodomain protein BRD4 regulates splicing during heat shock. Nucleic Acids Res. *45*, 382–394. https://doi.org/10.1093/nar/gkw729.

Jamrich, M., Greenleaf, A.L., and Bautz, E.K. (1977). Localization of RNA polymerase in polytene chromosomes of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA *74*, 2079–2083.

Jonkers, I., Kwak, H., and Lis, J.T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. Elife *3*, e02407. https://doi.org/10.7554/eLife.02407.

Kaida, D., Berg, M.G., Younis, I., Kasim, M., Singh, L.N., Wan, L., and Dreyfuss, G. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. Nature *468*, 664–668. https://doi.org/10.1038/nature09479.

Kamieniarz-Gdula, K., Gdula, M.R., Panser, K., Nojima, T., Monks, J., Wiśniewski, J.R., Riepsaame, J., Brockdorff, N., Pauli, A., and Proudfoot, N.J. (2019). Selective roles of vertebrate PCF11 in premature and full-length transcript termination. Mol. Cell *74*, 158–172.e9. https://doi.org/10.1016/j.molcel.2019.01.027.

Kamieniarz-Gdula, K., and Proudfoot, N.J. (2019). Transcriptional control by premature termination: A forgotten mechanism. Trends Genet. *35*, 553–564. https://doi.org/10.1016/j.tig.2019.05.005.

Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S., and Karolchik, D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics *26*, 2204–2207. https://doi.org/10.1093/bioinformatics/btq351.

Kettenberger, H., Eisenführ, A., Brueckner, F., Theis, M., Famulok, M., and Cramer, P. (2006). Structure of an RNA polymerase II-RNA inhibitor complex elucidates transcription regulation by noncoding RNAs. Nat. Struct. Mol. Biol. *13*, 44–48. https://doi.org/10.1038/nsmb1032.

Korotkevich, G., Sukhov, V., and Sergushichev, A. (2019). Fast gene set enrichment analysis. bioRxiv, bioRxiv:10.1101/060012.

Krajewska, M., Dries, R., Grassetti, A.V., Dust, S., Gao, Y., Huang, H., Sharma, B., Day, D.S., Kwiatkowski, N., Pomaville, M., et al. (2019). CDK12 loss in cancer cells affects DNA damage response genes through premature cleavage and polyadenylation. Nat. Commun. *10*, 1757. https://doi.org/10.1038/s41467-019-09703-y.

Kulak, N.A., Geyer, P.E., and Mann, M. (2017). Loss-less nano-fractionator for high sensitivity, high coverage proteomics. Mol. Cell. Proteomics *16*, 694–705. https://doi.org/10.1074/mcp.O116.065136.

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for computing and annotating genomic ranges. PLoS Comput. Biol. *9*, e1003118. https://doi.org/10.1371/journal.pcbi.1003118.

Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L., and Steitz, J.A. (1980). Are snRNPs involved in splicing? Nature *283*, 220–224. https://doi.org/10.1038/283220a0.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Li, W., You, B., Hoque, M., Zheng, D., Luo, W., Ji, Z., Park, J.Y., Gunderson, S.I., Kalsotra, A., Manley, J.L., and Tian, B. (2015). Systematic profiling of poly(A)+ transcripts modulated by core 3′ end processing and splicing factors reveals regulatory rules of alternative cleavage and polyadenylation. PLoS Genet. *11*, e1005166. https://doi.org/10.1371/journal.pgen.1005166.

Liang, K., Smith, E.R., Aoi, Y., Stoltz, K.L., Katagi, H., Woodfin, A.R., Rendleman, E.J., Marshall, S.A., Murray, D.C., Wang, L., et al. (2018). Targeting processive transcription elongation via SEC disruption for MYC-induced cancer therapy. Cell *175*, 766–779.e17. https://doi.org/10.1016/j.cell.2018.09.027.

Lindquist, S. (1986). The heat-shock response. Annu. Rev. Biochem. *55*, 1151–1191. https://doi.org/10.1146/annurev.bi.55.070186.005443.

Lindquist, S., and Craig, E.A. (1988). The heat-shock proteins. Annu. Rev. Genet. *22*, 631–677. https://doi.org/10.1146/annurev.ge.22.120188.003215.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

Mahat, D.B., Salamanca, H.H., Duarte, F.M., Danko, C.G., and Lis, J.T. (2016). Mammalian heat shock response and mechanisms underlying its genome-wide transcriptional regulation. Mol. Cell *62*, 63–78. https://doi.org/10.1016/j.molcel.2016.02.025.

Mariner, P.D., Walters, R.D., Espinoza, C.A., Drullinger, L.F., Wagner, S.D., Kugel, J.F., and Goodrich, J.A. (2008). Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. Mol. Cell *29*, 499–509. https://doi.org/10.1016/j.molcel.2007.12.013.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.j. *17*, 3. https://doi.org/10.14806/ej.17.1.200.

Mason, P.B., and Struhl, K. (2005). Distinction and relationship between elongation rate and processivity of RNA polymerase II *in vivo*. Mol. Cell *17*, 831–840.

McDowell, J.C., Roberts, J.W., Jin, D.J., and Gross, C. (1994). Determination of intrinsic transcription termination efficiency by RNA polymerase elongation rate. Science *266*, 822–825. https://doi.org/10.1126/science.7526463.

Miguel, A., Montón, F., Li, T., Gómez-Herreros, F., Chávez, S., Alepuz, P., and Pérez-Ortín, J.E. (2013). External conditions inversely change the RNA polymerase II elongation rate and density in yeast. Biochim. Biophys. Acta *1829*, 1248–1255. https://doi.org/10.1016/j.bbagrm.2013.09.008.

Morimoto, R.I. (1998). Regulation of the heat shock transcriptional response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators. Genes Dev. *12*, 3788–3796. https://doi.org/10.1101/gad.12.24.3788.

Nabet, B., Roberts, J.M., Buckley, D.L., Paulk, J., Dastjerdi, S., Yang, A., Leggett, A.L., Erb, M.A., Lawlor, M.A., Souza, A., et al. (2018). The dTAG system for immediate and target-specific protein degradation. Nat. Chem. Biol. *14*, 431–441. https://doi.org/10.1038/s41589-018-0021-8.

Naftelberg, S., Schor, I.E., Ast, G., and Kornblihtt, A.R. (2015). Regulation of alternative splicing through coupling with transcription and chromatin structure. Annu. Rev. Biochem. *84*, 165–198. https://doi.org/10.1146/annurev-biochem-060614-034242.

Oh, J.M., Di, C., Venters, C.C., Guo, J., Arai, C., So, B.R., Pinto, A.M., Zhang, Z., Wan, L., Younis, I., and Dreyfuss, G. (2017). U1 snRNP telescripting regulates a size-function-stratified human genome. Nat. Struct. Mol. Biol. *24*, 993–999. https://doi.org/10.1038/nsmb.3473.

Papasaikas, P., and Valcárcel, J. (2016). The spliceosome: the ultimate RNA chaperone and sculptor. Trends Biochem. Sci. *41*, 33–45. https://doi.org/10.1016/j.tibs.2015.11.003.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). c-Myc regulates transcriptional pause release. Cell *141*, 432–445. https://doi.org/10.1016/j.cell.2010.03.030.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187–W191. https://doi.org/10.1093/nar/gku365.

Rawat, P., Boehning, M., Hummel, B., Aprile-Garcia, F., Pandit, A.S., Eisenhardt, N., Khavaran, A., Niskanen, E., Vos, S.M., Palvimo, J.J., et al. (2021). Stress-induced nuclear condensation of NELF drives transcriptional downregulation. Mol. Cell *81*, 1013–1026.e11. https://doi.org/10.1016/j.molcel.2021.01.016.

Ray, J., Munn, P.R., Vihervaara, A., Lewis, J.J., Ozer, A., Danko, C.G., and Lis, J.T. (2019). Chromatin conformation remains stable upon extensive

transcriptional changes driven by heat shock. Proc. Natl. Acad. Sci. USA 116, 19431–19439. https://doi.org/10.1073/pnas.1901244116.

Reppas, N.B., Wade, J.T., Church, G.M., and Struhl, K. (2006). The transition between transcriptional initiation and elongation in E. coli is highly variable and often rate limiting. Mol. Cell 24, 747–757. https://doi.org/10.1016/j.molcel.2006.10.030.

Rosa-Mercado, N.A., Zimmer, J.T., Apostolidi, M., Rinehart, J., Simon, M.D., and Steitz, J.A. (2021). Hyperosmotic stress alters the RNA polymerase II interactome and induces readthrough transcription despite widespread transcriptional repression. Mol. Cell 81, 502–513.e4. https://doi.org/10.1016/j.molcel.2020.12.002.

Saldi, T., Cortazar, M.A., Sheridan, R.M., and Bentley, D.L. (2016). Coupling of RNA polymerase II transcription elongation with pre-mRNA splicing. J. Mol. Biol. 428, 2623–2635. https://doi.org/10.1016/j.jmb.2016.04.017.

Santos, A., Colaço, A.R., Nielsen, A.B., Niu, L., Geyer, P.E., Coscia, F., Wewer Albrechtsen, N.J., Mundt, F., Jensen, L.J., and Mann, M. (2021). Clinical knowledge graph integrates proteomics data into clinical decision-making. bioRxiv, bioRxiv:10.1101/2020.05.09.084897.

Schopf, F.H., Biebl, M.M., and Buchner, J. (2017). The HSP90 chaperone machinery. Nat. Rev. Mol. Cell Biol. 18, 345–360. https://doi.org/10.1038/nrm.2017.20.

Schwalb, B., Michel, M., Zacher, B., Frühauf, K., Demel, C., Tresch, A., Gagneur, J., and Cramer, P. (2016). TT-seq maps the human transient transcriptome. Science 352, 1225–1228. https://doi.org/10.1126/science.aad9841.

Seifert, A., Schofield, P., Barton, G.J., and Hay, R.T. (2015). Proteotoxic stress reprograms the chromatin landscape of SUMO modification. Sci. Signal. 8, rs7. https://doi.org/10.1126/scisignal.aaa2213.

Shalgi, R., Hurt, J.A., Lindquist, S., and Burge, C.B. (2014). Widespread inhibition of posttranscriptional splicing shapes the cellular transcriptome following heat shock. Cell Rep. 7, 1362–1370. https://doi.org/10.1016/j.celrep.2014.04.044.

Shi, Y., and Manley, J.L. (2007). A complex signaling pathway regulates SRp38 phosphorylation and pre-mRNA splicing in response to heat shock. Mol. Cell 28, 79–90. https://doi.org/10.1016/j.molcel.2007.08.028.

Shin, C., Feng, Y., and Manley, J.L. (2004). Dephosphorylated SRp38 acts as a splicing repressor in response to heat shock. Nature 427, 553–558. https://doi.org/10.1038/nature02288.

So, B.R., Di, C., Cai, Z., Venters, C.C., Guo, J., Oh, J.M., Arai, C., and Dreyfuss, G. (2019). A complex of U1 snRNP with cleavage and polyadenylation factors controls Telescripting, regulating mRNA transcription in human cells. Mol. Cell 76, 590–599.e4. https://doi.org/10.1016/j.molcel.2019.08.007.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA 102, 15545–15550. https://doi.org/10.1073/pnas.0506580102.

Teves, S.S., and Henikoff, S. (2011). Heat shock reduces stalled RNA polymerase II and nucleosome turnover genome-wide. Genes Dev. 25, 2387–2397. https://doi.org/10.1101/gad.177675.111.

Tian, B., and Manley, J.L. (2017). Alternative polyadenylation of mRNA precursors. Nat. Rev. Mol. Cell Biol. 18, 18–30. https://doi.org/10.1038/nrm.2016.116.

Tibshirani, R., Seo, M.J., Chu, G., Narasimhan, B., and Li, J. (2018). samr: SAM: significance analysis of microarrays, R package version 3.0 https://CRAN.R-project.org/package=samr.

Tufegdžić Vidaković, A., Mitter, R., Kelly, G.P., Neumann, M., Harreman, M., Rodríguez-Martínez, M., Herlihy, A., Weems, J.C., Boeing, S., Encheva, V.,

et al. (2020). Regulation of the RNAPII pool is integral to the DNA damage response. Cell 180, 1245–1261.e21. https://doi.org/10.1016/j.cell.2020.02.009.

Venters, C.C., Oh, J.M., Di, C., So, B.R., and Dreyfuss, G. (2019). U1 snRNP telescripting: suppression of premature transcription termination in introns as a new layer of gene regulation. Cold Spring Harb. Perspect. Biol. 11, a032235. https://doi.org/10.1101/cshperspect.a032235.

Vera, M., Pani, B., Griffiths, L.A., Muchardt, C., Abbott, C.M., Singer, R.H., and Nudler, E. (2014). The translation elongation factor eEF1A1 couples transcription to translation during heat shock response. Elife 3, e03164. https://doi.org/10.7554/eLife.03164.

Vihervaara, A., Duarte, F.M., and Lis, J.T. (2018). Molecular mechanisms driving transcriptional stress responses. Nat. Rev. Genet. 19, 385–397. https://doi.org/10.1038/s41576-018-0001-6.

Vihervaara, A., Mahat, D.B., Guertin, M.J., Chu, T., Danko, C.G., Lis, J.T., and Sistonen, L. (2017). Transcriptional response to stress is pre-wired by promoter and enhancer architecture. Nat. Commun. 8, 255. https://doi.org/10.1038/s41467-017-00151-0.

Vilborg, A., Passarelli, M.C., Yario, T.A., Tycowski, K.T., and Steitz, J.A. (2015). Widespread inducible transcription downstream of human genes. Mol. Cell 59, 449–461. https://doi.org/10.1016/j.molcel.2015.06.016.

Vilborg, A., Sabath, N., Wiesel, Y., Nathans, J., Levy-Adam, F., Yario, T.A., Steitz, J.A., and Shalgi, R. (2017). Comparative analysis reveals genomic features of stress-induced transcriptional readthrough. Proc. Natl. Acad. Sci. USA 114, E8362–E8371. https://doi.org/10.1073/pnas.1711120114.

Vorlová, S., Rocco, G., Lefave, C.V., Jodelka, F.M., Hess, K., Hastings, M.L., Henke, E., and Cartegni, L. (2011). Induction of antagonistic soluble decoy receptor tyrosine kinases by intronic polyA activation. Mol. Cell 43, 927–939. https://doi.org/10.1016/j.molcel.2011.08.009.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature 456, 470–476. https://doi.org/10.1038/nature07509.

Wang, R., Nambiar, R., Zheng, D., and Tian, B. (2018). PolyA_DB 3 catalogs cleavage and polyadenylation sites identified by deep sequencing in multiple genomes. Nucleic Acids Res. 46, D315–D319. https://doi.org/10.1093/nar/gkx1000.

Wang, R., Zheng, D., Wei, L., Ding, Q., and Tian, B. (2019). Regulation of intronic polyadenylation by PCF11 impacts mRNA expression of long genes. Cell Rep. 26, 2766–2778.e6. https://doi.org/10.1016/j.celrep.2019.02.049.

Williamson, L., Saponaro, M., Boeing, S., East, P., Mitter, R., Kantidakis, T., Kelly, G.P., Lobley, A., Walker, J., Spencer-Dene, B., et al. (2017). UV irradiation induces a non-coding RNA that functionally opposes the protein encoded by the same gene. Cell 168, 843–855.e13. https://doi.org/10.1016/j.cell.2017.01.019.

Wu, C.H., Yamaguchi, Y., Benjamin, L.R., Horvat-Gordon, M., Washinsky, J., Enerly, E., Larsson, J., Lambertsson, A., Handa, H., and Gilmour, D. (2003). NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila. Genes Dev. 17, 1402–1414. https://doi.org/10.1101/gad.1091403.

Zatreanu, D., Han, Z., Mitter, R., Tumini, E., Williams, H., Gregersen, L., Dirac-Svejstrup, A.B., Roma, S., Stewart, A., Aguilera, A., and Svejstrup, J.Q. (2019). Elongation factor TFIIS prevents transcription stress and R-loop accumulation to maintain genome stability. Mol. Cell 76, 57–69.e9. https://doi.org/10.1016/j.molcel.2019.07.037.

Zovoilis, A., Cifuentes-Rojas, C., Chu, H.P., Hernandez, A.J., and Lee, J.T. (2016). Destabilization of B2 RNA by EZH2 activates the stress response. Cell 167, 1788–1802.e13. https://doi.org/10.1016/j.cell.2016.11.041.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Polyclonal to HSF1	Enzo Life Sciences	Cat# ADI-SPA-901 RRID:AB_10616511
Monoclonal to Vinculin	Sigma-Aldrich	Cat# V9131 RRID:AB_477629
Polyclonal to CPSF73	Bethyl	Cat# A301-090A RRID: AB_873009
Monoclonal to HA	Abcam	Cat# ab236632 RRID:AB_2864361
Anti-Mouse secondary antibody (HRP)	Santa Cruz	Cat# sc-516102 RRID:AB_2687626
Anti-Rabbit secondary antibody (HRP)	Jackson Immunoresearch	Cat# 711-035-152 RRID:AB_10015282
IRDye 680RD Donkey anti-Mouse secondary antibody	LI-COR Biosciences	Cat# 926-68072 RRID:AB_10953628
IRDye 800CW Goat anti-Rabbit secondary antibody	LI-COR Biosciences	Cat# 926-32211 RRID:AB_621843
Monoclonal to CTD repeat RNAPII (8WG16)	The Francis Crick Institute Core facility	N/A
Chemicals, Peptides, and Recombinant Proteins		
4-thiouridine	Glentham Life Sciences	Cat# GN6085
4-thiouracil	Sigma-Aldrich	Cat# 440736
MTSEA biotin-XXlinker ((MTSEA Biotincacap; 2-((6-((6-((biotinoyl)amino)hexanoyl)amino)hexanoyl)amino)ethylmethanethiosulfonate))	Biotium	Cat# BT90066
DRB (5,6-dichloro-1-β-D-ribofuranosylbenzimidazole)	Sigma-Aldrich	Cat# D1916
Flavopiridol hydrochloride hydrate	Sigma-Aldrich	Cat# F3055
Polyethylenimine	Polysciences, Inc.	Cat# 23966-1
Hexadimethrine bromide	Sigma-Aldrich	Cat# H9268
dTAG7	Tocris, Bio-Techne Ltd	Cat# 6912
Blasticidin	TOKU-E	Cat# B007-20ml
Puromycin	Gibco™	Cat# A1113803
Crystal violet	Sigma-Aldrich	Cat# C3886
Critical Commercial Assays		
RNeasy mini kit	QIAGEN	Cat# 74106
RNA minElute clean-up kit	QIAGEN	Cat# 74204
PureLink RNA Mini kit	Thermo Fisher Scientific	Cat# 12183020
μMACS Streptavidin Kit	Miltenyi Biotec	Cat# 130-074-101
Taqman Reverse Transcriptase Reagents	Thermo Fisher Scientific	Cat# N8080234
KAPA RNA HyperPrep kit	Roche	Cat# 08098107702
KAPA mRNA HyperPrep kit	Roche	Cat# 08098123702
Deposited Data		
Raw FASTQ files and unscaled bigwig files available at NCBI's Gene Expression Omnibusunder	This study	GEO: GSE165368
Sequencing data are available under GEO number GSE165368	This study	GEO: GSE165368
Images available at Mendeley	This study	https://doi.org/10.17632/x7hcmjjpj7.1

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Models: Cell Lines		
HEK293T	The Francis Crick Institute Cell Services	N/A
MRC5VA	The Francis Crick Institute Cell Services	N/A
MRC5VA CPSF73 (dTAG)	This study	N/A
MRC5VA shHSF1	This study	N/A
MRC5VA shCtrl	This study	N/A
Experimental Models: Organisms/Strains		
S. cerevisiae (strain BY4741, MATa, his3D1, leu2D0, met15D0, ura3D0)	Euroscarf	BY4741(Y00000)
Oligonucleotides		
All oligonucleotides used in this study are listed in Table S4	This paper	N/A
Recombinant DNA		
pLKO.1 shHSF1 TRCN0000007480	GE healthcare Life Sciences	Cat# RHS3979-201739753
pLKO.1 shCtrl	(Williamson et al., 2017)	N/A
pLP1 RRE (GAG POL)	Kind gift from Simon Boulton	N/A
pLP2 RVE (REV)	As above	N/A
pLP/VSVG (VSVG)	As above	N/A
GW223-pX330A-sgX-sgPITCh	Kind gift from Andreas Mayer	N/A
Software and Algorithms		
Cutadapt v1.9.1	(Martin, 2011)	https://cutadapt.readthedocs.io/en/stable/index.html
STAR v2.5.2a	(Dobin et al., 2013)	https://github.com/alexdobin/STAR
DESeq2	(Love et al., 2014)	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
GenomicAlignments	(Lawrence et al., 2013)	https://bioconductor.org/packages/release/bioc/html/GenomicAlignments.html
Picard v2.1.1	http://broadinstitute.github.io/picard	http://broadinstitute.github.io/picard
BEDtools v2.27	(Quinlan and Hall, 2010)	https://bedtools.readthedocs.io/en/latest/
KentTools	(Kent et al., 2010)	http://hgdownload.soe.ucsc.edu/admin/exe/
deepTools v2.5.3	(Ramírez et al., 2014)	https://www.deeptools.readthedocs.io/en/develop/#
Trim Galore! v0.4.4	N/A	https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
SAMtools v1.9	(Li et al., 2009)	http://www.htslib.org/
MaxQuant v1.6.15.0	(Cox and Mann, 2008)	http://www.maxquant.org/
Other		
High glucose DMEM	Thermos Fisher Scientific	Cat# 11965118
4-15% TGX gels	BioRad	Cat# 56711084
Complete EDTA-free protease inhibitor cocktail	Sigma-Aldrich	Cat# 05056489001
PhosSTOP™	Sigma-Aldrich	Cat# 04906837001
Nitrocellulose membrane	GE healthcare Life Sciences	Cat# 10600002
SuperSignal™ West Pico PLUS ECL reagent	Thermo Fisher Scientific	Cat# 34577
Micro Bio-Spin™ P-30 Gel Columns	BioRad	Cat# 7326223
iTaq™Universal SYBR® Green Supermix	BioRad	Cat# 172-5124
BaseMuncher	Expedeon	Cat# BM0100
AMPureXP beads	Beckman Coulter	Cat# A63881
TRIzol™ Reagent	Thermo Fisher Scientific	Cat# 15596026

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Lipofectamine 3000	Thermo Fisher Scientific	Cat# L3000015
Lyticase	Sigma-Aldrich	Cat# L2524
Immobilon-PVDF membrane	Merck Life Science	IPFL00010
Intercept(PBS) Blocking Buffer	LI-COR	Cat# 927-70001
QuickExtract DNA Extraction solution	Epicentre Technologies/ Lucigen	Cat# QE09050
ExoSapIT	Applied Biosystem	Cat# 78201.1.ML
Q5 High Fidelity 2X Master Mix	New England BioLabs	Cat# M0492S
Fast Flow Q Sepharose	Cytiva	Cat# 17051001
Dynabeads MyOne Streptavidin T1	Invitrogen	Cat# 65601
Amicon Ultra 100kDa NMWCO	MerckMillipore	Cat# UFC8100

RESOURCE AVAILABILITY

Lead contact
Relevant material and any information required to reanalyze the data reported in this paper are available from the lead contact, Jesper Svejstrup, at jsvejstrup@sund.ku.dk.

Materials availability
Cell lines generated in this study are available from the lead contact without restriction.

Data and code availability
Raw FASTQ files and unscaled bigwig files were deposited to the NCBI's Gene Expression Omnibus under accession number GEO: GSE165368. The original images of the study are at Mendeley https://doi.org/10.17632/x7hcmjjpj7.1. Sequencing data have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GEO: GSE165368 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE165368). The code for the mathematical modelling of transcription is available at https://doi.org/10.5281/zenodo.5814979 and also at Github https://github.com/FrancisCrickInstitute/babs_uv_polymerase/releases/tag/v0.1.0.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines and culture conditions
HEK293T (human embryonic kidney epithelial, female origin), MRC5-VA (human fetal lung fibroblast, male origin) and derived cell lines were cultured in high glucose DMEM (Thermo fisher scientific) at 37°C with 5% CO_2. Culture media were supplemented with 10% v/v FBS, 100U/ml penicillin and 100mg/ml streptomycin. All cell lines were confirmed to be mycoplasma-free by the Francis Crick Institue Cell Services. For heat shock (HS), cells were subjected to immediate HS by replacing the media with pre-warmed media at 43°C and kept in an incubator at 43°C for the indicated period of time. All the sequencing experiments were carried out in conditioned medium.

METHOD DETAILS

Generation of stable RNAi cell line
Viral particles were generated by co-transfecting HEK293T cells with 1 μg third-generation lentiviral packaging vectors (pLP1, pLP2 and pLP/VSVG, a kind gift from the Boulton lab (Crick Institute)) and 1 μg lentiviral vectors (pLKO.1) containing either a HSF1 specific shRNA (TCRN0000007480) or a Non-targeting Control shRNA (shCtrl) using polyethylenimine (PEI) (Polysciences, Inc., 23966-1) as a transfection reagent (3:1 PEI/DNA ratio). 48h from transfection, viral particles were collected and used for transduction of MRC5-VA. Hexadimethrine bromide (Sigma-Aldrich, H9268) was used to increase infection efficiency. Cells were selected in 1 μg/ml puromycin and the knockdown efficiency was tested by western blot.

Generation of CPSF73 (dTAG) cell line
The CPSF73 gene was tagged with a C-terminal FKBP degron by the MMEJ homology method described by Nabet et al. (2018). CPSF73 homology arms were inserted into the plasmid pCRIS-PITChv2-dTAG-BSD (BRD4) according to the protocol provided by Nabet et al. The gRNA (GGCTGCACAGAGACTGTACG) was inserted into the Cas9 vector GW223-pX330A-sgX-sgPITCh. Both

plasmids were a kind gift from the Mayer Lab. The plasmids were transfected into MRC5-VA cells using Lipofectamine 3000 (Invitrogen, ThermoFisher Scientific) and selected with 5ug/ml Blasticidin (Cambridge Bioscience). Single clones were isolated and the genomic DNA prepared using QuickExtract (Epicentre Technologies) and ExoSapIT (Applied Biosystems). The region surrounding the tag insertion site was amplified with primers ATTAGGACCGTGCTGCTGTC and CCTGTAACACCCACGAGGAC using Q5 polymerase (New England bioLabs). Clones with a PCR product of 3036 bp were expanded and analysed by Western blot using either antibodies to HA (ab 236632 Abcam Ltd) or CPSF73 (A301-090 Bethyl Laboratories Inc). Genomic DNA was sequenced to ensure correct insertion of the degradation tag. Homozygous clones were treated with dTAG-7 (Tocris, Bio-Techne Ltd.) at 250nM for 2h and the degradation of CPSF73 confirmed by Western blot.

Clonogenic survival assay
shHSF1 stable cells were seeded at 400 cells/well and shCtrl cells were seeded at 200 cells/well into 6-well plates. Cells were exposed to HS for the indicated amount of time and then kept in incubator at 37°C for 11 to 13 days.

Colonies were fixed by 4 % (v/v) formaldehyde and stained with a 0.1 % (w/v) crystal violet solution. Crystal violet was extracted with a 10% acetic acid solution and quantified. Colonies from three biological replicates (each seeded into triplicate wells) were used for the quantification.

Western blot
For whole cell extracts, cells were resuspended and incubated for 30 minutes at 4°C in lysis buffer (250 U/ml BaseMuncher Benzonase (Expedeon, BM0100), 150 mM NaCl, 20mM TRIS-HCl pH 8, 0.1% (v/v) NP-40, 10% (v/v) Glycerol, 1.5 mM $MgCl_2$) supplemented with protease (Sigma-Aldrich, 05056489001) and phosphatase (PhosSTOP™, Sigma-Aldrich, 04906837001) inhibitors. The lysates were spun at 14000 RPM at 4°C for 15 minutes and the supernatant was used for western blot analysis. Proteins were separated on 4%–15% TGX gels (BioRad, 5671084) and transferred either to a nitrocellulose membrane (GE Healthcare Life Sciences, 10600002) or to a PVDF membrane (Merck Life Science, IPFL00010). Membranes were blocked respectively in 5% (w/v) skimmed milk in PBS supplemented with 0.1% (v/v) Tween20 (PBST) or in Intercept (PBS) blocking buffer (LI-COR, 927-70001) for 1 h at room temperature. Incubation with primary antibodies was carried out in PBST or Intercept (PBS) blocking buffer supplemented with 0.1% (v/v) Tween20 overnight at 4°C. Membranes were washed several times in PBST, incubated with either with HRP-conjugated or LI-COR fluorescent dye-conjugated secondary antibody diluted respectively in PBST or Intercept (PBS) blocking buffer supplemented with 0.1% (v/v) Tween20 for 45 min at room temperature and washed several times in PBST. Signal detection was obtained either using SuperSignal West Pico PLUS (Thermo Fsher Scientific, 34577) Substrate ECL reagent or Odyssey CLx imaging system (LI-COR). All antibodies used are listed in the key resources table.

RT-qPCR
Total RNA was extracted using RNeasy mini kit (QIAGEN, 74106) following manufacturer instructions including an on-column DNase digestion (QIAGEN, 79254). 1 to 1.5 μg RNA was used for Reverse transcription with TaqMan Reverse Transcription Reagents (Thermo Fisher Scientific, N8080234) according to manufacturer instructions. For the 4SU-labelled mRNA, approximately 30 to 100ng of material were used for the reverse transcription. For detection of nascent RNA, random hexamers were used in the reverse transcription reaction and for the detection of mRNA, oligo dT primers. cDNA was amplified on a CFX384 Touch Real-Time PCR Detector (BioRad 1855485) using iTAq Universal SYBR Green Supermix (BioRad, 172-5124) with the following conditions: 5 min denaturation at 95°C and 39 cycles of 10 s denaturation at 95°C, 10 s annealing at 58°C, and 20 s extensions at 72°C. Primers amplifying GAPDH mRNA were used for normalization purposes. All primer sequences are listed in Table S4.

Nucleus-cytoplasm fractionation
10×10^6 cells were exposed to HS for two hours, washed with cold PBS twice and collected. The cells were then resuspended in 200 μl cytoplasmic lysis buffer (10 U RNase inhibitor (Thermo Fisher Scientific, N8080119), 10 mM Tris-HCl pH 7, 150 mM NaCl, 0.15% NP40) and incubated in ice for 5 min. The samples were then layered over 500 μl sucrose buffer freshly made (20 U RNase inhibitor, 10 mM Tris-HCl pH 7, 150 mM NaCl, 25% sucrose) and centrifuged at 16,000 g for 10 min at 4 °C. The supernatant was collected and used for cytoplasmic RNA extraction with RNeasy mini kit (QIAGEN, 74106) after adding 3.5X RLT Buffer (QIAGEN, 79216) and 2.5X ice cold 100% ethanol. The nuclei pellets were washed with 200 μl sucrose buffer and then used for RNA extraction with RNeasy mini kit (QIAGEN, 74106) following manufacturer instructions including an on-column DNase digestion (QIAGEN, 79254).

TT$_{chem}$-seq
The TT$_{chem}$-seq was carried out as described in Gregersen et al. (2020) in at least two biological replicates. 8×10^6 cells were exposed to HS and the RNA was labelled *in vivo* with 1 mM 4SU (Glentham Life Sciences, GN6085) for 15 minutes prior to the addition of TRIzol (Thermo Fisher Scientific) that was used to stop the reaction at the desired time point. RNA was extracted according to manufacturer instructions.

As a control for equal sample preparation, we spiked-in 4-thiouracile (4TU) labelled RNA from *S. cerevisiae* (strain BY4741, MATa, his3D1, leu2D0, met15D0, ura3D0). The yeast culture was grown overnight in YPD medium, diluted to OD600 of 0.1 and grown to

mid-log phase (OD600 of 0.8) and labelled 5 min with 5 mM 4TU (Sigma-Aldrich, 440736). Total RNA was extracted with PureLink RNA Mini kit (Thermo Fisher Scientific, 12183020) following the enzymatic protocol.

100 µg of human 4SU labelled RNA was spiked-in with 1 µg of 4TU-labelled yeast RNA and brought to a total volume of 100 µl with water. The mix was fragmented by adding 20 µl 1M NaOH and incubating in ice for 20 min. 80 µl of 1M Tris-HCl pH 6.8 were added to stop the fragmentation and samples were cleaned up twice with Micro Bio-Spin P30 Gel Columns (BioRad, 7326250) following manufacturer instructions. Biotinylation of 4SU and 4TU residues was carried out in a total volume of 250 µl 10 mM Tris-HCl pH 7.4 and 1 mM EDTA, containing MTSEA biotin-XXlinker (Biotium, BT90066) for 30 min at room temperature in the dark. The RNA was then purified by phenol:chloroform extraction, denatured 10 min at 65°C and added to 200 µl µMACS Streptavidine MicroBeads (Miltenyi Biotec, 130-074-101). After 15 min incubation at room temperature, the mix was loaded to a µColumn in the magnetic field of a µMACS magnetic separator. The beads were washed twice in a buffer containing 100 mM Tris-HCl pH7.4, 10 mM EDTA, 1M NaCl and 01% Tween20. Biotinylated RNA was eluted twice in 100 mM DTT (200 µl final volume) and cleaned up with the RNeasy MinElute kit (QIAGEN, 74204) using 1050 µl 100%ethanol and 750 µl RLT buffer to precipitate RNA <200nt.

DRB/TT$_{chem}$-seq
The DRB/TT$_{chem}$-seq was carried out as described in Gregersen et al. (2020) in two biological replicates. One 150 mM dish was used for each sample, except the samples not released from DRB which required two 150 mM dishes. Cells were incubated in 100 µM DRB (Sigma-Aldrich, D1916) for 1,5 hours. For the HS treatment, the medium in the cells was replaced with medium pre-warmed at 43°C supplemented with 100 µM DRB and the incubation was continued in a 43°C incubator for additional 2 hours. The cells were then washed twice in PBS and fresh medium at 43°C was added to restart transcription. 4SU labelling was carried out for 10 min prior to the addition to TRIzol to stop the reaction at the desired time point between 10 and 40 min after DRB release. Samples were then processed following the TT$_{chem}$-seq protocol.

FP/TT$_{chem}$-seq
Cells were kept at 43°C for 2h, the untreated samples (not treated with FP) were labelled with 4SU for 10 min and collected in TRIzol exactly after 2 hours of HS. The remaining samples were kept at 43°C, FP (Sigma-Aldrich, F3055) was added to the cells to a final concentration of 300nM and the 10 min 4SU labelling was timed to obtain five time points between 15 and 40 minutes after FP addition. The cells were disrupted in TRIzol at the desired time point and the samples were then processed following the TT$_{chem}$-seq protocol.

4SU/mRNA-seq
Cells were exposed to HS for 1h and 45 min, RNA was labelled in vivo with 0.7 mM 4SU for 90 min in HS conditions. The cells were harvested by scraping in pre-warmed PBS and span down. The pellets were then snap-frozen in liquid nitrogen and quickly defrosted for RNA extraction with RNeasy mini kit (QIAGEN, 79254) following manufacturer instructions including an on-column DNase digestion (QIAGEN, 79254). 100 µg of 4SU labelled RNA were used for the biotinylation reaction that was carried out in a total volume of 250 µl 10 mM Tris-HCl pH 7.4 and 1 mM EDTA, containing MTSEA biotin-XXlinker (Biotium, BT90066) for 1h at room temperature in the dark. The purification and pull-out of biotinylated RNA was carried out as in the TT$_{chem}$-seq protocol. The material was either used for RT-qPCR or library preparation.

Library preparation
For all the TT$_{chem}$-seq experiments and its derivatives (DRB/ and FP/TT$_{chem}$-seq), 30 to 100 ng of 4SU/4TU labelled RNA were used for library preparation with the KAPA RNA HyperPrep kit (Roche, 08098107702) following manufacturer instructions. The fragmentation step was omitted and the RNA, resuspended in FPE Buffer, was denatured at 65°C for 5 min. The libraries of the DRB/TT$_{chem}$-seq and the FP/TT$_{chem}$-seq experiments were prepared with modifications as previously described (Tufegdzic Vidakovic et al., 2020). Briefly, the two SPRI bead purifications were carried out, respectively, with a bead-to-sample volume ratio of 0.95x and 1x. For the 4SU/mRNAseq experiment, 200ng of purified 4SU labelled RNA per sample were used to prepare polyA+ mRNA libraries with KAPA mRNA HyperPrep kit (Roche, 08098123702) following manufacturer instructions. The libraries were then sequenced with single end 75bp reads on the Hiseq2500, with 50,000,000 reads per sample.

Mass spectrometry analysis
Cells were exposed to HS for 2 hours and let recover at 37°C either for 3 hours (T1) or 6 hours (T2). Cell pellets were lysed with lysis buffer (1% (w/v) Sodium Deoxycholate, 100 mM Tris, pH 8.5) and incubated for 10 min at 95°C followed by sonication using a Bioruptor pico (30 cycles, 30s on/off, ultra-low frequency). Samples were cleared by centrifugation reduced with 5 mM of TCEP for 15 min at 55°C and alkylated with 20 mM CAA for 30 minutes at RT. 100 µg of reduced and alkylated lysates were digested adding Trypsin/LysC 1:100 (enzyme/protein ratio). Peptides were cleaned up using StageTips packed with SDB-RPS resin, and resuspended in 50 µl TEAB 100 mM, pH 8,5. 50 µg of each sample was labeled with 0.5 mg TMTpro labeling reagent according to the manufacturer's instructions. Labeled peptides were combined and cleaned up using C18-E (55 µm, 70 Å, 100 mg) cartridges (Phenomenex). Labeled desalted peptides were resuspended in Buffer A* (5% acetonitrile, 1% TFA), and 30 µg fractionated into 24 fractions, by high-pH fractionation (Kulak et al., 2017).

Each fraction was measured on an EASY-nLC 1200 ultra-high-pressure system coupled through a nano-electrospray source to a Tribrid Eclipse mass spectrometer (Thermo Fisher Scientific). Peptides were loaded in buffer A (0.1 % formic acid) onto a 25 cm Aurora HPLC column (Ionopticks) and separated with a non-linear gradient of 5 – 44 % buffer B (0.1 % formic acid, 80 % acetonitrile) at a flow rate of 400 nl/min over 91 min. The column temperature was kept at 40° C. Data acquisition switched between a full scan (120 K resolution, 50 ms max. injection time, AGC target 100%) and data-dependent MS/MS scans (50K resolution, 120 ms max. injection time, AGC target 200%) over a fixed cycle time of 3 s. The isolation window was set to 0.7 (m/z), and normalized collision energy to 35. Precursors were filtered by precursor envelope fit of 70% with fit window of 1.2 m/z, charge state of 2-6 and multiple sequencing of peptides was minimized by excluding the selected peptide candidates for 45 s.

Purification of bovine RNAP II from calf thymus

BOVINE RNAP II was prepared as previously described with minor modification (Zatreanu et al., 2019). Unless otherwise noted, all steps were completed at 4° C. Calf thymus was homogenized for 3 min in buffer A (50 mM Tris, pH 7.9 at 4 °C, 10 μM $ZnCl_2$, 10% glycerol, 1 mM EDTA, protease inhibitors) using a 2L blender (Waring). The homogenized material was centrifuged and the supernatant filtered through two layers of mesh nylon filter. A 10% solution of polyethyleneimine, pH 7.8, was added to a final concentration of 0.05%, and the material was stirred for 30 min, then centrifuged for 30 min at 12,000g. The resulting pellets were redissolved in buffer B (50 mM Tris, pH 7.9 at 4° C, 10 μM $ZnCl_2$, 10% glycerol, 150 mM $(NH_4)_2SO_4$, 1 mM EDTA, protease inhibitors). After centrifugation, the supernatant was loaded on a 120 mL Fast Flow Q Sepharose column (Cytiva 17-0510-01), equilibrated in buffer B, by using a peristaltic pump at 5 ml/min. The column was washed with three column volumes of buffer B, followed by elution with buffer C (50 mM Tris, pH 7.9 at 4° C, 10 μM $ZnCl_2$, 500 mM $(NH_4)_2SO_4$, 1 mM EDTA, protease inhibitors). The material was further purified using a 5 mL gravity flow column of 8WG16 (αRPB1 CTD) antibody-coupled Sepharose. The input was loaded overnight using a peristaltic pump at 0.1 ml/min. After application of the input material, the antibody column was washed with ten column volumes of buffer C, sealed, and allowed to equilibrate to room temperature (20–25° C) for 15 min. RNAP II was eluted using elution buffer (40% 1,2 propanediol, 500 mM $(NH_4)_2SO_4$, 50 mM Tris pH7.8, 10 μM $ZnCl_2$, 1 mM EDTA), collecting 4x 10 mL fractions at room temperature. The RNAPII containing fractions were dialyzed against dialysis buffer (50 mM Tris pH 7.8, 10 μM $ZnCl_2$, 5 mM DTT, 10% glycerol, 150 mM $(NH_4)_2SO_4$) and concentrated using a 100-kDa cut-off Amicon concentrator (MerckMillipore, UFC8100) to a final concentration of around 1 mg/ml.

In vitro transcription assay

The elongation complex was assembled as previously described with minor modification (Han et al., 2017). First, 5 pmol of pre-annealed RNA:DNA (template strand) hybrid was mixed with equimolar amount of RNAPII, followed by adding 10 pmol of 5' biotin-labelled non-template strand DNA. The assembled elongation complex was then immobilized onto streptavidin beads (Dynabeads MyOne Streptavidin T1 (Invitrogen, 65601) and washed with transcription buffer (TB) containing 20 mM Tris-HCl pH 7.5, 100 mM NaCl, 8 mM $MgCl_2$, 10 μM $ZnCl_2$, 10% glycerol, 2 mM DTT, then with TB/0.1% Triton, TB/0.5 M NaCl and finally TB.

A long transcription template (around 1 kb) was PCR-amplified from RPL41 termination region by using primers containing BbsI restriction digestion sites. The PCR product was then digested with Bbs I and ligated to the pre-assembled RNAPII elongation complex by T4 DNA ligase.

After ligation, the transcript was first pulse radiolabelled with alpha-32P UTP, then chased with 100 μM NTPs in transcription buffer under 37° C vs 43° C. Reactions were stopped at indicated timepoint by mixing with 25 mM EDTA (final concentration) and separated into supernatant and bead fraction. The bead fractions were resuspended in 8 μl of loading buffer (1× TBE, 8 M urea) and boiled for 5 min at 95° C, and RNAs in the supernatant fractions were ethanol precipitated and resuspended in 8 μl of loading buffer. The samples were subjected to 6% (w/v) denaturing PAGE (8 M urea) and the gel was visualized by phosporimaging using a Typhoon scanner (GE healthcare).

RNA: UUUUUCGACCAGGA

hplc purified from IDT

Template DNA: containing 5' phosphor modification in order to ligate to long transcription templates

GTTGTGCAGGCCGGGTGCGGCCGCCCGTGTGGAGATGGGTGAGAGATGTTGAGGATCCTGGTCGTTTCCTATAGTTTGTTTCCTGAGTAAGTCTTCATCG

page purified from IDT.

Non-template DNA:

CTAGCGATGAAGACTTACTCAGAAACACGACTTAGGTAGACGACCAGGATCCTCAACATCTCTCACCCATCTCCACACGGGCGGCCGCACCCGGCCTGCA

page purified from IDT.

QUANTIFICATION AND STATISTICAL ANALYSIS

TT_{chem}-seq alignment and quantification

Illumina adapter sequences were trimmed from reads using cutadapt v1.18 (Martin, 2011) with the following settings: -a AGATCG GAAGAGC -e 0.1 -q 10 -m 25 -O 1. Reads were aligned against the Homo sapiens GRCh38 genome build using STAR v2.5.2a (Dobin et al., 2013) with Ensembl release 86 transcript annotations. A yeast spike-in was used to normalise sequencing depth between

samples. Scale factors were derived by aligning reads to the Saccharomyces cerevisiae sacCer3 genome build and generating a gene counts matrix using the standard Ensembl gene definitions. The counts matrix was then passed to DESeq2's estimateSizeFactors function (Love et al., 2014). Note that yeast scale-factors were applied to all TT_{chem}-seq datasets except for the mRNA and the CPSF73-dTAG datasets, where the default human factors were used.

Differential expression

Human gene abundance estimates were calculated using the 3' 50% of the gene-level intervals using the GenomicAlignments (Lawrence et al., 2013) package's summarizeOverlaps function (mode="Union", ignore.strand=FALSE). Yeast derived scale replaced the default size factors from the analysis of the human data. Differential expression analysis was conducted between treatment groups using the Bioconductor package DESeq2 (Love et al., 2014). Genes were filtered for significance based on an FDR<=0.01, a fold-change of +/- 2, and a base-mean expression of >=10.

BAM files merging

For the purposes of visualisation, genome alignment BAM files were merged across biological replicates, sorted and indexed using Picard v2.1.1 (http://broadinstitute.github.io/picard).

BigWig files

BigWig files were generated by converting BAM files to bedGraph format using BEDtools' genomecov function (Quinlan and Hall, 2010) in a strand-specific manner. Where applicable, yeast scale factors were applied to normalize for differences in library size. bedGraph files were in turn converted to bigWig format using the bedGraphToBigWig function from the KentTools package (Kent et al., 2010). All the genome-wide visualizations are strand-specific and represent a merge of at least two biological replicates.

Read depth profiles

Read depth profiles were created directly from the scaled bigwig files using deepTools v.2.5.3 (Ramírez et al., 2014) for protein coding genes from standard chromosomes that were between 5-300kb in width. Additional gene sets were created by sub-setting genes based on their genomic width: >=5kb, >=30kb, >=60kb, >=90kb and >=120kb and also on differential expression status.

For the metagene profiles the computeMatrix function was used to calculate read coverage over the regions 5kb upstream of the TSS, the gene-body divided into 100 equally sized bins and the 5kb downstream of the TSS (computeMatrix scale-regions -m 15000 -a 5000 -b 5000 –binSize 100). For the CPSF73-dTAG dataset, the metagene profiles were additionally transformed to bring the curves into alignment and better highlight relative changes in expression across the genic interval. This was done by scaling each curve by the sum of sense coverage upstream of the TSS, effectively aligning the y-axes based on the promoter reads which is presumed to be background signal.

For the extended TSS profiles the computeMatrix function was used to calculate read coverage over the regions 2kb upstream to 120kb downstream of the TSS (computeMatrix reference-point –referencePoint TSS -a 120000 -b 2000 –binSize 100).

Coverage matrices were imported into R for further processing and visualisation. The mean read-depth at each bin was calculated across contributing genes having trimmed the 5% most extreme values.

Travel ratios

Travel ratios were calculated as in Rahl et al. (2010). Briefly, each gene was divided into i) a promoter-proximal bin −30 bp to +300 bp around its TSS and ii) a gene body bin extending to the TTS. The trave ratio is the ratio of RNAPII density in the promoter-proximal bin to that in the gene body. A count of 1 was added to the promoter and gene body counts prior to calculating density in order to mitigate for zero and infinite travel ratios. All protein coding genes from standard chromosomes that were between 5-300kb in width were considered for the analysis, but subsets were taken corresponding to genes significantly up or down regulated between heat-shock and control samples based on the DESeq2 analysis.

DRB/TT_{chem}-seq: wave-front and elongation rate

DRB/TT_{chem}-seq analysis was conducted as in Gregersen et al. (2020). Briefly, bp-resolution read-depth coverage was calculated over intervals representing non-overlapping, protein-coding genes at least 200kb in width, extended -2kb:+200kb around the TSS region. Scale factors derived from the yeast spike-in were used to normalise counts for library size before being converted to read-counts-per-million (RPM). Meta-profiles were created by taking a trimmed mean across genes at each bp location. A smoothing spline was fitted to each meta-profile using R's smooth.spline function (spar=0.3). Wave peaks were subsequently calculated as the maximum point on the spline for each sample, ensuring that wave peaks advanced with time. Elongation rates in kb/min were estimated by fitting a linear model to the wave peak positions of all samples (i.e. 0, 10, 20, 30, and 40 minutes post release) as a function of time.

FP/TT_{chem}-seq: wave-front and elongation rates

Extended TSS (-2k:+200kb) sense read depth profiles were smoothed using a spline. The slope of the transcriptional wave indicative of polymerase speed was defined as the point at which the curve was approximately linear. This was estimated by calculating the gradient of the first differential at the point the second differential crossed zero.

Exemplar single gene level wave positions were calculated from scaled read-depth coverages converted to densities over the region TSS:+200kb. The wave was defined at the position at half the height of the maximum observed density within the first 150kb. Elongation rates in kb/min were estimated by fitting a linear model to the wave peak positions of all samples (i.e. 0, 15, 20, 25, 40 and 40 minutes post release) as a function of time.

FP/TT$_{chem}$-seq: read-through-ratio
The read-depth-ratio was defined as the ratio of read density of terminal exons divided by the read density in the downstream 20kb region. Only protein coding genes from standard chromosomes were considered. A single representative transcript was selected per gene, prioritising better transcript support level then larger transcript width.

TTchem-seq: defective transcript elongation after heat shock
Non-overlapping, protein-coding genes were divided into two intervals: a 5' region representing the first 5%' of a gene's width downstream of the TSS and a 3' region representing the final 25% of the gene up to the TES. Sense reads were counted over these intervals and converted to densities by dividing by interval width. Genes containing an interval with zero counts in either the HS or ctrl condition were discarded, leaving n=4876 genes. A ratio of each gene's 3' and 5' interval densities was calculated for each treatment condition. Finally, a ratio-of-ratios between the HS and ctrl samples was calculated: HS(3'/5')/ctrl(3'/5'). The data were stratified based on gene width into 6 groups (<15kb, 15<30kb, 30<60kb, 60<90kb, 90<120kb, >=120kb) and the number of genes showing at least a 1.5-fold reduction in each width group was recorded.

mRNA-seq first exon enrichment analysis
This analysis attempted to measure the enrichment of expression of short isoforms within HS or Ctrl treatment groups by comparing the number of reads mapping to first exons relative to all remaining exons. A single transcript was selected per gene (Gencode Basic v33) as being representative of each locus, giving priority to transcripts with the strongest support level and then by longest genomic interval. Genes comprising just a single exon were discarded. Exon level read counts were generated from BAM files using the GenomicAlignments package's summarizeOverlaps function in a strand-specific manner (ignore.strand=FALSE, inter.feature=TRUE). The matrix was used to generate sample scale factors using DESeq's estimateSizeFactorsForMatrix function, with default settings. The exon level counts matrix was then recoded such that there were 2 paired columns per samples representing i) first exon counts and ii) remaining exon counts. This new matrix was further analysed using DESeq2. Size factors generated from the above all exon analysis were applied to this new matrix such that the same size factor for a given sample was applied to both the "first exon" and "remaining exon" columns. DESeq2's Wald test was then run with the following design formula: $\sim$ HS_status + HS_status:pair + HS_status:sample. The resulting logFC from the test equates to the ratio of first/remaining exon counts for each treatment, i.e. (HS first/ HS remaining) / (CTRL first/CTRL remaining). Thus, a positive logFC indicates a preference for shorter isoforms in the HS relative to the CTRL and a negative logFC indicates a preference for shorter isoforms in the CTRL relative to the HS. Figure 3B depicts these logFC values for all expressed genes, i.e. a count of at least 1 in i) exon1 and ii) remaining exons across all HS and ctrl samples (n=8744).

Genomic features analysis
The shrunken-logFC estimates from the DESeq2 analysis of the TT$_{chem}$-seq MRC5-VA differential expression analysis were used to assess treatment-specific differences associated with various genomic features. In each case, the logFC estimates were stratified on gene differential expression status in the HS samples relative to the control and presented as boxplots. Gene width was defined as the genomic interval from TSS to TES. U1 motifs and polyadenylation site (PAS) analysis was conducted as in Krajewska et al. (2019). Briefly, U1 motifs ("GGTGAG", "GGTAAG" or "GTGAGT") and PAS motifs ("AATAAA") were counted within the genic intervals. The "U1/PAS" ratio was defined as the ratio of these counts, i.e. (GGTGAG + GGTAAG + GTGAGT)/(AATAA). Significance differences in the distribution of logFCs between HS down-regulated genes and all other genes was assessed using a Wilcoxon rank sum test.

Gene-width was plotted against the shrunken-logFC estimates from the DESeq2 analysis of the HS vs ctrl samples for all protein-coding genes from standard chromosomes. Genes significantly up and down regulated in the HS samples were highlighted. A loess line was fitted to the data to highlight the trend for down-regulated genes to be generally wider.

Abundance estimates for the HS and ctrl samples were plotted against the mean-centred $\log_2$ gene-widths. Abundance was rlog transformed to minimises differences between samples for genes with small counts and to normalise with respect to library size. Linear regression was applied to calculate gene-width coefficients defined by the slope of each line. A t-test was then used to assess the significance of differences in the coefficients due to the HS treatment.

mRNA-seq first exon enrichment analysis (U2OS cells)
Published RNA-seq data from untreated U2OS cells and those exposed to heat stress (43°C) for 4 hours were downloaded from ArrayExpress under accession E-MTAB-2349. Reads were aligned to the hg38 genome as per the previously mentioned TTchem-seq methodology. First-exon enrichment analysis was conducted as for the MRC5-VA cells.

HEK293 Pro-seq data analysis

Published HEK293 Pro-seq data were downloaded from the NCBI Gene Expression Omnibus (GSE112379) (Aprile-Garcia et al., 2019). Reads were trimmed to remove adapter sequence using trimGalore v0.4.4 (trim_galore –quality 20 -e 0.05 -a TGGAATTCTCGGGTGCCAAGG –length 10) before alignment to GRCh38 using HiSat2 v2.0.4 allowing for up to 5 primary alignments (k=5). Aligned reads were filtered to remove low quality alignments using SAMtools (Li et al., 2009) [-q 20] and used to create strand-specific BigWig files of the 5' positions of each read. Metagene profiles were created covering the region 500bp downstream of the TSS to the TES to focus attention on the nascent RNA differences over gene bodies. No correction for differing library size between samples was performed.

Correlation of HS expression and U1 induced PCPA sites

Gene Set Enrichment Analysis (GSEA) was used to assess whether there was a significant, concordant difference in HS responsive gene expression in genes previously shown to exhibit premature cleavage and polyadenylation (PCPA) sites after 8h of U1 inhibition (Oh et al., 2017). Genes from the TTchem-seq differential expression analysis were ranked ordered based on their DESeq2 Wald test statistic generated from the comparison of heat-shock to untreated samples. The expression related enrichment of U1 responsive genes within that list was assessed using the Bioconductor package "fgsea" (Korotkevich et al., 2019).

Computer modeling of transcription

Adapting the approach previously described (Tufegdzic Vidakovic et al., 2020), we simulated polymerase dynamics in response to a discrete set of events (including initiation, termination, processivity errors - for a full list see Tufegdzic Vidakovic et al. [2020]) happening in continuous time in model cells expressing a mix of short (5kb), medium (63kb) and long (100kb) genes.

To model the behaviour of transcription after HS, where transcription ends much earlier in genes on a genome-wide scale, we replaced the stochastic damage-and-repair mechanism from Tufegdzic Vidakovic et al. (2020) with a fixed site at 10 kb into the model genes upon reaching which, polymerases will terminate and be recycled for new transcription.

In the model, a dynamic equilibrium is first achieved under the normal conditions of cells having the mix of active short (5 kb), medium (63 kb) and long (100 kb) genes as in Tufegdzic Vidakovic et al. (2020). New HS-induced transcription termination sites (TSSs) are then for simplicity introduced at 10 kb for all genes, at which point longer genes start to lose polymerases at both the new 10 kb TSS and at the previously normal TTS (at 63 or 100 kb), until the region between them is devoid of active polymerases. The model assumes that the RNAPII pool is limiting and that transcription initiation is attempted following a Poisson process whose rate is proportional to the number of free polymerases, with the constant of proportionality being such that expected times between transcription events on a gene is 2.5s in normal conditions. We then plot how the average density (100 simulation runs, arbitrary scale consistent across runs and gene-lengths) of active polymerases changes across time from the instantaneous introduction of the 10 kb TSS. The increase in transcription rate, enabled by the larger pool of free RNAPII, is most easily deduced from the initiation point on the Y-axis. The Julia code to recreate this is at Github (https://github.com/FrancisCrickInstitute/babs_uv_polymerase/releases/tag/v0.1.0).

Data analysis, mass spectrometry
Short peptide database

A library of short peptide sequences for mass-spectrometry screening was created based on Gencode (v33) non-overlapping, protein-coding transcripts from standard chromosomes. A simple, relatively straightforward approach was taken to avoid the considerable complexity of assembling a brand-new transcriptome, gene-by-gene, based on the thousands of individual new mRNA transcripts that had been revealed by our mRNA analysis. In this virtual library, translation starts at the canonical ATG of every gene considered in the analysis and is then allowed to proceed through the first exon, into the adjacent intron, and up to the first stop codon (examples shown in Figure S9). It thus assesses only incompletely but proteome-wide, the likelihood that the new mRNAs produced after HS are translated efficiently.

Raw mass spectrometry data were analyzed with MaxQuant (v1.6.15.0). Peak lists were searched against the human Uniprot FASTA database combined with 262 common contaminants by the integrated Andromeda search engine. The false discovery rate was set to 1 % for both peptides (minimum length of 7 amino acids) and proteins.

All statistical analysis of TMT derived protein expression data was performed using the automated analysis pipeline of the Clinical Knowledge Graph (Santos et al., 2021). Protein entries referring to potential contaminants, proteins identified by matches to the decoy reverse database, and proteins identified only by modified sites, were removed, as well as proteins not found in all 9 channels (samples). Reporter ion intensities were $\log_2$ transformed and normalized to adjust the TMT experiment to an equal signal per channel (equal mean in all channels).

Principal component analysis (PCA) was performed and visualized with 2 principal components, showing clear separation between control samples and the other groups. Differentially expressed proteins were identified with two different methods: One-way Anova followed by pairwise comparison posthoc tests (unpaired t-tests) with Benjamini-Hochberg correction for multiple hypothesis (alpha=0.05, Fold Change (FC)=1.5); and SAMR Multiclass test followed by pairwise comparison posthoc tests (two class unpaired tests) with permutation-based FDR correction (FDR = 0.05, s0=1, permutations=250, FC=1) (Tibshirani et al., 2018). Significantly

regulated proteins are colored in red and blue in the volcano plots for up and downregulated hits, respectively. Hierarchical clustering was performed using all significant proteins (up- and down-regulated). The selected proteins were then mean normalized.

Functional enrichment analysis was performed in the Clinical Knowledge Graph [2], applying Fisher's exact test and Benjamini-Hochberg FDR correction, and using Gene Ontology Biological Process (GOBP), Gene Ontology Molecular Function (GOMF), Gene Ontology Cellular Component (GOCC) and Reactome annotations. All proteins quantified in the experiment were used as background, and statistically significant proteins (posthoc tests) were considered foreground. Enrichment analysis is performed for up- and down-regulated proteins separately.