



# Comparative analysis reveals genomic features of stress-induced transcriptional readthrough

Anna Vilborg<sup>a,b,1</sup>, Niv Sabath<sup>c</sup>, Yuval Wiesel<sup>c</sup>, Jenny Nathans<sup>a,b</sup>, Flonia Levy-Adam<sup>c</sup>, Therese A. Yario<sup>a,b</sup>, Joan A. Steitz<sup>a,b</sup>, and Reut Shalgi<sup>c,1</sup>

<sup>a</sup>Department of Molecular Biophysics and Biochemistry, Boyer Center for Molecular Medicine, Yale University School of Medicine, New Haven, CT 06536; <sup>b</sup>Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, CT 06536; and <sup>c</sup>Department of Biochemistry, Rappaport Faculty of Medicine, Technion-Israel Institute of Technology, Haifa 31096, Israel

Edited by Jasper Rine, University of California, Berkeley, CA, and approved July 14, 2017 (received for review July 10, 2017)

**Transcription is a highly regulated process, and stress-induced changes in gene transcription have been shown to play a major role in stress responses and adaptation. Genome-wide studies reveal prevalent transcription beyond known protein-coding gene loci, generating a variety of RNA classes, most of unknown function. One such class, termed downstream of gene-containing transcripts (DoGs), was reported to result from transcriptional readthrough upon osmotic stress in human cells. However, how widespread the readthrough phenomenon is, and what its causes and consequences are, remain elusive. Here we present a genome-wide mapping of transcriptional readthrough, using nuclear RNA-Seq, comparing heat shock, osmotic stress, and oxidative stress in NIH 3T3 mouse fibroblast cells. We observe massive induction of transcriptional readthrough, both in levels and length, under all stress conditions, with significant, yet not complete, overlap of readthrough-induced loci between different conditions. Importantly, our analyses suggest that stress-induced transcriptional readthrough is not a random failure process, but is rather differentially induced across different conditions. We explore potential regulators and find a role for HSF1 in the induction of a subset of heat shock-induced readthrough transcripts. Analysis of public datasets detected increases in polymerase II occupancy in DoG regions after heat shock, supporting our findings. Interestingly, DoGs tend to be produced in the vicinity of neighboring genes, leading to a marked increase in their antisense-generating potential. Finally, we examine genomic features of readthrough transcription and observe a unique chromatin signature typical of DoG-producing regions, suggesting that readthrough transcription is associated with the maintenance of an open chromatin state.**

transcriptional readthrough | stress response | transcription regulation

**S**tress can affect transcription in numerous ways. Gene-specific transcription factors such as CREB (cAMP response element-binding protein) are activated by various stress signaling pathways, causing stress-dependent activation of transcription initiation of target genes (reviewed in refs. 1 and 2). Stress is also known to affect pause release, which is the transition of RNA polymerase II (Pol-II) from a pause site located within 100 bp of the promoter into efficient transcription elongation. A hallmark example of promoter proximal pausing is that of the *Hsp70* gene in *Drosophila*, where pause release is triggered by heat shock, leading to a rapid boost in *Hsp70* transcription (3, 4); this phenomenon has been recently shown to be a major part of the mammalian transcriptional response to heat shock (5). Other transcription-related processes such as chromatin modification (6) and splicing (7) are also regulated by stress (reviewed in refs. 1 and 2).

Recent studies reveal that stress can reduce the efficiency of transcription termination (8–10). Transcription termination involves elongation through the cleavage and polyadenylation (polyA) signal and subsequent cleavage of the nascent RNA, followed by the addition of nontemplated A residues to the 3' end to produce an mRNA. The actual dissociation of Pol-II from the RNA and from the template DNA can occur as much as 10 kb downstream of the polyA signal (11). Typically, the extended RNA is quickly

degraded by exonucleases that access the unprotected 5' end generated by cleavage at the polyA site (12, 13). However, recent studies show that various stress and disease states, including osmotic stress (10), HSV-1 infection (9), and renal carcinoma (8), increase both the levels and length of transcripts mapping to regions downstream of the cleavage and polyadenylation sites. Our previous study (10) showed that these transcripts are continuous with the RNAs generated from the upstream protein-coding gene, suggesting that they result from alterations in cleavage and polyadenylation and/or termination events. Two studies found a correlation between readthrough transcription and lower frequency of polyA sites, either at the end of gene (9) or in the first 5 kb downstream of the gene end (10). Further, Vilborg et al. (10) found that induction of readthrough transcripts, referred to as downstream of gene-containing transcripts (DoGs), after osmotic stress is regulated by calcium signaling mediated by the IP3 receptor (IP3R) in human neuroblastoma SK-N-BE(2)C cells. However, the exact mechanism through which osmotic or viral stress affect readthrough, as well as the function of readthrough transcripts, remains unclear.

Preliminary data from Vilborg et al. (10) suggested a role for readthrough transcription in maintenance of euchromatin after stress. Even without conclusive evidence of function, the stress-mediated induction of thousands of readthrough transcripts, many up to 100 kb or more in length, is important to consider when performing

## Significance

**Cells and organisms live in constantly changing environments. Therefore, cells have evolved complex mechanisms to cope with physiological and environmental stresses. Many of these mechanisms involve transcriptional responses facilitating survival and adaptation. Recent evidence documents extensive transcriptional readthrough beyond annotated gene ends in response to stress, but the role and regulation of these downstream of gene-containing transcripts (DoGs) remain elusive. Here we report that induction of transcriptional readthrough is a hallmark of the mammalian stress response. We explore its causes and consequences in a genome-wide fashion, identifying thousands of readthrough transcripts that are induced in three different stress conditions. Our results suggest potential roles for this class of transcripts in the maintenance of open chromatin under stress.**

Author contributions: A.V. and R.S. designed research; A.V., N.S., Y.W., J.N., F.L.-A., and T.A.Y. performed research; N.S., Y.W., and R.S. analyzed data; Y.W. wrote the DoG discovery pipeline; N.S. performed all computational analyses in the manuscript; and A.V. and R.S. wrote the manuscript with input from N.S., Y.W., and J.A.S.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, <https://www.ncbi.nlm.nih.gov/geo> (accession no. GSE98906).

<sup>1</sup>To whom correspondence may be addressed. Email: vilborg.anna@gmail.com or reutshalgi@technion.ac.il.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1711120114/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1711120114/-DCSupplemental).

transcriptome analyses. Rutkowski et al. (9) demonstrated that many quiescent genes previously reported to be activated by HSV-1 infection in fact appeared to be activated as a result of readthrough into downstream genes by upstream DoGs; these “readthrough-induced genes,” however, did not generate functional mRNAs, as demonstrated by the absence of these mRNAs in accompanying ribosome profiling data (9). Likewise, noncoding RNAs mapping downstream of genes that are detected specifically after stress may in fact be readthrough transcripts, suggesting caution should be taken when interpreting changes in noncoding RNA expression after stress. Assessing the extent of stress-induced transcriptional readthrough is therefore critical.

In this study, we document widespread induction of readthrough transcription by heat shock and oxidative stress in mouse NIH 3T3 fibroblast cells. We compare the patterns and kinetics of readthrough induction between these two stress conditions and osmotic stress. Importantly, we provide evidence that stress-induced transcriptional readthrough is not a result of random failure of cleavage and polyadenylation or of subsequent transcription termination, but instead is differentially regulated by different stress conditions. We explore genomic features of readthrough transcripts and show they exhibit a unique chromatin signature, suggesting readthrough transcription is associated with the maintenance of an open chromatin state.

## Results

**Comparative Genome-Wide Stress-Dependent Regulation of Transcriptional Readthrough.** In previous work, several osmotic-stress-induced human DoGs were also induced by heat stress (10). This suggested readthrough might not be specific to osmotic stress but, rather, be a more general phenomenon occurring in other proteotoxic stress conditions. Indeed, when we reanalyzed our previously published RNA-Seq data for heat shock-treated mouse NIH 3T3 fibroblasts (7), as well as available data for oxidative-stress-treated human skin fibroblasts (14), we found evidence for widespread induction of transcriptional readthrough in both datasets (*SI Appendix, Fig. S1 A and B*). To obtain a comprehensive understanding of quantitative and qualitative differences in transcriptional readthrough induction in response to proteotoxic stress, we performed a genome-wide comparison of three different stresses in the same mammalian cell type.

To produce high-resolution transcription readthrough data, we extracted nuclear RNA from NIH 3T3 cells treated with heat shock (44 °C), osmotic stress (200 mM KCl), or oxidative stress (0.2 mM H<sub>2</sub>O<sub>2</sub>) for 2 h. We then performed strand-specific RNA-Seq on nuclear RNA after removal of ribosomal RNA. The resulting data were subjected to a DoG discovery pipeline (outlined in *SI Appendix, Fig. S1C*). In short, the pipeline demands 80% coverage in the first 4 kb downstream-of-gene ends to define a minimal DoG unit and then extends the DoG in running windows of 200 nucleotides with a requirement of 80% coverage for the entire DoG. Importantly, we defined the gene 3' ends very stringently by combining all annotated transcript definitions from RefSeq, UCSC, and Ensembl databases and defining the gene boundaries to be the longest possible; this approach should eliminate the discovery of DoGs that in fact represent alternative 3' gene ends (*Materials and Methods*). Our results were robust with respect to the choice of pipeline parameters (*Materials and Methods*).

We used our pipeline for DoG discovery in each condition independently and then compared the resulting DoGs in the different stress conditions. The length of DoGs indicates how far readthrough polymerases continue transcribing beyond the annotated gene ends, while DoG expression levels represent the degree of readthrough (i.e., how many of the polymerases transcribing a gene fail to terminate properly). After data filtering (*Materials and Methods*), the discovery pipeline identified a total of 4,852 DoGs expressed in at least one condition. NIH 3T3 DoGs are distributed along the entire mouse genome (*SI Appendix,*

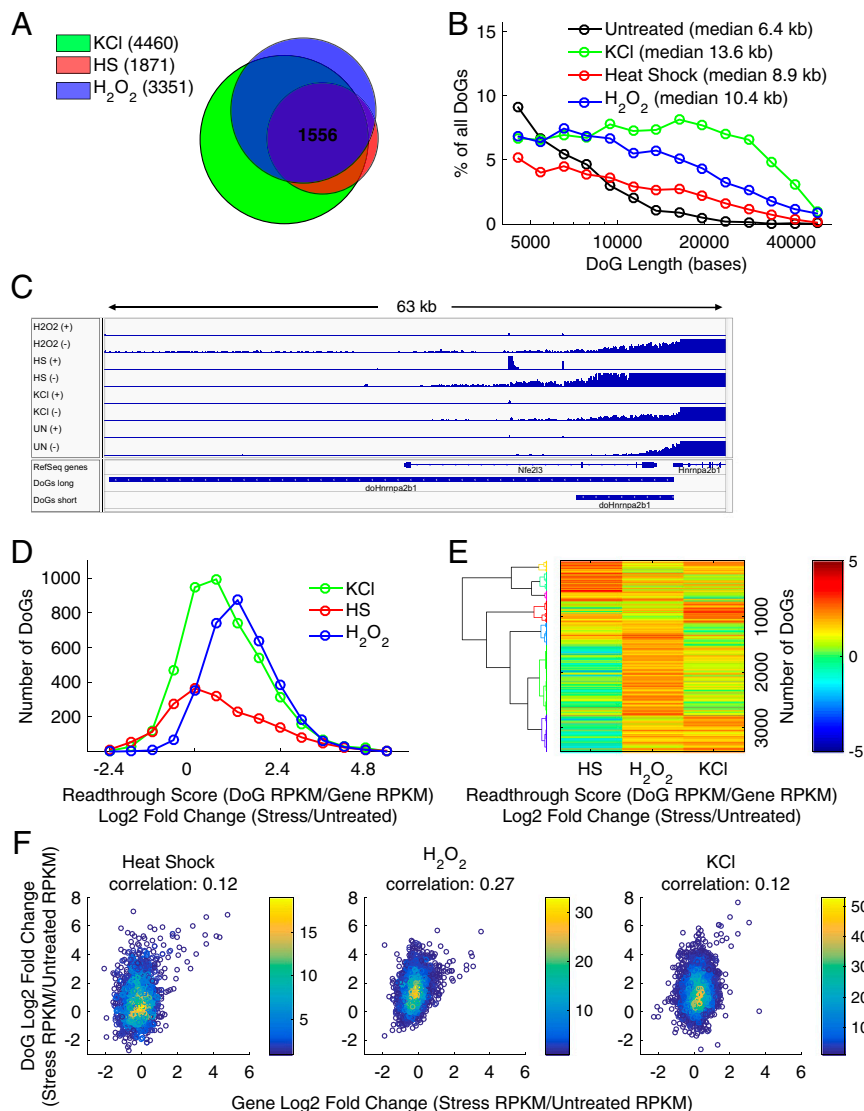
*Fig. S1D* and *Dataset S1*). Notably, we found significant overlap between DoGs identified in the different stress conditions, with 1,556 DoGs common to all three stresses ( $P$  value  $<10^{-300}$ , calculated using ref. 15), while other DoGs were induced by one or two stresses (Fig. 1A). The pipeline detected 1,860 transcripts extending more than 4 kb beyond gene 3' ends in untreated cells. As transcription termination has been shown to occur on average around 1.5–3.3 kb downstream of annotated 3' gene ends and can extend up to 10 kb (11, 16), these transcripts probably represent true transcription termination events that occur more than 4 kb beyond the annotated 3' end (16). Indeed, only 14 such cases were exclusively identified in untreated cells, while the rest were identified in at least one additional stress condition. Importantly however, DoGs were almost universally longer in stress conditions than in untreated cells (10) (Fig. 1B and *SI Appendix, Fig. S1E*). For example, the DoG *doHnmpa2b1*, for downstream of *Hnmpa2b1*, was found in all conditions (Fig. 1C); however, it terminated around 10 kb downstream of the gene end in unstressed cells, while in stressed cells it extended up to 47–58 kb. In contrast, *doTxn1* was specific to osmotic stress and extended up to 27 kb (*SI Appendix, Fig. S3A*; see *SI Appendix, Fig. S3 B and C* for additional examples).

We then calculated DoG expression levels (as reads per kilobase per million mapped reads, RPKMs), using either the shortest or longest DoG definitions (the most proximal and most distal endpoints found by the DoG discovery pipeline) from all conditions (*Materials and Methods*), and normalized them to the RPKM of their upstream genes [using Cufflinks (17); see *Materials and Methods*]. This normalization generated a readthrough score for each DoG, representing the ratio of readthrough transcription in each condition. Using this metric, we calculated the fold change in readthrough score for all identified DoGs in each stress compared with untreated cells. We found massive readthrough induction in all stress conditions, with DoGs induced on average over twofold in the different stress conditions (mean of readthrough score fold changes based on the longest DoG definition was 2, 2.15, and 3.07 with SDs of 2.65, 2.28, and 1.93 for heat shock, osmotic stress, and oxidative stress, respectively; Fig. 1D). The majority of DoGs were up-regulated more than 1.5-fold (56%, 65%, and 88% for heat shock, osmotic stress, and oxidative stress, respectively; Fig. 1D). Importantly, we observed a similar trend when calculating readthrough score fold changes using the shortest DoG definition (*SI Appendix, Fig. S1F*). Together, these observations indicate that both the number of polymerases that read through during stress and the distance they continue transcribing increase in stress conditions (Fig. 1D and *SI Appendix, Fig. S1F*). These results were recapitulated in a biological replicate experiment (*SI Appendix, Fig. S2* and *Dataset S1*).

We selected three DoGs, *doHspa8* (downstream of the constitutive HSP70-encoding gene), *doHnmpa2b1*, and *dolfum2* (Fig. 1C and *SI Appendix, Fig. S3 B–D*), and confirmed their stress-mediated induction by qRT-PCR in NIH 3T3 cells (*SI Appendix, Fig. S3 E–G*).

We compared the changes in DoG levels for all DoGs discovered in our data. To this end, we subjected the readthrough score fold changes of DoGs to hierarchical clustering analysis (*Materials and Methods*), selecting for analysis 3,431 DoGs with a readthrough score induction of at least twofold in at least one condition. Clustering analysis showed that DoGs are often markedly induced in a stress-specific manner (Fig. 1E). Nevertheless, DoGs that are highly induced by any single stress are often also mildly induced by oxidative and osmotic stresses, but not necessarily by heat shock. Interestingly, the most highly induced DoGs tended to be stress specific ( $P$  value  $1.5 \times 10^{-8}$  for heat shock and osmotic stress, and 0.001 for oxidative stress; *Materials and Methods*). These DoGs were enriched for pathways related to regulation of catabolic processes (*Materials and Methods*).

Finally, we asked whether readthrough induction merely reflects random failure of transcription termination, whereby polymerases fail to terminate at a higher rate under stress conditions.

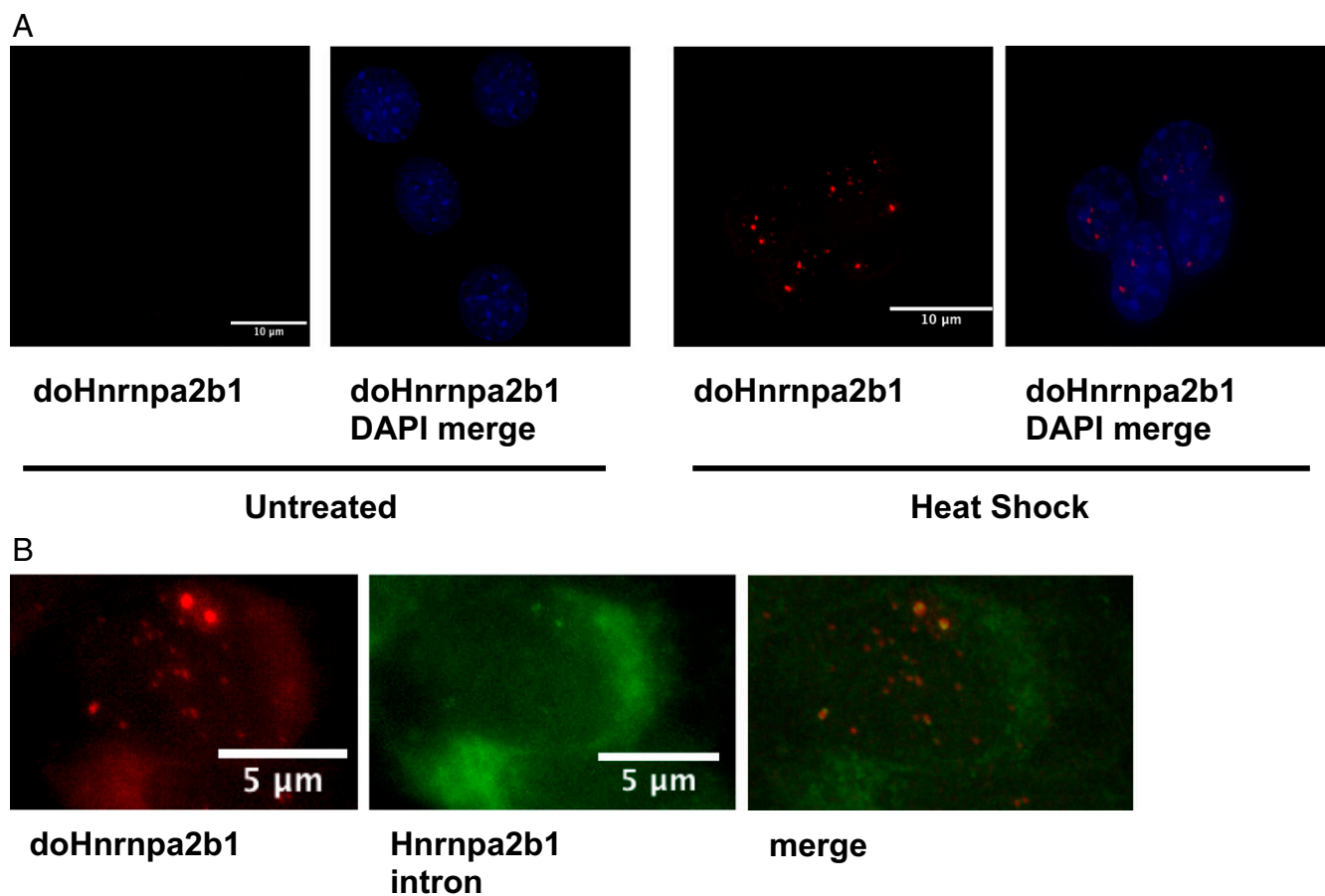


**Fig. 1.** Comparative genome-wide stress-dependent regulation of transcriptional readthrough. (A) Venn diagram of DoGs discovered in each of the three stress conditions shows high overlap. (*Inset*) Numbers of DoGs discovered in each condition; 1,860 DoGs were found in untreated cells. There are 1,556 DoGs that are common to all three stresses ( $P$  value  $< 10^{-300}$ , calculated using ref. 15). DoGs common to two of three stresses: 42 for heat shock and  $H_2O_2$ , 203 for heat shock and KCl, 1,487 for  $H_2O_2$  and KCl. There are 70, 266, and 1,214 DoGs unique to heat shock,  $H_2O_2$ , and KCl, respectively. (B) Length distribution of DoGs discovered in the three stresses and untreated cells. The percentage of DoGs of each length relative to the entire set of 4,852 DoGs is shown on the y axis. (C) Mapped read density for *Hnmpa2b1* DoG in all stress conditions. Panels shows mapped reads (scale, 0–150), separately for forward and reverse strands (“+” and “-,” respectively). (*Bottom*) RefSeq gene annotation tracks, and the longest and shortest DoG annotations. (D) Readthrough score log<sub>2</sub> fold change distribution for the three stresses. Readthrough score was defined as DoG RPKM/Gene RPKM, using the longest DoG annotations. (E) Hierarchical clustering analysis of the readthrough score fold changes across conditions (log<sub>2</sub> scale) for 3,431 DoGs with changes of at least twofold in at least one stress condition. (*Right*) A heat map of readthrough score fold changes (log<sub>2</sub> scale). (*Left*) Corresponding dendrogram resulting from the clustering procedure. (F) DoG RPKM log<sub>2</sub> fold change versus gene RPKM log<sub>2</sub> fold change in the three stresses show that readthrough induction is poorly explained by increase in gene expression. Spearman rho correlation coefficients are shown.

If this were the case, we would expect that the induction of readthrough should be highly correlated with the induction of the upstream gene, and that the more highly abundant the mRNA is, the higher the readthrough should be during stress. However, we found a low correlation between the fold changes of DoG RPKMs and the RPKM fold changes of their upstream genes (Fig. 1F). There was also a low correlation between DoG and gene expression levels (*SI Appendix, Fig. S4A*), as well as between the gene fold change and the length of DoGs (*SI Appendix, Fig. S4B*). The fold changes of 270 genes that were significantly up-regulated in stress conditions [significance assessed using DESeq2 (18)] showed moderate correlation with the fold changes of their respective DoGs (0.3–0.36; *SI Appendix, Fig. S4 C and D*). However, these

represent a very small fraction (5.6%) of the 4,852 DoG-associated transcripts. In addition, fold changes of DoGs derived from unchanged genes were frequently as high as those of induced genes. These observations indicate that, overall, gene induction cannot explain the majority of readthrough induction that occurs in stress conditions and argue that readthrough is differentially regulated under stress conditions.

**DoGs Localize to Nuclear Punctae.** To investigate the subcellular localization of DoG transcripts in NIH 3T3 cells, we used single-molecule FISH for *doHnmpa2b1* and *doHspa8*. This method confirmed their heat shock- and KCl-mediated induction (Fig. 2A and *SI Appendix, Fig. S5 A–C*). Once induced, DoGs localized to multiple nuclear punctae, in agreement with the pattern observed



**Fig. 2.** Single-molecule RNA-FISH confirms induction of nuclear DoGs. (A) Single-molecule Stellaris RNA-FISH confirms the heat shock-mediated induction of *doHnrnpa2b1* and *doHspa8* (SI Appendix, Fig. S5). In both cases, DoGs appear in nuclear puncta. (B) Costaining with the *doHnrnpa2b1* probe and a probe targeting the first intron of *Hnrnpa2b1* in heat shock-treated NIH 3T3 cells confirms that some of the nuclear dots containing *doHnrnpa2b1* staining represent *Hnrnpa2b1* transcription sites.

in our previous study of DoG induction in human neuroblastoma cells (10). However, our previous data showed a maximum of four dots per nucleus. These dots represented DoG transcription sites for one or two alleles, confirmed by costaining with an intronic probe, which labels the nascent transcript at the site of transcription, as splicing is predominantly cotranscriptional (10). In contrast, here we found more than four puncta per nucleus after both heat shock (Fig. 2) and KCl (SI Appendix, Fig. S5) treatments, suggesting DoGs can also localize to other sites within the nucleus. This pattern was particularly pronounced after heat shock treatment.

To investigate whether some of the observed dots coincide with transcription sites, we performed costaining of the *doHnrnpa2b1* probe with a probe targeting intron 1 of *Hnrnpa2b1* in heat shock-treated NIH 3T3 cells (Fig. 2B and SI Appendix, Fig. S6 A–C). We detected colocalization of the two probes, even though the *doHnrnpa2b1* probe also localized to other nuclear puncta. This pattern suggests that some but not all of the *doHnrnpa2b1* puncta induced by heat shock and by osmotic stress represent transcription sites; other DoGs may leave the site of transcription and spread within the nucleus.

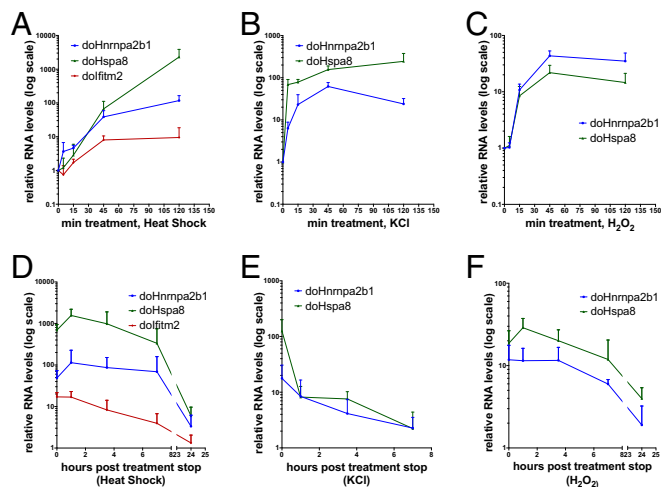
To gain further understanding of the subcellular localization of DoGs in a transcriptome-wide fashion, we took advantage of a ribosome footprint profiling dataset previously performed on NIH 3T3 cells in control and heat shock conditions (19) and calculated the footprint density over the first 4 kb of heat shock DoG regions. More than 79% of heat shock DoGs showed no evidence of translation (zero ribosome footprint reads), while the remaining DoGs

contained spurious reads in both heat shock and untreated cells, with very low RPKM (Materials and Methods). We conclude that DoGs are not translated, in agreement with previous evidence on their nuclear localization, as confirmed by cell fractionation (10). Thus, current evidence suggests that DoGs remain nuclear and do not reach the cytoplasm.

#### The Kinetics of DoG Induction Differ with Different Stress Conditions.

To explore similarities and differences between the induction of DoGs by the three stress conditions, we performed time course experiments in NIH 3T3 cells. After heat shock, DoG levels start to increase at 15 min after stress and continue to increase through the duration of the experiment (2 h) (Fig. 3A), with a similar extent of induction at 2 and 4 h (SI Appendix, Fig. S3E). In contrast, DoG induction by osmotic stress is evident already after 5 min, and levels do not rise significantly after 45 min (Fig. 3B). After oxidative stress, DoG levels start to increase at 15 min after treatment and reach a maximum at 45 min (Fig. 3C).

Next, we examined the kinetics of DoG level changes during recovery from each of the stress conditions. We treated cells for 2 h with heat shock, osmotic stress, or oxidative stress, and then removed the stress and harvested samples at subsequent times. We found striking differences between osmotic stress on the one hand and heat shock and oxidative stress on the other (Fig. 3D–F). Upon reversal of osmotic stress, DoG levels were markedly reduced already at 1 h and had returned to baseline at 3.5 h after removal of KCl (Fig. 3E). In contrast, during recovery from heat



**Fig. 3.** The kinetics of DoG induction and decay vary for different stress conditions. Time courses of DoG induction by heat shock (A), KCl (B), and H<sub>2</sub>O<sub>2</sub> (C), using qRT-PCR, demonstrating DoG induction already after 5 min by KCl treatment, while DoG induction after heat shock and H<sub>2</sub>O<sub>2</sub> is noticeable after 15 min. DoG induction by KCl and H<sub>2</sub>O<sub>2</sub> reaches a maximum after 45 min but keeps increasing at 2 h after heat shock. The duration of DoG expression after stress removal also differs for heat shock (D), KCl (E), and H<sub>2</sub>O<sub>2</sub> (F) stress. While DoG levels quickly return to baseline after removal of KCl stress (200 mM 2 h), DoG levels remain high as long as 7 h upon return to 37 °C after heat shock (44 °C for 2 h) and after recovery from H<sub>2</sub>O<sub>2</sub> stress. Mean and SD of 3 (A–D) or 4 (E–F) biological replicate experiments are presented.

shock (Fig. 3D) or oxidative stress (Fig. 3F), DoG levels remained high as long as 7 h after reversal. This difference in the kinetics of induction and duration of DoG levels during recovery from different stress conditions suggests different modes of regulation.

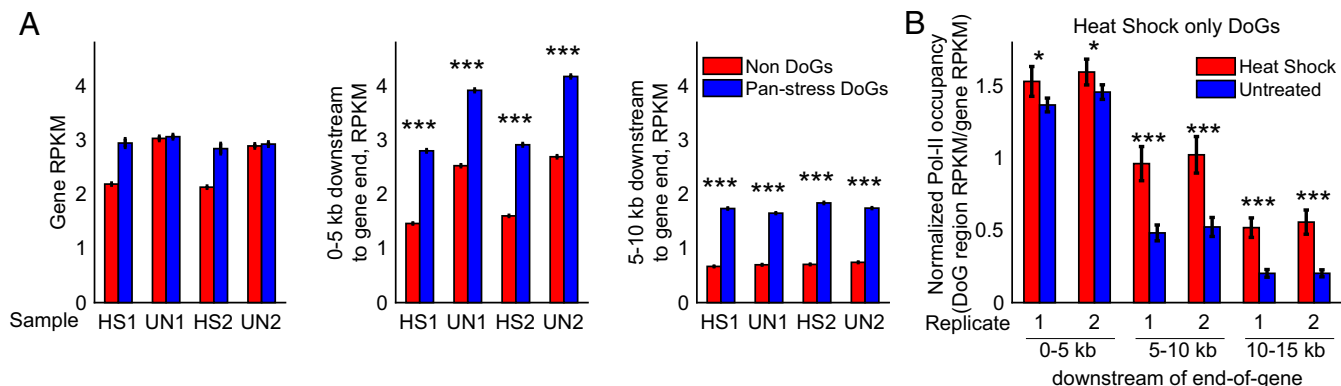
**Detecting Heat Shock-Induced Readthrough at the Level of RNA Pol-II Occupancy.** We next explored the landscape of RNA polymerase (Pol-II) occupancy in readthrough regions. We used a recently published high-resolution dataset of genome-wide Pol-II occupancy (PRO-seq) performed in MEFs (mouse embryonic fibroblast cells) before and after heat shock (5). We first quantified Pol-II occupancy in different regions of genes associated with pan-stress DoGs, the 1,556 DoGs common to all three stress conditions (*Materials*

and *Methods*). For controls, we defined a set of non-DoG genes, i.e., transcripts with very low readthrough levels in all conditions (*Materials and Methods* and *SI Appendix, Fig. S7A*). Importantly, we found a marked increase in Pol-II occupancy in pan-stress DoGs compared with non-DoG downstream regions (*SI Appendix, Fig. S7B*). Moreover, the difference was highly significant when we matched the expression of pan-stress and non-DoG genes using the PRO-seq gene body data (Fig. 4A and *SI Appendix, Fig. S7C*), such that both sets show the same distribution of Pol-II occupancy in the body of the genes in untreated cells (Fig. 4A), as well as in heat shock (*SI Appendix, Fig. S7C*), indicative of equal transcription level distributions. Thus, we observed substantially more Pol-II occupancy downstream of pan-stress DoG genes than downstream of non-DoG genes, even though their Pol-II occupancy in the gene body is the same.

We next normalized the Pol-II occupancy in DoG regions to the Pol-II occupancy of the corresponding gene. Interestingly, when we examined DoGs present in heat shock, but not in untreated cells, we observed significantly higher normalized Pol-II occupancy immediately downstream of gene ends, and this difference increased the further downstream we looked (Fig. 4B). When we considered all the heat shock DoGs, including the ones also present in untreated cells, we observed a similar increase in normalized Pol-II occupancy in heat shock, but occurring farther downstream (5–10 kb and 10–15 kb downstream of gene ends; *SI Appendix, Fig. S7D*). This observation mirrors our analyses showing some natural downstream termination sites for certain genes; however, these genes generate readthrough transcripts that extend significantly beyond their natural termination regions after stress (Fig. 1B).

Taken together, our analyses indicate similar heat shock-induced readthrough patterns in the two fibroblast cell types, NIH 3T3 and MEFs, even considering the two different, yet complementary, experimental methodologies: nuclear RNA-Seq and PRO-seq. These results further substantiate the notion that stress-induced transcriptional readthrough is a conserved, widespread phenomenon, which can be directly observed at the polymerase level.

**Heat Shock-Induced Readthrough Is Affected by HSF1.** Since DoGs are differentially induced in different stress conditions, we hypothesized that transcription regulators might play a role in stress-induced readthrough. We thus examined the potential involvement of the bona fide transcription regulator of the heat shock response, the transcription factor HSF1. We analyzed a published dataset of



**Fig. 4.** Pol-II occupancy landscape in DoG regions. (A) Mean and SE of Pol-II occupancy [PRO-seq (5) RPKM] in genes and DoG regions (first 5 kb and 5–10 kb) in heat shock (HS) and untreated (UN) samples are shown for pan-stress DoGs (blue) versus non-DoGs (red). Significant differences between pan-stress DoGs and non-DoG regions are shown for pan-stress and non-DoG sets expression-matched using PRO-seq untreated cell data RPKMs of gene regions. Significance was estimated as the FDR-corrected 95th percentile of 1,000 ranksum test *P* values for 1,000 expression-matched subsamples. (B) Higher normalized Pol-II occupancy in heat shock compared with untreated cells for heat shock DoGs. Mean and SE of Pol-II occupancy in DoG regions (first 5 kb, 5–10 kb, and 10–15 kb downstream to gene ends) normalized by Pol-II occupancy in the corresponding upstream gene are shown for DoGs present in heat shock but not in untreated cells. Significance was assessed by ranksum test and controlled for false-discovery rate (A and B, \**P* < 0.05; \*\*\**P* < 0.001).

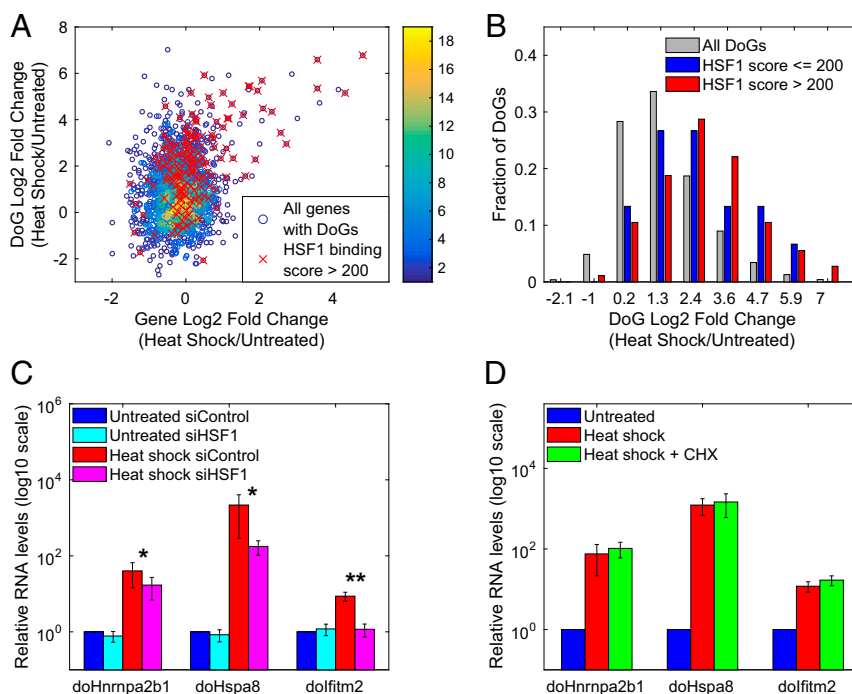
a genome-wide HSF1 ChIP-seq performed in MEFs, before and after heat shock (5). HSF1 showed both basal and heat shock-induced binding to chromatin mainly in promoter regions and was rarely present in regions specifying the 3' ends of transcripts or downstream (*SI Appendix, Fig. S8 A and B*). Surprisingly, we observed a correlation between heat shock-enhanced HSF1 binding in a gene's promoter and the extent of readthrough; HSF1 promoter-bound genes exhibited higher correlation between gene and DoG induction levels ( $\rho = 0.41$ ; red crosses, Fig. 5A), and DoGs were more induced the higher their gene's HSF1 binding scores were (Fig. 5B and *SI Appendix, Fig. S8C*). Thus, genes with increased HSF1 binding demonstrated greater heat shock-induced readthrough (Fig. 5A and B). Analysis of an additional heat shock HSF1 ChIP-seq dataset (20) showed identical trends (*SI Appendix, Fig. S8C*).

To validate these findings, we examined the effects of HSF1 depletion in NIH 3T3 cells on the three DoGs *doHspa8*, *doHnmpa2b1*, and *dolfitm2*. We found DoG induction to be significantly reduced in cells transfected with siRNA targeting HSF1 48 h before heat shock treatment compared with control siRNA-transfected cells (Fig. 5C). While heat shock-increased levels of *doHspa8* and *doHnmpa2b1* were partially dependent on HSF1, that of *dolfitm2* was completely abrogated after HSF1 knockdown (Fig. 5C). HSF1 knockdown did not affect DoG induction by osmotic stress or oxidative stress (*SI Appendix, Fig. S8 D and E*). HSF1 mRNA knockdown was efficient (*SI Appendix, Fig. S8 F–H*) and, most important, did not alter the levels of the corresponding transcripts from the upstream DoG-associated genes (*SI Appendix, Fig. S8 I–K*), indicating that the reduction in heat shock-mediated DoG in-

duction after HSF1 knockdown was independent of the level of the DoG-associated mRNAs.

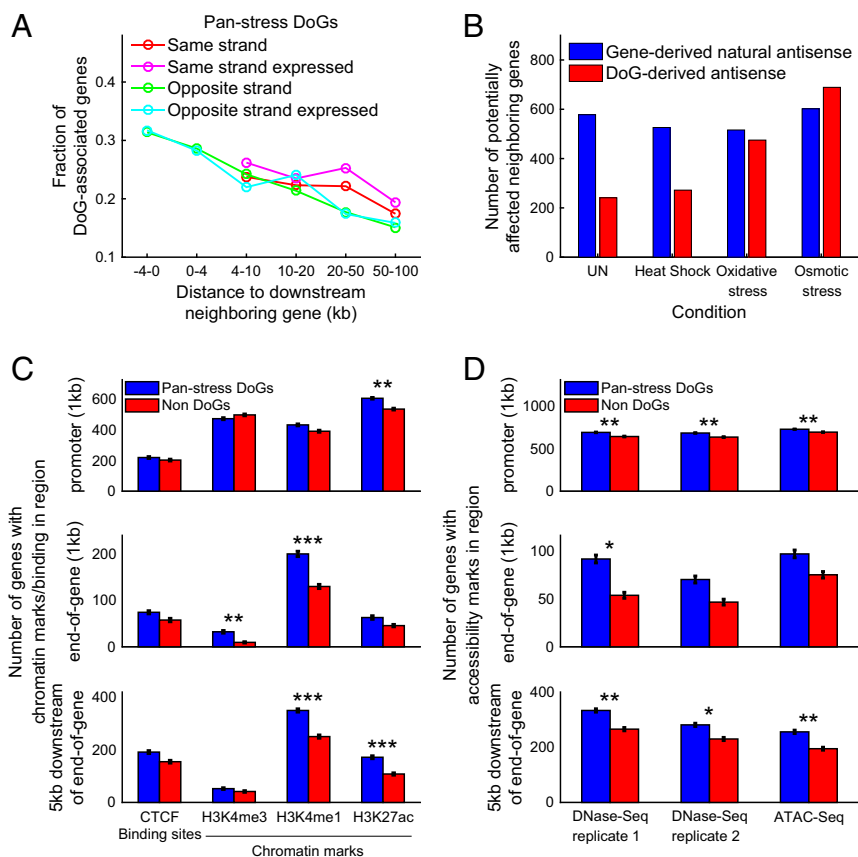
Since HSF1 is a transcription factor that regulates a large number of genes upon heat shock, one possibility was that HSF1 activates the transcription of a gene encoding a protein involved in DoG generation. We therefore investigated whether DoG induction after heat shock is prevented by cycloheximide, which blocks protein translation. Cycloheximide treatment for 15 min before the onset of stress did not affect DoG induction after heat shock (Fig. 5D), whereas the RNA levels of the NMD (nonsense-mediated mRNA decay) target *Uhg* (U22 host gene) transcript increased, as expected (21) (*SI Appendix, Fig. S9A*). These observations suggest that, for the subset of genes with enhanced HSF1 binding, DoG induction upon heat shock involves HSF1 and does not involve protein synthesis.

**DoG Induction Is Affected by Intracellular Calcium.** Calcium signaling through the IP3R was previously reported to play a major role in transcriptional readthrough after osmotic stress in human SK-N-BE(2)C cells (10). We therefore pretreated NIH 3T3 cells with the membrane-permeable calcium chelator BAPTA for 30 min before the onset of stress and quantified DoG levels by qPCR. We found that calcium chelation partially reduced the levels of DoG induction in most, but not all, cases; the greatest effect was observed after heat shock, where the induction of all tested DoGs (*doHspa8*, *doHnmpa2b1*, and *dolfitm2*) was reduced (*SI Appendix, Fig. S9 B–D*). However, pretreatment with the IP3R inhibitor 2-APB did not prevent DoG induction by the stresses tested (*SI Appendix, Fig. S9E*). We conclude that calcium signaling contributes, at least in part, to DoG induction in NIH 3T3 cells in



**Fig. 5.** Heat shock DoG induction is affected by HSF1. (A) Genes with induced HSF1-promoter binding in heat shock [Mahat et al. (5), with binding-score differences between heat shock and control, greater than 200] show a higher correlation between DoG RPKM log<sub>2</sub> fold change and Gene FPKM log<sub>2</sub> fold change. Spearman  $\rho$  of all Gene-DoG pairs = 0.12, HSF1 promoter-bound Gene-DoG pairs (marked in red crosses) = 0.41. (B) DoGs with heat shock-induced HSF1 promoter binding show increased DoG expression levels, as seen by a shifted distribution of DoG RPKM log<sub>2</sub> fold changes in heat shock compared to control. Both blue (binding scores between 1 and 200; mean = 1.96), and red (binding score above 200; mean = 2.2) show a shift toward higher induction levels compared with all DoGs (gray). (C) Knockdown of HSF1 48 h before heat shock shows that DoG induction is dependent on HSF1. Data represent the average and SD (log<sub>10</sub> scale) of 7 biological replicate experiments of qRT-PCR. Significant differences between siHSF1 and siControl in heat shock are noted. \* $P < 0.05$ ; \*\* $P < 0.01$ . (D) DoG induction does not require de novo protein synthesis after heat shock, as pretreatment with cycloheximide (CHX) did not prevent DoG induction by heat shock. Data represent average and SD (log<sub>10</sub> scale) of 5 biological replicate experiments.





**Fig. 7.** Stress-induced readthrough transcription is associated with open chromatin features. (A) DoG frequency is negatively correlated with the distance to the nearest downstream neighboring gene. The fraction of DoG-associated genes out of all expressed genes is plotted on the y axis as a function of the distance to the downstream nearest neighboring gene. Gene pairs were grouped according to strand orientation of the downstream neighbor (same strand or opposite strand) and whether the downstream neighboring gene is expressed (RPKM  $\geq 4$ ). (B) The number of genes potentially affected by antisense transcription of overlapping expressed opposite-strand neighboring genes (blue) or DoGs (red). (C–D) Number of genes with chromatin marks in promoter regions (1 kb upstream of transcription start site), gene ends (1 kb upstream of gene end), and regions downstream of gene ends (5 kb), among pan-stress DoG and non-DoG groups. As groups differ in size and expression, 1,000 expression-matched subsamples were generated. The mean and SE for the number of gene regions with at least one peak are shown for (C) ChIP-seq of three histone marks (H3K4me3, H3K4me1, H3K27ac) and CTCF insulator in MEF cells (24, 25), and (D) two chromatin accessibility assays, ATAC-Seq in MEF cells (28) and DNase-Seq in NIH 3T3 cells (24, 27). In each subsample, Fisher exact test was applied to assess significance (*Materials and Methods*). \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ .

### Stress-Induced Readthrough Transcription Is Associated with Features of Open Chromatin.

As DoGs tend to be expressed in the proximity of neighboring genes regardless of their neighbors' expression, we hypothesized that perhaps a certain chromatin environment favors stress-induced readthrough transcription. To examine this possibility, we used histone modification and CTCF ChIP-seq data performed in MEFs [from mouse ENCODE (24, 25)] and compared the chromatin modification pattern in promoters, gene ends, and downstream regions for pan-stress DoG genes versus non-DoG genes (*Materials and Methods*). Since expression levels and chromatin status are related, here too expression-matched pan-stress DoG and non-DoG gene sets of the same size were used, including shuffled controls to assess significance (*Materials and Methods*). We found that promoters, as well as downstream regions, of pan-stress DoG genes were no different from those of non-DoG genes with respect to the histone mark H3K4me3 (Fig. 7C), a typical promoter mark (26). The same was true for regions bound by the chromatin insulator CTCF (Fig. 7C and *SI Appendix, Fig. S11C*). However, DoG-associated genes showed a significant enrichment in histone marks typical of accessible regulatory chromatin, H3K4me1 and H3K27ac, especially in their downstream regions (Fig. 7C and *SI Appendix, Fig. S11C*). Analysis of additional histone modification data from C2C12 mouse myoblast cells [generated by the Wold lab-

oratory, Caltech, Pasadena, CA (24)] showed a marked enrichment in elongation histone marks (H3K36me3 and H3K79me2) both near the gene ends and downstream of pan-stress DoGs, but no significant difference was found for the repressive mark H3K27me3 (*SI Appendix, Fig. S11D*). These analyses suggest that stress-induced transcriptional readthrough is enhanced in regions of open chromatin. To look more closely at chromatin accessibility, we analyzed publicly available DNase hypersensitivity data (DNase-seq) from NIH 3T3 cells (24, 27) and ATAC-seq data from MEFs (28). Importantly, we found significant enrichment in both DNase hypersensitive sites and ATAC-seq peaks, both in the promoters and downstream of pan-stress DoGs compared with non-DoG genes (Fig. 7D and *SI Appendix, Fig. S11E*). Together, these results indicate that even before stress, stress-induced readthrough gene loci have a distinct chromatin signature indicative of an open chromatin state.

### Discussion

In this study, we established transcriptional readthrough as a hallmark of the transcriptional response to proteotoxic stress. We observed that transcriptional readthrough is induced by heat shock and oxidative stress, as well as by osmotic stress, as reported previously (10). By comparing the genome-wide readthrough patterns after heat shock, oxidative stress, and osmotic stress



using deep nuclear RNA-Seq in mouse NIH 3T3 cells, we found 1,556 DoGs (pan-stress DoGs) that were common to all three stress conditions. Nevertheless, numerous DoGs show stress specificity. Importantly, induction of DoGs did not correlate with enhanced transcription of the DoG-generating gene, arguing against a random failure mechanism of transcription termination.

Interestingly, some readthrough transcripts were down-regulated in heat shock compared with untreated cells. In our RNA-Seq dataset, we found no difference in the reduction of the mRNA levels of the 91 genes with reduced heat shock readthrough (two-fold or more, average mRNA RPKM fold change upon heat shock of 0.84 compared with 0.87, respectively;  $P$  value = 1). However, PRO-Seq data (5) showed slightly reduced Pol-II occupancy in these genes (average PRO-seq RPKM fold change upon heat shock of 0.61 compared with 0.77, respectively;  $P$  value =  $1.4 \times 10^{-5}$ ). Taken together with the moderate correlation observed between the induction of DoGs and their associated mRNAs for genes significantly induced by heat shock (*SI Appendix, Fig. S4 C and D*), it is possible that marked up- or down-regulation of a gene would subsequently affect its DoG levels in a similar direction. However, our evidence suggests that, overall, DoGs are regulated independently of their upstream genes.

By comparing the DoGs discovered here to the set of human osmotic-stress-induced DoGs identified previously in neuroblastoma SK-N-BE(2)C cells (10), we found that 62–64% of human DoGs are also found in mouse, even though different cell types were used for discovery ( $P$  value <  $2.2 \times 10^{-125}$ ; *SI Appendix, Table S2*). Similar to human osmotic-stress-induced DoGs, mouse pan-stress DoGs were not significantly enriched in any particular gene ontology pathway.

Using FISH on human cells subjected to osmotic stress, we previously found the DoG *doSERBP1* exclusively at its transcription sites (10). Interestingly, the use of a more powerful microscope in the present study allowed us to detect the localization of both *doHspa8* and *doHnnpa2b1* in additional punctae outside of transcription sites, suggesting they spread to distinct positions within the nucleus. Nevertheless, in both mouse and human, our evidence suggests that DoGs remain nuclear and do not reach the cytoplasm. These comparisons further support stress-induced transcriptional readthrough as an intrinsic part of a cellular stress response that is conserved across different stress conditions, as well as between mammalian species.

In the current study, we characterized various properties of stress-induced readthrough induction, which suggest potential mechanisms underlying the regulation of readthrough under stress. Interestingly, we found that HSF1 affects DoG induction, suggesting that HSF1 may be part of an effector pathway for heat shock-activated DoG transcription. Protein synthesis was not required for DoG induction, and increased HSF1 binding to gene promoters correlated with downstream DoG induction in heat shock. HSF1 promoter-bound genes showed higher correlation between their gene and DoG fold induction in heat shock ( $\rho$  of 0.41, Fig. 5A) and demonstrated increased Pol-II occupancy in heat shock, (average PRO-seq gene RPKM fold change of 1.4 versus 0.75 for all other DoG-related genes in heat shock compared to control). Nonetheless, a large fraction of the HSF1 promoter-bound genes showed a DoG induction of twofold or more, with only a marginal influence (less than 25% increase) on their corresponding mRNA level (half of HSF1 promoter-bound genes) and gene Pol-II occupancy (a third of HSF1 promoter-bound genes). Moreover, HSF1 knockdown reduced heat shock-mediated DoG induction, either partially or completely (Fig. 5C), for the candidate DoGs tested, without affecting the overall levels of the upstream gene transcript (*SI Appendix, Fig. S8J*). Together, these data indicate that HSF1 may play a direct role in DoG induction, rather than activating a target gene whose protein product regulates transcription termination or by affecting transcription levels of DoG-hosting genes. One hypothesis is that

HSF1 binding to the gene promoter may recruit termination regulators that travel with the transcribing Pol-II to the end of the gene where they affect termination, analogous to some elongation regulators (29). However, it is possible that part of the effect of HSF1 on a subset of DoGs is a result of induction of their upstream genes in heat shock. Moreover, while our results on HSF1 suggest the involvement of transcription factors in the induction of a subset of DoGs under certain conditions, such a mechanism cannot fully explain all DoG induction. Further detailed mechanistic studies are required to unravel the involvement of HSF1 and other transcription factors in particular, and of molecular changes in general, that allow pervasive transcriptional readthrough in response to stress.

Our current study confirms previous findings that readthrough transcription is correlated with weak polyA sites in the region downstream of gene ends. Further, we found a significant difference in sequence composition downstream of pan-stress DoG genes compared with non-DoG genes, including depletion of AC stretches, GU stretches, and other sequence motifs in DoG regions. Interestingly, GU elements have been implicated in efficient cleavage and polyadenylation in mammals, and are bound by several cleavage and polyadenylation factors (reviewed in ref. 30). It is tempting to speculate that cleavage and polyadenylation factors binding to low complexity tracts downstream of genes promote efficient termination under normal conditions, and that this requirement becomes more stringent after stress. Consistently, our previous study found a fourfold induction in the readthrough of the human *Cxhc4* after knockdown of the cleavage and polyadenylation factor CPSF73 (10). In addition, some of the significantly depleted motifs we identified are similar to those of known RNA binding proteins (*Materials and Methods* and *Dataset S2*), which may recruit the polyadenylation machinery to nascent RNAs for efficient cleavage and polyadenylation under stress. Indeed, the polyadenylation machinery was shown to be inhibited in heat shock (31), and if it becomes limiting in stress conditions, perhaps recruitment by specific RNA-binding proteins could promote efficient termination specifically for non-DoG-producing genes.

Another fascinating open question concerns the consequences of stress-induced transcriptional readthrough. We have shown that DoG production is not merely a random failure phenomenon, but rather involves specific genes under different conditions, supporting the hypothesis that stress-induced readthrough might function in cellular programs coping with stress. Our findings suggest that a combination of transcription regulators, sequence signals, and chromatin environment may govern this specificity. Most significantly, our analyses of Pol-II occupancy data (Fig. 4) mirror the readthrough detected at the RNA level (from our nuclear RNA-Seq) in heat shock conditions, further establishing the generality of the readthrough phenomenon. In addition, our observation of Pol-II enrichment in farther downstream regions of heat shock DoGs after heat shock compared with untreated cells further generalize our previous observation (10) that readthrough transcript accumulation does not reflect failure to degrade, but rather to increase the synthesis of these RNA molecules.

One possibility is that readthrough transcription may serve to sequester Pol-II molecules upon stress, thereby effectively reducing overall transcription, as part of an adaptive response. However, the overall fraction of Pol-II reads in gene bodies (calculated from PRO-seq data) showed only a slight decrease in heat shock (from 45.2% in untreated to 42.6% in heat shock). Moreover, readthrough is induced in different gene loci under different conditions. Thus, the effect of sequestration is probably minor, and the readthrough selectivity for different genes in different conditions warrants additional explanation.

Since DoGs remain nuclear, the levels of the mature cytoplasmic mRNAs of their associated genes are somewhat lower than measured by RNA-Seq. An analogous nuclear retention mechanism has been shown for splicing-inhibited transcripts in heat shock (7).

In addition, the lengthening of readthrough transcripts in stress leads to a marked increase in antisense transcription potential (Fig. 7B). Thus, readthrough may also fine-tune the expression of many genes upon stress, contributing to changes in the cytoplasmic transcriptome.

Interestingly, we found that DoG transcription is negatively correlated with distance to neighboring genes. While DoG-associated genes are on average 63 and 17 kb away from their closest 3' neighbor on the same and opposite strand, respectively, non-DoG genes tend to be much further away from their 3' neighbors (77 and 54 kb on average, respectively). Furthermore, our analyses show that readthrough genes bear a unique chromatin signature, indicative of open structure in the absence of stress. Induced readthrough occurs in plants with defects in the 3'-end processing machinery, and a link has been shown to chromatin modifications (32). Thus, there are two possible, not mutually exclusive, explanations for the observed correlation between readthrough transcription and open chromatin. First, the presence of open chromatin under normal growth conditions may facilitate stress-induced readthrough transcription. Second, readthrough transcription itself may help to keep chromatin regions open in stressful environments. Supporting the latter possibility, our previous study (10) found evidence of a role for DoGs in maintaining the integrity of the nuclear scaffold after stress. Thus, the favored expression of DoGs in proximity to neighboring genes, as well as a specific chromatin landscape,

supports the idea that transcriptional readthrough serves to keep specific genome regions open during stress, thereby maintaining nuclear organization in an otherwise compromised state.

## Materials and Methods

For stress treatments, cells were cultured at 44 °C (heat shock), or in the presence of 200 mM KCl (osmotic stress), or of 0.2 mM H<sub>2</sub>O<sub>2</sub> (oxidative stress) for 2 h if not otherwise stated. For siRNA accession numbers and primers, see *SI Appendix, Table S3*. Data were deposited in GEO, accession number GSE98906. Single-molecule RNA FISH was done using Stellaris probes. See *SI Appendix, SI Materials and Methods* for full details of all experimental procedures. A detailed description of the DoG discovery pipeline, and all subsequent analyses including hierarchical clustering analysis, highly induced DoGs, pan-stress DoG and non-DoG groups, Pol-II occupancy analysis, HSF1 ChIP-seq data analysis, sequence motifs (6-mers) analysis, functional enrichment analysis (33), and chromatin environment analysis can be found in the *SI Appendix*.

**ACKNOWLEDGMENTS.** We thank Dig B. Mahat from the J. T. Lis lab for providing normalized bed files of mapped PRO-seq data. We are grateful to Minju Ha, Deborah Lincow, and Nicole Rosa-Mercado for critical review of the manuscript. We thank Ortal Orbach for help with figure visual design and Angela Miccinello for expert editorial assistance. This project has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme Grant 677776 (to R.S., N.S., and Y.W.) and NIH Grant GM026154. J.A.S. is an investigator of the Howard Hughes Medical Institute.

- de Nadal E, Ammerer G, Posas F (2011) Controlling gene expression in response to stress. *Nat Rev Genet* 12:833–845.
- Vilborg A, Steitz JA (2016) Readthrough transcription: How are DoGs made and what do they do? *RNA Biol* 14:632–636.
- Adelman K, Lis JT (2012) Promoter-proximal pausing of RNA polymerase II: Emerging roles in metazoans. *Nat Rev Genet* 13:720–731.
- Rougvie AE, Lis JT (1988) The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of *D. melanogaster* is transcriptionally engaged. *Cell* 54:795–804.
- Mahat DB, Salamanca HH, Duarte FM, Danko CG, Lis JT (2016) Mammalian heat shock response and mechanisms underlying its genome-wide transcriptional regulation. *Mol Cell* 62:63–78.
- Zovolis A, Cifuentes-Rojas C, Chu HP, Hernandez AJ, Lee JT (2016) Destabilization of B2 RNA by EZH2 activates the stress response. *Cell* 167:1788–1802.e13.
- Shalgi R, Hurt JA, Lindquist S, Burge CB (2014) Widespread inhibition of post-transcriptional splicing shapes the cellular transcriptome following heat shock. *Cell Reports* 7:1362–1370.
- Grosso AR, et al. (2015) Pervasive transcription read-through promotes aberrant expression of oncogenes and RNA chimeras in renal carcinoma. *eLife* 4:e09214.
- Rutkowski AJ, et al. (2015) Widespread disruption of host transcription termination in HSV-1 infection. *Nat Commun* 6:7126.
- Vilborg A, Passarelli MC, Yario TA, Tycowski KT, Steitz JA (2015) Widespread inducible transcription downstream of human genes. *Mol Cell* 59:449–461.
- Core LJ, Waterfall JJ, Lis JT (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 322:1845–1848.
- Porrúa O, Libri D (2015) Transcription termination and the control of the transcriptome: Why, where and how to stop. *Nat Rev Mol Cell Biol* 16:190–202.
- Richard P, Manley JL (2009) Transcription termination by nuclear RNA polymerases. *Genes Dev* 23:1247–1269.
- Giannakakis A, et al. (2015) Contrasting expression patterns of coding and noncoding parts of the human genome upon oxidative stress. *Sci Rep* 5:9737.
- Wang M, Zhao Y, Zhang B (2015) Efficient test and visualization of multi-set intersections. *Sci Rep* 5:16923.
- Schwalb B, et al. (2016) TT-seq maps the human transient transcriptome. *Science* 352:1225–1228.
- Trapnell C, et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28:511–515.
- Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550.
- Shalgi R, et al. (2013) Widespread regulation of translation by elongation pausing in heat shock. *Mol Cell* 49:439–452.
- Takii R, et al. (2015) ATF1 modulates the heat shock response by regulating the stress-inducible heat shock factor 1 transcription complex. *Mol Cell Biol* 35:11–25.
- Tycowski KT, Shu MD, Steitz JA (1996) A mammalian gene with introns instead of exons generating stable RNA products. *Nature* 379:464–466.
- Khorkova O, Myers AJ, Hsiao J, Wahlestedt C (2014) Natural antisense transcripts. *Hum Mol Genet* 23:R54–R63.
- Katayama S, et al.; RIKEN Genome Exploration Research Group; Genome Science Group (Genome Network Project Core Group); FANTOM Consortium (2005) Antisense transcription in the mammalian transcriptome. *Science* 309:1564–1566.
- Yue F, et al.; Mouse ENCODE Consortium (2014) A comparative encyclopedia of DNA elements in the mouse genome. *Nature* 515:355–364.
- Pope BD, et al. (2014) Topologically associating domains are stable units of replication-timing regulation. *Nature* 515:402–405.
- Zhou VW, Goren A, Bernstein BE (2011) Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet* 12:7–18.
- Vierstra J, et al. (2014) Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. *Science* 346:1007–1012.
- Maza I, et al. (2015) Transient acquisition of pluripotency during somatic cell transdifferentiation with iPSC reprogramming factors. *Nat Biotechnol* 33:769–774.
- Lin S, Coutinho-Mansfield G, Wang D, Pandit S, Fu XD (2008) The splicing factor SC35 has an active role in transcriptional elongation. *Nat Struct Mol Biol* 15:819–826.
- Tian B, Graber JH (2012) Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdiscip Rev RNA* 3:385–396.
- Di Giannardino DC, Shi Y, Manley JL (2013) PARP1 represses PAP and inhibits polyadenylation during heat shock. *Mol Cell* 49:7–17.
- Sonmez C, Dean C (2012) Transcription beyond borders has downstream consequences. *RNA Biol* 9:143–147.
- Reimand J, et al. (2016) g:Profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res* 44:W83–W89.
- Ray D, et al. (2013) A compendium of RNA-binding motifs for decoding gene regulation. *Nature* 499:172–177.