


ARTICLE

<https://doi.org/10.1038/s41467-019-11536-8>

OPEN

The pause-initiation limit restricts transcription activation in human cells

Saskia Gressel ^{1,2}, Björn Schwalb ^{1,2} & Patrick Cramer ¹

Eukaryotic gene transcription is often controlled at the level of RNA polymerase II (Pol II) pausing in the promoter-proximal region. Pausing Pol II limits the frequency of transcription initiation ('pause-initiation limit'), predicting that the pause duration must be decreased for transcriptional activation. To test this prediction, we conduct a genome-wide kinetic analysis of the heat shock response in human cells. We show that the pause-initiation limit restricts transcriptional activation at most genes. Gene activation generally requires the activity of the P-TEFb kinase CDK9, which decreases the duration of Pol II pausing and thereby enables an increase in the productive initiation frequency. The transcription of enhancer elements is generally not pause limited and can be activated without CDK9 activity. Our results define the kinetics of Pol II transcriptional regulation in human cells at all gene classes during a natural transcription response.

¹Department of Molecular Biology, Max-Planck-Institute for Biophysical Chemistry, Am Faßberg 11, 37077 Göttingen, Germany. ²These authors contributed equally: Saskia Gressel, Björn Schwalb. Correspondence and requests for materials should be addressed to B.S. (email: bjorn.schwalb@mpibpc.mpg.de) or to P.C. (email: patrick.cramer@mpibpc.mpg.de)

Gene transcription is regulated at multiple levels^{1–3}. A critical point of transcription regulation in human cells is the phase of early elongation by RNA polymerase (Pol) II^{4–6} (reviewed in ref. 7). After Pol II escapes from the promoter, it often pauses in the promoter-proximal region, and this represents a major regulatory step at protein-coding genes^{8,9}. Paused Pol II is stabilized by the factors DSIF¹⁰ and NELF¹¹, and is released into active elongation by the CDK9-containing kinase complex P-TEFb^{12,13}. Recent studies revealed structures of the Pol II elongation complex in the paused and activated state, and provided mechanistic insights into the P-TEFb dependent switch to active elongation^{14,15}. Pol II pauses also at noncoding genes that produce enhancer RNAs¹⁶, upstream antisense RNAs (uaRNAs)¹⁷, and other long noncoding RNAs^{18,19}. However, whether and to which extent pausing can restrict the transcriptional output at different gene classes was not quantified and compared *in vivo*.

Most available studies estimated the degree of Pol II pausing as the relative ratio of Pol II occupancy in the promoter-proximal region and the gene body, which has been termed the traveling ratio²⁰, or the pausing index^{21,22}. Pol II occupancy can be mapped using DNA by chromatin immunoprecipitation (IP) assays²³, and along nascent RNA transcripts by nuclear run-on assays^{8,24} or by native elongating transcript sequencing (NET-seq)^{25,26}. The latter methods are powerful tools to locate engaged Pol II in a strand-specific manner and at high resolution. Although, Pol II occupancy depends on pausing it does not directly relate to the kinetics of pausing²⁷. This is because the Pol II occupancy signal at a given time depends on the number of polymerases and their speed, and cannot be used in isolation to distinguish between these two. Indeed, when a Pol II occupancy peak increases, this can be due to an increase in the number of pausing polymerases or due to an increase in the duration of pausing, or both.

As a consequence, the time polymerases reside in a pause window can only be estimated by factoring in the number of polymerases that pass through the pause window. This number we have called the productive initiation frequency I , which is defined as the number of Pol II enzymes that initiated at the promoter and were successfully released from the pause window into productive elongation²⁸. I is therefore independent of a putative unknown fraction of polymerases that may be terminating in the pause window. The productive initiation frequency I can be measured by transient transcriptome sequencing (TT-seq), which is a sensitive genome-wide assay that captures RNA that is newly transcribed within 5 min of metabolic labeling with 4-thiouridine (4sU)²⁹. After labeling, RNA is fragmented, and only newly synthesized RNA fragments are purified and sequenced. When TT-seq is combined with mammalian NET-seq (mNET-seq), the pause duration d can also be obtained²⁸ (Fig. 1a). This multiomics approach provided evidence for a “pause-initiation limit” that restricts the productive initiation frequency at a given pause duration²⁸. Such a limit was predicted based on steric considerations²⁷, and an independent study also concluded that pausing can impair initiation³⁰.

These studies suggested that transcriptional activation of genes near the pause-initiation limit must be enabled by a decrease in the duration of Pol II pausing, but this was not investigated by a genome-wide kinetic analysis before. In order to achieve this, we first defined transcription kinetics for protein-coding and non-coding genes in steady state in human cells. We then used our multiomics approach to follow changes of productive initiation frequency I and pause duration d in a quantitative manner during the dynamic transcriptional response to heat shock over time. The heat shock response in the human hematopoietic cell line K562 was chosen because it provides a well-established model system^{31–36}, and it involves global transcriptional mechanisms

that are conserved across species (reviewed in ref. 37). It was also shown that recruitment of P-TEFb is critical for Pol II release upon heat shock from selected model genes^{38–41}.

Results presented here show that transcriptional activation of protein-coding genes involves an increase in the productive initiation frequency, as expected, and is indeed restricted by the pause-initiation limit, as predicted. We also show that inhibition of the P-TEFb kinase CDK9 impairs full activation upon heat shock, confirming that a decrease in Pol II pause duration is a critical step in gene activation. In contrast, we find that transcription of enhancers, which also involves Pol II pausing¹⁶, is generally not restricted by the pause-initiation limit. Instead, enhancer transcription can be strongly upregulated and down-regulated by changes of the productive initiation frequency alone, even when the P-TEFb kinase CDK9 is inhibited. Taken together, we describe the kinetic basis for gene regulatory strategies underlying a transcription response in human cells, and demonstrate in a quantitative manner that promoter-proximal Pol II pausing defines a pause-initiation limit that restricts gene activation by limiting the increase in productive initiation frequency.

Results

Multiomics analysis and annotation of transcription units. We first aimed at defining genome-wide Pol II kinetic parameters during steady-state conditions. To obtain the productive initiation frequency I and the promoter-proximal pause duration d ²⁸, we carried out TT-seq and mNET-seq of total Pol II (with Empigen BB) in human K562 cells (Fig. 1a). We generated TT-seq data for two independent biological replicates after 5 min of metabolic labeling with 4sU (Spearman correlation $\rho = 1.00$) as well as mNET-seq data for two independent biological replicates (Spearman correlation $\rho = 0.99$) (Supplementary Table 1; Supplementary Fig. 1). We then used the TT-seq data to create a genome-wide transcription unit (TU) annotation with the segmentation algorithm GenoSTAN⁴² (Fig. 1b). As a strategy for identifying accurate transcription start sites (TSSs) for the TT-seq derived TUs, we used published GRO-cap data⁴³, which recovers nascent RNAs with 5' caps of transcriptionally engaged Pol II. To be eligible for further analysis, annotated TUs needed a GRO-cap signal in a window of 250 bp around the GenoSTAN-derived start site, and an expression of >5 reads per base in the sum of both replicates of the TT-seq signal. We then sorted each capped TU (cTU) into one of the following seven classes using GENCODE annotation, respective location, and chromatin state annotation⁴² for enhancer and promoter classification (Fig. 2a; “Methods”): protein-coding (m), long intergenic noncoding (linc), antisense (as), convergent (con), upstream antisense (ua), short intergenic noncoding (sinc), and putative enhancer (e) RNAs.

Subsequently, we used the mNET-seq data to extract the position of paused polymerases for all cTUs in each class that showed mNET-seq signal peaks above background (“Methods”). The called pause sites were distributed around a maximum located ~50 bp downstream of the TSS (Fig. 1c), in contrast to the pause sites that were previously derived based on the TSS annotation from RefSeq, which were located ~30 bp further downstream (Supplementary Fig. 5c)²⁸. This agrees well with recently published data in K562 cells⁴⁴. We did not observe any substantial differences in the positions of called pause sites among the different classes. This resulted in a total of 10363 expressed cTUs, for which a pause site call was successful, encoding 604 eRNA, 471 asRNA, 1314 sincRNA, 965 uaRNA, 445 conRNA, 209 lincRNA, and 6355 mRNAs (Figs. 1b–d and 2a). Below we will refer to these cTUs simply as “genes” or with the respective RNA transcript class they give rise to.

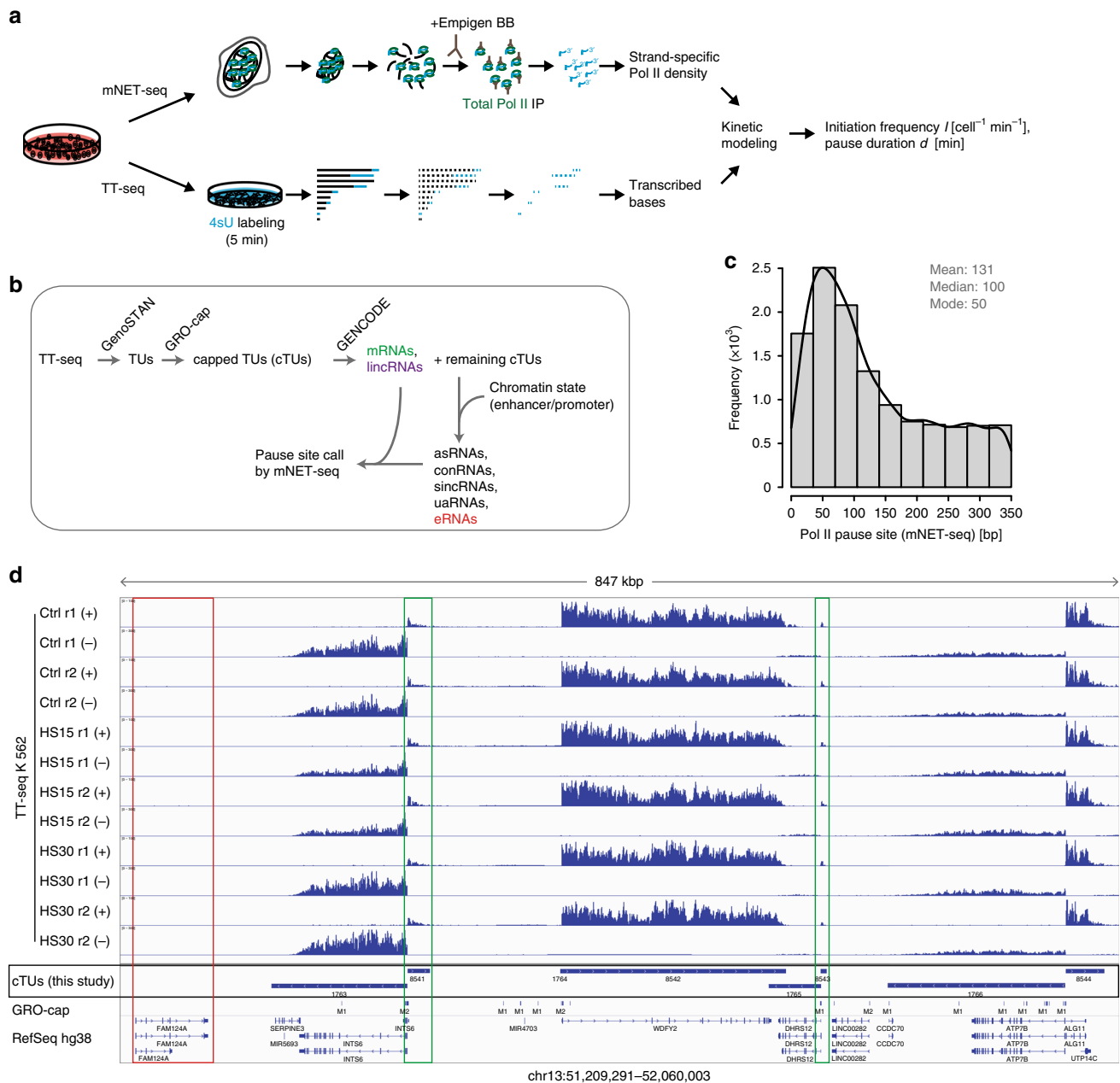


Fig. 1 Multiomics analysis and annotation of different gene classes in human K562 cells. **a** Experimental schematic of multiomics approach. Human K562 cells were subjected to TT-seq to estimate the productive initiation frequency I [$\text{cell}^{-1}\text{min}^{-1}$]. By combining TT-seq data with mNET-seq the pause duration d [min] is calculated (“Methods”)28. **b** Schematic overview of capped transcription unit (cTU) annotation in K562 cells. First, TT-seq data ($n = 6$) were subjected to the genomic segmentation algorithm GenoSTAN42. The transcription start sites (TSSs) were further refined by published GRO-cap data generated in K562 cells43. To call promoter-proximal pause sites mNET-seq data ($n = 6$) generated in K562 cells were used. (Supplementary Table 1, Supplementary Fig. 1). **c** Distribution of pause site distance [bp] for 12,160 investigated cTUs measured by mNET-seq depicted as a histogram with respect to GRO-cap refined TSS (mean 131 [bp], median 100 [bp], and mode 50 [bp]). **d** Genome browser view of a 847-kbp region located in chromosome 13 (chr13: 51,209,291-52,060,003) visualized with the Integrative Genomics Viewer (IGV, version 2.4.10; human hg38)81. From top to bottom, tracks represent: TT-seq coverage in K562 cells ($n = 6$) on respective plus (+) or minus (–) strand, GRO-cap data43, new cTU annotation, RefSeq GRCh38 annotation. cTUs missing in the RefSeq GRCh38 annotation are highlighted in a green. RefSeq transcript not present in K562 cells is highlighted in red

Transcription kinetics differ between gene classes. We next used TT-seq and mNET-seq data in combination with our previously described kinetic modeling28 to derive estimates of I and d for all annotated genes (Fig. 2a; “Methods”). We observed a reciprocal behavior of I and d for all classes of genes except enhancers (Fig. 2b–d). When I was high, d was generally low, and the other way around, consistent with an anticorrelation between these two parameters28,30. Protein coding and lincRNAs

showed the highest I value (with median initiation events of 1 and $0.3 \text{ cell}^{-1}\text{min}^{-1}$) and the lowest d values (median 1 and 2.7 min), consistent with high expression levels (Fig. 2b, d; Supplementary Figs. 2a and 6a, b). On average, lincRNAs show significantly longer pausing compared with mRNAs (p -value $< 2.2 \times 10^{-16}$, Wilcoxon rank sum test) (Fig. 2c, d; Supplementary Fig. 2a), in contrast to a recent study45. Thus, genes encoding lincRNAs initiate on average about half as frequently as protein-coding

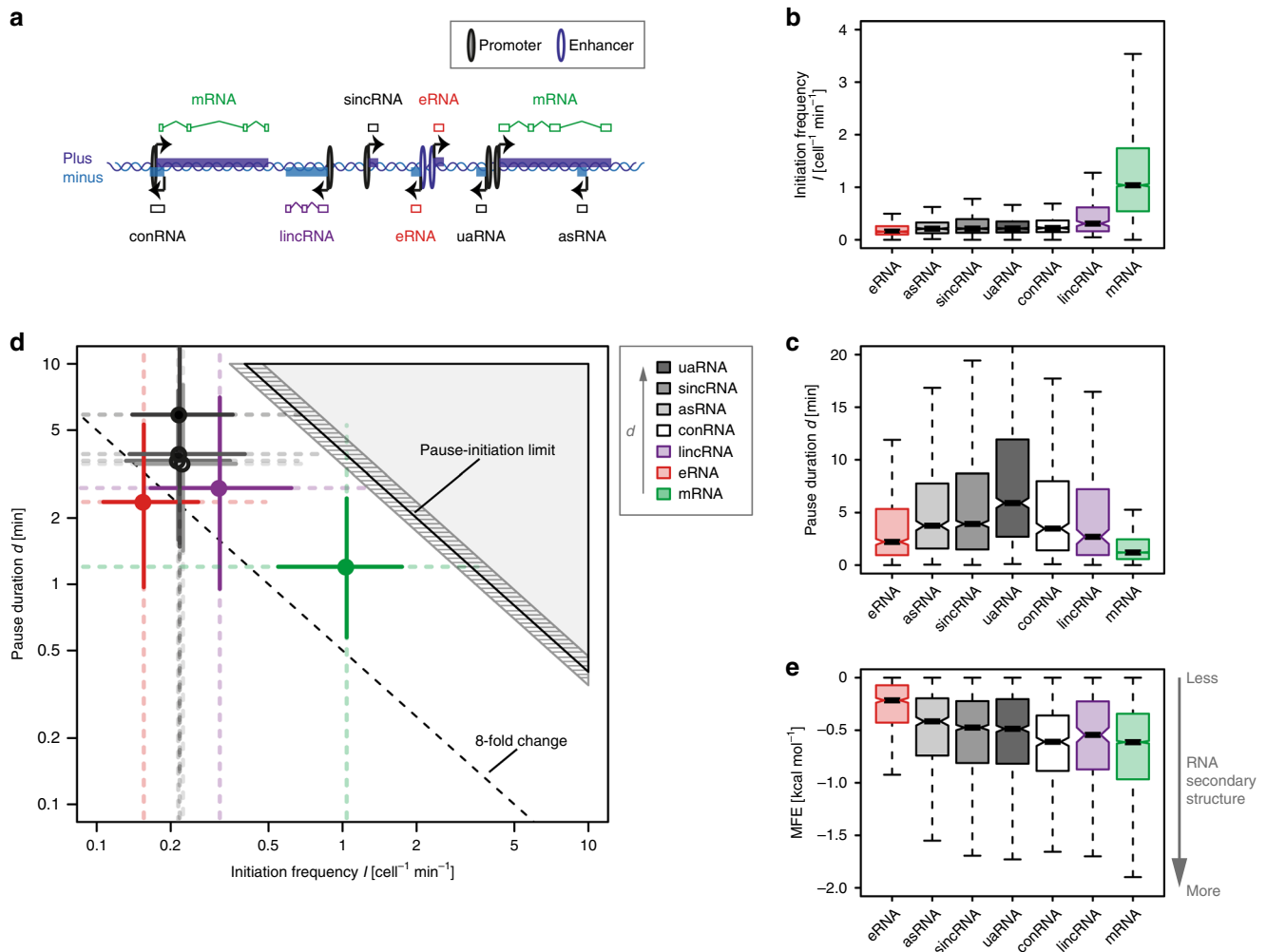


Fig. 2 Transcription kinetics of different gene classes in steady state. **a** Schematic representation of seven major transcript classes annotated in this study: 604 eRNAs (in light red), 471 asRNAs, 1314 sincRNAs, 965 uaRNAs, 445 conRNAs (in different shades of gray), 209 lincRNAs (in purple), and 6355 mRNAs (in green), depicted on plus (dark blue) or minus (light blue) strand (“Methods”). **b** Boxplots of productive initiation frequency I [$\text{cell}^{-1}\text{min}^{-1}$] for transcript classes defined in **a**. Black bars represent medians, boxes span from upper to lower quartiles, and whiskers represent 1.5 times the interquartile range. **c** Boxplots of pause duration d [min]. **d** Plot shows the median productive initiation frequency I [$\text{cell}^{-1}\text{min}^{-1}$] depicted against the median pause duration d [min] for all transcript classes (circles) in log scale. The two solid perpendicular lines define the interquartile range, and the dotted whiskers represent 1.5 times the interquartile range of the respective estimates for the entire transcript class. The gray shaded area depicts impossible combinations of I and d ^{27,28}. Striped area shows confidence intervals of the pause-initiation limit. The dotted line defines an eightfold possible fold change until a gene would be restricted by the pause-initiation limit. **e** Boxplots of minimum free energy MFE [kcal mol⁻¹]. MFE was calculated in a window of [-15, -65] bp upstream of the pause site to predict RNA secondary structure⁷⁹ for each transcript in a transcript class

genes, and smaller mNET-seq peaks for these genes indicate longer pause durations, which is counterintuitive but clearly revealed by the kinetic analysis (Supplementary Fig. 6a, b). Generally, all noncoding RNA classes (except eRNAs) showed high d values, with a median ranging around 4–6 min, explaining their low levels of expression. The low pause durations at enhancers might be explained by the fact that eRNAs do generally not adopt stable secondary structure in the nascent RNA exiting from the Pol (Fig. 2e; Supplementary Fig. 6c), which is associated with pausing²⁸. However, other chromatin features and factor availability at different loci might contribute to pausing as well (Supplementary Fig. 2b). A particularly long pause duration was observed for uaRNAs (median 6 min), and this could impair initiation events in the noncoding direction of bidirectional promoters (Fig. 2c; Supplementary Fig. 6d).

Transcription kinetics of a natural transcription response. We next investigated how transcription activation kinetics change

upon response to heat shock (Fig. 3). The optimal time points were determined by quantitative reverse transcription PCR (RT-qPCR) of the major cell stress protein HSPA1A (human Hsp70) and cell viability assays (Supplementary Fig. 3). We performed TT-seq, RNA-seq, and mNET-seq in K562 cells that were maintained under optimal growth conditions at 37 °C (control), or placed in a 42 °C water bath for 15 or 30 min (heat shock) (Fig. 3b). TT-seq, RNA-seq, and mNET-seq libraries were prepared in two biological replicates that were highly reproducible (TT-seq: $\rho > 0.99$ for all time points; RNA-seq: $\rho > 0.99$ for all time points; and mNET-seq: $\rho > 0.99$ for all time points) (Supplementary Fig. 1; Supplementary Table 1). In order to capture global changes in transcription profiles, we used a spike-in normalization strategy (“Methods”) ²⁹. This revealed that 899 genes were significantly upregulated, whereas 2614 genes were downregulated after 30 min of heat shock (Supplementary Fig. 4a). To normalize the respective mNET-seq signals in the heat shock conditions versus the control, we identified 3416 genes

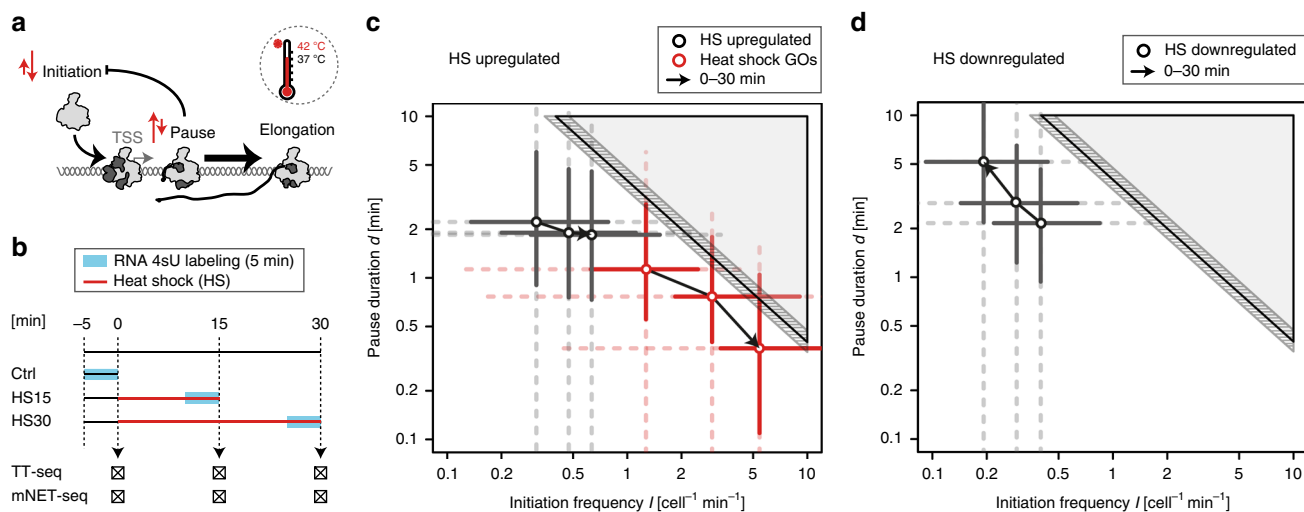


Fig. 3 Regulation of transcription upon heat shock timecourse in human K562 cells. **a** Illustration of the regulatory changes (red arrows) at a model gene upon heat shock. Pol II is depicted in silver, decorated with initiation, pausing or elongation factors (dark gray). **b** Experimental setup of the heat shock (HS) timecourse (0–30 min at 42 °C) in human K562 cells. TT-seq and mNET-seq experiments were performed in two independent biological samples (Supplementary Table 1; Supplementary Fig. 1). **c** Pause-initiation trajectories upon heat shock timecourse. Shown are 706 significantly upregulated cTUs (in gray), and 33 significantly upregulated RefSeq-TUs (in red) of overrepresented GO categories linked to heat shock (unfolded protein binding, regulation of cellular response to heat, response to unfolded protein, and chaperone binding) (Supplementary Fig. 7). Median estimates per time point (0, 15, and 30 min) are depicted as circles. The two solid perpendicular lines define the interquartile range, and the dotted whiskers represent 1.5 times the interquartile range of the respective estimates for the entire transcript class. The gray shaded area depicts impossible combinations of I and d ^{27,28}. Striped area shows confidence intervals of the pause-initiation limit. **d** Pause-initiation trajectories as in **c** upon heat shock timecourse of 1907 significantly downregulated cTUs (in gray)

that were unchanged in their TT-seq signal, and globally calibrated the mNET-seq data to show no change on these genes during the heat shock response (“Methods”).

The pause-initiation limit restricts transcription activation.

Next, we calculated the productive initiation frequency I and the pause duration d for all genes after 15 min and after 30 min of heat shock. Our results show that upon heat shock, activated genes generally show increased I and decreased d (Fig. 3c; Supplementary Fig. 5f), suggesting that gene activation requires a decrease in pause duration, which in turn allows for higher productive initiation frequencies. This behavior is even more evident at protein-coding genes linked to heat shock (Fig. 3c). This can be exemplified at the HSPA1A (HSP70) gene, which is a prominent model gene for the heat shock response (Supplementary Fig. 4b). The pause duration d of HSPA1A changes from ~30 min at steady state to ~0 min upon 30 min of heat shock which agrees well with literature estimates⁴⁶. Consistent with an anticorrelation between d and I , I increased upon heat shock to 87 productive initiation events ($\text{cell}^{-1}\text{min}^{-1}$). The results also revealed that downregulated genes exhibited decreased I and increased d (Fig. 3d; Supplementary Fig. 5g). Thus, this multiomics analysis reveals that the pause-initiation limit restricts gene regulation at genes which are located close to the limit in steady state. The use of TT-seq in this respect is critical because it directly monitors RNA synthesis activity and productive initiation frequency (see also Supplementary Note 1; Supplementary Fig. 6e).

CDK9 activity lowers the pause-initiation limit. Our kinetic modeling revealed that the activation of genes restricted by the pause-initiation limit requires a decrease in the pause duration. To corroborate this, we utilized the highly specific and rapid inhibition of an analog-sensitive CDK9 (CDK9^{as}) Raji B cell line using the bulky ATP analog 1-NA-PP1²⁸ (Fig. 4a). Raji B and

K562 cell lines are predicted to show a conserved response to heat shock with respect to timing of HSPA1A upregulation, cell viability, and GO terms of upregulated or downregulated TUs (Supplementary Figs. 3 and 7). We generated TT-seq data for two independent biological replicates after 5 min of metabolic labeling with 4sU to measure changes in I upon specific inhibition of the P-TEFb kinase CDK9 prior and during heat shock (Supplementary Table 1). TT-seq data were highly reproducible (Spearman correlation $\rho = 1.00$) (Supplementary Fig. 8a) and CDK9 kinase inhibition was very rapid (Supplementary Fig. 8b–d). We again annotated TUs genome wide with GenoSTAN (Supplementary Fig. 9; “Methods”), and this resulted in 6990 mRNAs, 3451 eRNAs, 243 lincRNAs, 1398 asRNAs, 326 conRNAs, 565 uaRNAs, and 3479 sincRNAs.

We now tested our hypothesis by investigating the changes in I upon heat shock after CDK9 inhibition (Fig. 4). We derived estimates of the productive initiation frequency I for all 12958 expressed, non-ambiguously classified genes after spike-in normalization (“Methods”). Changes in I for upregulated genes confirmed a strong dependence of transcription activation on CDK9 kinase activity. Genes were significantly less inducible during heat shock when CDK9 was inhibited (p -value = 9×10^{-9} , Wilcoxon rank sum test), confirming an obligatory decrease in pause duration for upregulation of the productive initiation frequency (Fig. 4b, left). Productive initiation events of genes encoding mRNAs decreased to 75% after CDK9 inhibition (Fig. 4c, left). Downregulation of genes was overall stronger upon CDK9 inhibition (p -value < 2.2×10^{-16} , Wilcoxon rank sum test) (Fig. 4b, right), indicating less possible initiation events (33% for mRNAs) due to even longer pause durations compared with downregulation upon heat shock alone (Fig. 4c, right). Thus, CDK9 activity lowers the pause duration to allow for high gene activation that is restricted by the pause-initiation limit.

Enhancer transcription is generally not pause limited. Amongst all seven gene classes, enhancers showed the greatest distance

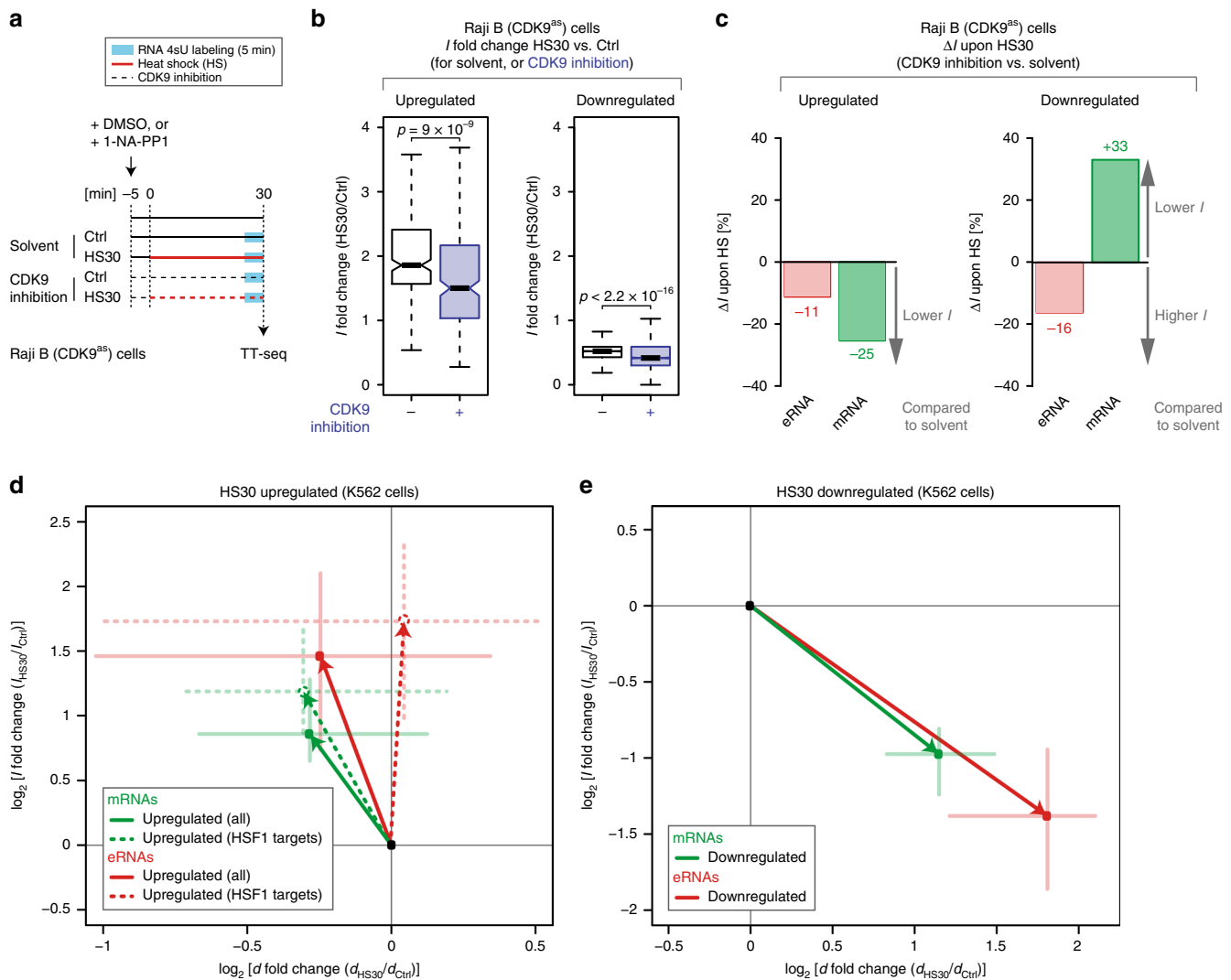


Fig. 4 Enhancer transcription is generally not pause limited and less dependent on CDK9. **a** Experimental setup. The adenine analog 1-NA-PP1 allows for rapid and highly specific inhibition of analog-sensitive CDK9⁸² in CRISPR-Cas9 engineered human Raji B (CDK9^{as}) cells²⁸ (Supplementary Fig. 8b–d). CDK9 was inhibited using 5- μ M 1-NA-PP1 (CDK9 inhibition) in combination with heat shock at 42 °C for 30 min in human Raji B (CDK9^{as}) cells. DMSO was used as solvent control. **b** Boxplots of productive initiation frequency *I* fold change before (in white) and after CDK9 kinase inhibition (in dark blue). Shown are 241 significantly upregulated genes (left boxplot), and 2795 significantly downregulated genes (right boxplot) annotated in Raji B (CDK9^{as}) cells (Supplementary Fig. 9). **c** ΔI upon upregulation or downregulation between heat shock with CDK9 inhibition (CDK9 inhibited HS30) and heat shock with solvent control (Solvent HS30) is shown in percent [%]. Left: bar plot comparing productive initiation frequency *I* change (ΔI) with and without CDK9 inhibition for 92 significantly upregulated mRNAs (in green), and 54 significantly upregulated eRNAs (in red) annotated in Raji B (CDK9^{as}) cells. Right: bar plot comparing productive initiation frequency *I* change (ΔI) with and without CDK9 inhibition for 2210 significantly downregulated mRNAs (in green), and 223 significantly downregulated eRNAs annotated in Raji B (CDK9^{as}) cells. **d** \log_2 fold change of pause duration *d* and initiation frequency *I* for 336 significantly upregulated mRNAs (green solid line), and 67 significantly upregulated eRNAs (red solid line) in K562 cells upon 30 min of heat shock (HS30) (Supplementary Fig. 4a). Dashed lines represent HSF1 driven (“Methods”) subsets of 91 mRNAs and 20 eRNAs. **e** \log_2 fold change of pause duration *d* and initiation frequency *I* for 1101 significantly downregulated mRNAs (in green), and 99 significantly downregulated eRNAs (in red) in K562 cells upon 30 min of heat shock (HS30) (Supplementary Fig. 4a)

from the pause-initiation limit at steady state in K562 cells, thus forming a notable exception from other gene classes (Fig. 2d). Enhancers showed a median pause duration of 2.2 min and a median productive initiation frequency of only one initiation event every 6–7 min. This indicated that enhancers were generally not restricted by the pause-initiation limit, i.e., *I* could increase several fold without any change in pause duration, until the pause-initiation limit would be reached. Indeed, upon heat shock, the productive initiation frequency *I* for enhancers was increased 1.5-fold more than for mRNAs given the same change in pause duration *d* (Fig. 4d, solid arrows). This difference is even stronger

when comparing heat shock factor 1 (HSF1) targeted mRNAs and eRNAs (Fig. 4d, dotted arrows). HSF1 is a major activator in heat shock induced transcription upregulation⁴⁷. HSF1 driven eRNAs can be activated without a change in pause duration, while HSF1 driven mRNAs still require a shortening of the pause duration.

Another implication of this exceptional kinetic behavior is that enhancer transcription can generally only be reduced by a strong increase (>8-fold, Fig. 2d) of the pause duration *d*. However, downregulation of enhancer transcription is in line with the general observation of prolonged pausing to inhibit new initiation events, as is the case for mRNA synthesis (Fig. 4e). In conclusion,

enhancers differ from protein-coding genes, because their productive initiation frequency appears generally not to be restricted by the pause duration.

Enhancer transcription is less dependent on CDK9. Although enhancer transcription is generally not limited by pausing in K562 cells at steady state (Fig. 2d), it remained unclear whether enhancer transcription is controlled by P-TEFb. Using mNET-seq data in Raji B (CDK9^{as}) cells²⁸ we found that the median pause duration of all transcript classes and the exceptional role of eRNAs is conserved in unperturbed Raji B cells (Supplementary Fig. 10a). We confirmed this also in the context of the upregulation of enhancer transcription upon heat shock in K562 cells. Enhancers showed higher initiation frequency fold changes provided the same fold change in pause duration as found for protein-coding genes (Fig. 4d). Consistent with the results in K562, activation of enhancer transcription in Raji cells was only reduced by 11% upon CDK9 inhibition (Fig. 4c; Supplementary Fig. 10c).

This shows that enhancer transcription can be activated even when CDK9 is inhibited. The overall behavior of impaired activation after CDK9 inhibition for all gene classes resembles the pause durations calculated for Raji cells²⁸ and strongly supports our estimates (Fig. 2d; Supplementary Fig. 10b). Surprisingly, downregulated enhancers were not repressed by inhibition of CDK9, consistent with our assumption that higher pause durations do not cause lower productive initiation frequencies at enhancers (Fig. 4c, right; Supplementary Fig. 10d). Taken together, enhancer transcription and thus eRNA synthesis can be upregulated and downregulated to a large extent without changes in pause duration.

Discussion

To understand genome function, the regulatory steps of gene transcription must be defined and it must be analyzed under which conditions they become rate limiting. A rate-limiting step may be defined as the slowest molecular transition in the process that limits the overall progression and the transcriptional output (reviewed in refs. 2,48). We recently showed that prolonged promoter-proximal pausing of Pol II impairs new initiation, and thus reduces the amount of mRNA synthesized per time (“pause-initiation limit”)²⁸. Others have also provided evidence that pausing controls initiation³⁰. However, changes in the kinetics of Pol II initiation and pausing have not been quantified genome-wide during a transcription response, which is required as definitive evidence that natural gene regulation is controlled by pausing kinetics.

Here we quantified transcription kinetics of protein-coding and noncoding genes in steady state and during the dynamic transcriptional response of human cells to heat shock. To this end, we annotated protein-coding RNAs (mRNAs), and six major non-coding transcript classes, i.e., lincRNAs, asRNAs, eRNAs, uaRNAs, conRNAs, and sincRNAs in human hematopoietic cell lines (K562 and Raji B) (Fig. 1; Supplementary Fig. 9a). We then used a multiomics approach to follow changes in productive initiation frequency I and pause duration d in a quantitative manner.

In cells at steady state, we observed a reciprocal behavior of I and d for genes encoding all transcript classes except eRNAs. Protein coding and lincRNAs were among the classes with the shortest pause durations, consistent with their high transcription levels. The longest median pause duration was observed for uaRNAs, presumably impairing initiation events in the noncoding direction of bidirectional promoters (Fig. 2). So far, it was proposed that transcriptional regulation upon heat shock is coordinated at the single step of promoter-proximal pause

release^{36,49}. Here we could show that this holds true for genes that are close to the pause-initiation limit such as protein-coding genes (Fig. 3). Using the transcriptional response to heat shock, we show that upregulation of the productive initiation frequency is restricted by the P-TEFb kinase CDK9. Enhancers form a notable exception to this rule, because changes in pause duration do not cause changes in eRNA production during heat shock (Fig. 4; Supplementary Fig. 10c, d).

More generally, upregulation of a gene requires an increase in initiation frequency, which leads to a higher number of polymerases loaded onto the gene, and a higher amount of RNA synthesis over time. At genes that are at the pause-initiation limit, pausing limits initiation, and a decrease in pause duration is required for upregulation of transcription by allowing for higher initiation frequencies. This is apparently often the case at protein-coding genes. In contrast, upregulation of enhancer transcription is often possible without changes in pause duration because enhancers are generally not pause limited. These observations lead to the following hypothetical model of gene activation. Gene activation would start with an upregulation of enhancer transcription by an increase in the frequency of initiation from enhancer regions. This would go along with a decrease in pause duration at the protein-coding target genes, which in turn allows for an increase of the productive initiation frequency at these genes. These mechanisms however rely on the availability of polymerases and transcription factors, and assume that transcription is generally processive.

Provided the critical role of decreasing the pause duration for gene activation, the mechanisms of Pol pause release should be studied further in the future. Note that our model holds true independent of the percentage of premature termination that might occur before the promoter-proximal pause site as it quantifies the effective pause between two initiation events that successfully lead to productive elongation. A role in pause release upon heat shock has been reported for the following factors: the TFIIF-associated kinase CDK7^{50–52}, the elongation factor TFIIS^{53,54}, the DNA-PK kinase, the ATM kinase, the 7SK snRNP recruitment factor TRIM28/KAP1, the pause stabilizing factor GRINL1A/GDOWN1^{55,56}, and the CTD phosphatase FCP1⁵⁷. Methods developed here and elsewhere²⁸ can now be used to study the kinetics underlying the mechanisms of P-TEFb delivery and activation in a quantitative and genome-wide manner, ultimately unraveling the nature of gene regulation in metazoan cells.

Methods

Cell culture. Human K562 erythroleukemia cells were obtained from DSMZ (Cat. # ACC-10). Human Raji Burkitt's (B) lymphoma (CDK9^{as}) cells carry homozygous mutation of phenylalanine (F) 103 to alanine (A) at the CDK9 gene loci and were generated using the CRISPR-Cas9 system²⁸. K562 and Raji B (CDK9^{as}) cells were cultured antibiotic free in accordance with the DSMZ cell culture standards in RPMI 1640 medium (Thermo Fisher Scientific) containing 10% heat inactivated fetal bovine serum (Thermo Fisher Scientific), and 1× GlutaMAX supplement (Thermo Fisher Scientific) at 37 °C in a humidified 5% CO₂ incubator. Both cell lines used in this study display the phenotypic properties, including morphology and proliferation rate, that have been described in the literature. Cells were verified to be free of mycoplasma contamination using Plasmoguard Mycoplasma Detection Kit (InvivoGen). Biological replicates were cultured independently.

Heat shock treatment. To avoid transcriptional changes by freshly added growth medium, fresh growth medium was added ~24 h prior to heat shock treatments. Heat shock treatments of K562 or Raji B (CDK9^{as}) cells were performed in T175 flasks in a volume of 50 mL at 0.6×10^6 cells mL⁻¹ in a water bath (LAUDA, Aqualine AL12) at 42 °C. Temperature was monitored by thermometer. It took 5 min until the cell suspension reached 42 °C. For RT-qPCR and cell viability assessment, cells were treated for a timecourse of 0–75 min. For TT-seq, RNA-seq and mNET-seq experiments, cells were treated for 0 (Ctrl), 15 (HS15), or 30 min (HS30).

Cell viability assessment by trypan blue exclusion test. Cell viability levels were evaluated by the trypan blue exclusion method as described by Strober⁵⁸. Cell viability assessment was performed with two biological replicates. Briefly, 5×10^5 cells were treated as above for 0, 15, 30, 45, 60, or 75 min. Cells were pelleted at $200 \times g$ for 5 min and resuspended in 1-mL DPBS prior to counting. Equal volumes of cell suspension and 0.4% trypan blue solution (Sigma-Aldrich) were mixed and incubated for 2 min. The solution was applied to a hemacytometer and viable cells were counted using light microscopy. For each treatment time, the cell count was duplicated (dilution factor for trypan blue) and the average value was obtained. Cell viability was calculated as the ratio of viable cells upon heat shock (15–75 min) to viable cells of control (0 min).

Total RNA isolation and RT-qPCR. K562 or Raji B (CDK9^{as}) cells were treated as above for 0–75 min. For each time point and biological replicate ($n = 2$), 5×10^6 cells were harvested at $3000 \times g$ for 2 min. Total RNA was isolated with QIAzol (QIAGEN) according to the manufacturer's instructions except for the addition of 10 ng RNA spike-in mix²⁹ together with QIAzol. To remove possible genomic DNA contamination, isolated RNA (10 μ g) was treated with TURBO DNase (Thermo Fisher Scientific) according to the manufacturer's instructions. For reverse transcription (RT), random hexamer priming (5'-d(NNNNNN)-3', N = G, A, T, or C) was used according to the manufacturer's instructions. Briefly, 1 μ g of DNase-treated RNA, random hexamer primers (final concentration of 5 ng/ μ L), and dNTPs mix (final concentration of 0.5 mM) were mixed and incubated at 65 °C for 5 min. Subsequently, Maxima H Minus reverse transcriptase (RT) (final concentration of 200 U) and $5 \times$ Maxima RT buffer (Thermo Fisher Scientific) were added (+RT reaction). For DNA contamination control, cDNA synthesis without RT (-RT reaction) was performed (RT was substituted with water). The (-/+RT) reactions were incubated in a PCR cycler at 25 °C for 10 min, 50 °C for 30 min, and 85 °C for 5 min. Primers for quantitative PCR (qPCR) were designed by using the online primer design software Primer3 v.0.4.0⁵⁹. Primer specificity (single product peak) was validated by melting profiles. Primer sequences, length, annealing temperature, amplicon length, and position on target are reported in Supplementary Table 3. cDNAs (50 ng) were amplified with SYBR[®] Select Master Mix (Thermo Fisher Scientific) according to the manufacturer's instruction with a final primer concentration of 400 nM. PCR reactions were run in 96-well optical plates sealed with optical adhesive cover on a qTOWER 2.0/2.2 instrument (Analytik Jena AG). The following thermal cycling conditions were used (SYBR Select Master Mix reference, standard cycling mode): 50 °C for 2 min, 95 °C for 2 min, 40 cycles of 95 °C for 15 s, and 60 °C for 1 min. Three synthetic RNA spike-ins were used for normalization. The 2(-delta delta Ct) method was applied to calculate the normalized target gene expression fold change, with the amplification efficiency (E) for each target gene, slope of standard curve (S), and mean threshold cycle (Ct)⁶⁰.

RNA spike-ins. Synthetic RNA spike-in controls are derived from selected RNAs of the ERCC RNA Spike-in Mix (Ambion) as described in ref. 29. Briefly, spike-ins (three unlabeled and three 4sU-labeled) are in vitro transcribed using the MEGAscript T7 kit (Ambion). In vitro transcription (IVT) of unlabeled spike-ins was performed following the manufacturer's instruction. For IVT of 4sU-labeled spike-ins, 10% of UTP was substituted with 4-thio-UTP (Jena Bioscience). This is to ensure at least similar 4sU incorporation rates in the IVT as has been observed in human cell lines^{61–63}. RNA spike-ins were purified with RNAClean XP beads (Beckman Coulter) following the manufacturer's instructions. The final RNA spike-in pool contained equal amounts of all RNA spike-ins.

TT-seq and RNA-seq. A detailed step-by-step protocol has been deposited in the protocols.io repository⁶⁴. Two biological replicates of reactions including RNA spike-ins were performed essentially as described²⁹. Experiments were performed using $\sim 3.6 \times 10^7$ cells per biological replicate. K562 cells were kept at optimal growth conditions (Ctrl), or subjected to heat shock at 42 °C as described above for 15 min (HS15) or 30 min (HS30). Cells were labeled with 4-thiouridine (4sU) (Sigma-Aldrich) for the last 5 min. Raji B (CDK9^{as}) cells were pretreated with solvent (0.05% v/v DMSO) or 5 μ M 1-NA-PP1 (EMD Millipore) at 37 °C for 5 min. Afterward, Raji B (CDK9^{as}) cells were kept at 37 °C (Ctrl) or shifted to 42 °C for additional 30 min (HS30), and 4sU-labeled for the last 5 min. An overview for all treatment conditions prior to 5 min of 4sU labeling is provided in Supplementary Table 1. Cells were harvested by centrifugation at $3000 \times g$ for 2 min. Total RNA was isolated with QIAzol reagent (QIAGEN) according to the manufacturer's instructions⁶⁵ except for the addition of 150 ng RNA spike-in mix together with QIAzol. RNAs (150 μ g) were sonicated to generate fragments of <10 kbp (total fragmented RNA) using AFAMicro tubes in a S220 Focused-ultrasonicator (Covaris Inc.). 4sU-labeled RNA was purified from 300 μ g (K562) or 600 μ g (Raji B CDK9^{as}) of total fragmented RNA. Separation of labeled RNA was achieved with streptavidin beads (Miltenyi Biotec) as described^{29,66}. Elution was performed using freshly prepared 100 mM DL-Dithiothreitol (Sigma-Aldrich). Prior to library preparation, 4sU-labeled RNA (TT-seq) and total fragmented RNA (RNA-seq) were purified using RNAClean XP beads (Beckman Coulter) and quantified by Qubit High Sensitivity RNA kit (Thermo Fisher Scientific). Input RNA was treated with HL-dsDNase (ArcticZymes) and used for strand-specific library preparation according to the Ovation Universal RNA-Seq System (NuGEN) with minor

modifications. For “first strand primer premix preparation” only random primers were used. The user's guide instructions were followed from “first strand synthesis using DNase-treated RNA” to “adaptor cleavage.” The precise number of PCR amplification cycles was determined using $\sim 10\%$ of the library and the KAPA HIFI Library Amp Real Time kit (Kapa Biosystems) following the manufacturer's instructions. For the remaining 45 μ L of library, the determined number of PCR cycles was used following the user guide's instructions for amplification as described in “using qPCR to determine the number of PCR cycles.” The purified and preamplified fragments were analyzed on a Fragment Analyzer (Agilent) before clustering and sequencing on a NextSeq 550 (Illumina) in paired-end mode with 75 bp read length.

TT-seq/RNA-seq preprocessing and normalization parameters. Paired-end 75 bp reads with additional six bases of barcodes were obtained for each of the samples (Supplementary Table 1). Reads were demultiplexed and mapped with STAR 2.3.0⁶⁷ to the hg20/hg38 (GRCh38) genome assembly (Human Genome Reference Consortium). Samtools⁶⁸ was used to quality filter SAM files, whereby alignments with MAPQ smaller than seven ($-q 7$) were skipped and only proper pairs ($-f 2$) were selected. Further data processing was carried out using the R/Bioconductor environment. We used a spike-in (RNAs) normalization strategy essentially as described²⁹ to allow observation of global shifts (sequencing depth) σ_j , cross-contamination rate ϵ_j (proportion of unlabeled reads purified in the TT-seq samples), and antisense bias ratio c_j (ratio of spurious reads originating from the opposite strand introduced by the RT reaction). Read counts (k_{ij}) for spike-ins were calculated using HTSeq⁶⁹. Calculations for each parameter are described in the following in more detail.

Antisense bias ratio c_j . Antisense bias ratios were calculated for each sample j according to

$$c_j = \text{median}_i \left(\frac{k_{ij}^{\text{antisense}}}{k_{ij}^{\text{sense}}} \right) \quad (1)$$

with read counts k_{ij} for all available spike-ins i in sample j .

Sequencing depth σ_j and cross-contamination rate ϵ_j . Sequencing depths were calculated for each sample j according to

$$\sigma_j = \text{median}_i \left(\frac{k_{ij}}{l_i} \right) \quad (2)$$

with read counts k_{ij} for all available spike-ins i in sample j for the RNA-seq samples and the labeled spike-ins i in sample j for the TT-seq samples. The cross-contamination rate ϵ_j was calculated for each sample j as

$$\epsilon_j = \frac{\text{median}_i \left(\frac{k_{ij}}{l_i} \right)}{\sigma_j} \quad (3)$$

using the unlabeled spike-ins i for TT-seq samples. Note that ϵ_j is set to 1 for the RNA-seq samples.

TUs based on the UCSC RefSeq (RefSeq-TUs). For each annotated gene, TUs were defined as the union of all existing inherent transcript isoforms (UCSC RefSeq GRCh38).

Isoform-independent exonic regions (constitutive exons). Isoform-independent exonic regions were determined using a model for constitutive exons⁷⁰ based on UCSC RefSeq annotation (GRCh38).

GenoSTAN annotation of TUs. Annotation of different transcript classes was done as in ref. 29 with minor differences. In brief, genome-wide coverage was calculated from all TT-seq fragment midpoints in consecutive 200-bp bins throughout the genome. In order to create a comprehensive annotation independent of heat shock induced length differences, two replicate tracks were constructed by taking the maximum of each bin over the first and second replicates, respectively, regardless of treatment. A two-state hidden Markov model with a Poisson-log-normal emission distribution was learned in order to segment the genome into “transcribed” and “untranscribed” states. Consecutive “transcribed” states were joined, if its gaps were smaller than 200 bp, within a validated GENCODE mRNA or lincRNA (version 22) or showed uninterrupted coverage supported by all TT-seq samples. Subsequently, TU start and end sites were refined to nucleotide precision by finding borders of abrupt coverage increase or decrease between two consecutive segments in the four 200-bp bins located around the initially assigned start and stop sites via fitting a piecewise constant curve to the TT-seq coverage profiles for both replicates using the segmentation method from the R/Bioconductor package “tilingArray”⁷¹.

GRO-cap TSS refinement of TUs (cTUs, K562). For all TUs i , the GRO-cap refined ts^* was determined as the closest nonzero GRO-cap signal⁴³ in a window of 500 bp around the start of the TUs. Note that all TUs without an assigned GRO-cap site were not used. It was recently shown that upon 1 h of heat shock the TSS architecture remains mostly unchanged⁴⁴. Thus, we assume changes of the TSS architecture upon 30 min of heat shock to be insignificant.

Transcript sorting (K562 and Raji). We sorted each gene (cTU for K562, TU for Raji) into one of the following seven classes: eRNA, sincRNA, asRNA, conRNA, uaRNA, lincRNA, and mRNA. First, (c)TUs reciprocally overlapping by at least 50% with a validated GENCODE mRNA or lincRNA (version 22) on the same strand were classified as mRNAs and lincRNAs. (c)TUs reciprocally overlapping by less than 50% with a validated GENCODE mRNA or lincRNA (version 22) on the same strand were not classified. Next, (c)TUs located on the opposite strand of either a mRNA or lincRNA were classified as asRNA if the TSS was located >1 kb downstream of the sense TSS on the opposite strand, as uaRNA if its TSS was located <1 kb upstream of the sense TSS, and as conRNA if its TSS was located <1 kb downstream of the TSS on the opposite strand. For K562, each of the remaining cTUs were classified as sincRNA. Every ncRNA (sincRNA, asRNA, conRNA, or uaRNA) was reclassified as eRNA if its TSS fell into a K562 enhancer state⁴². This resulted in 11324 non-ambiguously classified RNAs in K562 cells (825 eRNAs, 1581 sincRNAs, 564 asRNAs, 502 conRNAs, 1064 uaRNAs, 239 lincRNAs, and 6549 mRNAs). For Raji, each of the remaining TUs were classified into eRNA—if its TSS exhibited a high (>1) ratio H3K4me1/H3K4me3—or as sincRNA—if its TSS exhibited a low (<1) ratio of H3K4me1/H3K4me3. This resulted in 16452 non-ambiguously classified RNAs (3451 eRNAs, 3479 sincRNAs, 1398 asRNAs, 326 conRNAs, 565 uaRNAs, 243 lincRNAs, and 6990 mRNAs).

Calculation of the number of transcribed bases. Of all sequenced fragments, only those were kept that exhibited a positive inner mate distance. The number of transcribed bases (tb_{*j*}) for all samples was calculated as the sum of the coverage of evident (sequenced) fragment parts (read pairs only) for all fragments with an inner mate interval not entirely overlapping a Refseq annotated intron (UCSC RefSeq GRCh38, ~98.5% of all fragments) in addition to the sum of the coverage of nonevident fragment parts (entire fragment).

TT-seq/RNA-seq processing with normalization parameters. The number of transcribed bases (tb_{*j*}) or read counts (k_j) for all features ((c)TUs, constitutive exons, or RefSeq-TUs) were normalized and corrected for antisense bias c_j , sequencing depth σ_j (K562 and Raji), and cross-contamination rate ϵ_j (K562) as follows using the parameters calculated as described above. Note that cross-contamination rate estimates for TT-seq in Raji were very low and were thus not corrected for.

Antisense bias correction. The real number of read counts or coverage s_{ij} for transcribed unit i in sample j was calculated as

$$s_{ij} = \frac{S_{ij} - c_j A_{ij}}{1 - c_j^2} \quad (4)$$

where S_{ij} and A_{ij} are the observed numbers of read counts or coverage on the sense and antisense strand.

Sequencing depth and cross-contamination correction. The antisense bias corrected read counts or coverage s_{ij} of transcribed unit i in sample j was normalized for sequencing depth and cross-contamination as

$$t_{ij} = \frac{\frac{s_{ij}^{\text{TT-seq}}}{\sigma_j^{\text{TT-seq}}} - \epsilon_j \frac{s_{ij}^{\text{RNA-seq}}}{\sigma_j^{\text{RNA-seq}}}}{1 - \epsilon_j} \quad (5)$$

Note that σ_j was balanced between replicates via classical size factor normalization⁷² to gain statistical power in the differential expression analysis. Also note that, the cross-contamination rate ϵ_j was set to 0 in the formula above for TT-seq in Raji.

Read counts per kilobase (RPK). RPKs were calculated upon antisense bias corrected read counts (k_j) falling into the region of a cTU, constitutive exon or RefSeq-TU divided by its length in kilobases.

Number of transcribed bases per base (coverage). Coverages were calculated upon antisense bias corrected number of transcribed bases (tb_{*j*}) falling into the region of a cTU divided by its length in bases. Based on the antisense bias corrected coverages a subgroup of expressed cTUs was defined to comprise all cTUs with a coverage of five or higher in one of two summarized replicates of TT-seq (HS15, HS30, or control).

mNET-seq. Two biological replicates of reactions including Empigen BB detergent treatment during IP were performed essentially as described^{28,45,73}. See Supplementary Table 1 for an overview of experimental conditions and biological replicates. Briefly, cells were kept at optimal growth conditions (Ctrl) or subjected to heat shock at 42 °C as described above for 15 min (HS15) or 30 min (HS30). Experiments were performed using 1.0×10^8 K562 cells per biological replicate. Per reaction, biochemical fractionations of K562 cells were performed as described⁷⁴. All buffers were freshly prepared and complemented with Protease Inhibitor Cocktail (Sigma-Aldrich) and PhosSTOP (Roche). Isolated chromatin was digested with 50 U micrococcal nuclease (MNase) (NEB) at 37 °C and 1400 r.p.m. in a ThermoMixer (Eppendorf) for 2 min. To inactivate MNase, EGTA (Bioworld) was added to a final concentration of 25 mM. Digested chromatin was collected by centrifugation at 4 °C and $16,000 \times g$ for 5 min. The supernatant was diluted eightfold with IP buffer containing 50 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.05% v/v NP-40, and 1% v/v Empigen BB (~30% active substance, Sigma-Aldrich). For IP, 30 µg of antibody targeting total Pol II, i.e., hRPB1 (phosphorylated and unphosphorylated) CTD (MABI0601; BIOZOL), was conjugated to Dynabeads M-280 Sheep Anti-Mouse IgG (Thermo Fisher Scientific) (Supplementary Table 4). The IP was performed on a rotating wheel at 4 °C for 1 h. The beads were washed six times with IP buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05% NP-40, and 1% Empigen BB) and once with 500 µL of PNK reaction mix containing $1 \times T4$ polynucleotide kinase (PNK) buffer (NEB) and 0.1% v/v Tween-20 (Sigma-Aldrich). Beads were incubated in 100 µL of PNK reaction mix containing $1 \times$ PNK buffer, 0.1% v/v Tween-20, 1 mM ATP, and T4 PNK 3' phosphatase minus (NEB) at 37 °C for 10 min. Beads were washed once with IP buffer. RNA was extracted with TRIzol reagent (Thermo Fisher Scientific), precipitated in ethanol, and resolved on a 6% denaturing PAGE containing 7 M urea for size purification⁷⁵. Bromophenol blue and xylene cyanol were used as tracking dyes. To select fragments of 25–110 nt, the gel fragment was cut between the two tracking dyes, eluted from the gel by the crush-and-soak method using elution buffer (1 M NaOAc, 1 mM EDTA), and precipitated in ethanol. RNA libraries were prepared according to the TruSeq Small RNA Library Kit (Illumina) and as described in ref. ⁷³. For final size selection of the amplified library, 4% E-Gel™ High-Resolution Agarose Gels (Invitrogen) were used. The size-selected fragments were analyzed on a Fragment Analyzer (Agilent) before clustering and sequencing on a NEXTseq 550 (Illumina) in paired-end mode with 75bp read length.

mNET-seq data preprocessing. Paired-end 75 bp reads with additional six bases of barcodes were obtained for each of the samples (Supplementary Table 1). Reads were demultiplexed, trimmed for adapter content with cutadapt⁷⁶ (-O 12 -m 25 -a TGGAATTCTCGG -A GATCGTCGGACT), and mapped with STAR 2.3.0⁶⁷ to the hg20/hg38 (GRCh38) genome assembly (Human Genome Reference Consortium). Samtools⁶⁸ was used to quality filter SAM files, whereby alignments with MAPQ smaller than 7 (-q 7) were skipped and only proper pairs (-f 2) were selected. Further data processing was carried out using the R/Bioconductor environment. Antisense bias (ratio of spurious reads originating from the opposite strand introduced by the RT reactions) was determined using positions in regions without antisense annotation with a coverage of at least 100 according to RefSeq-TUs (UCSC RefSeq GRCh38). Coverage tracks for further analysis were restricted to the last nucleotide incorporated by the Pol in the aligned mNET-seq reads.

mNET-seq data normalization. We first identified a subgroup of RefSeq-TUs with unchanged behavior over the response to heat shock in the spike-ins normalized TT-seq data via k-means clustering. On the resulting 3416 RefSeq-TUs i , size factors for each sample j were determined as

$$\sigma_j = \text{median}_i \left(\frac{p_{ij}}{\left(\prod_{v=1}^m p_{ij} \right)^{\frac{1}{m}}} \right) \quad (6)$$

where m denotes the total number of antisense corrected mNET-seq samples (p_{ij}). This formula has been adapted⁷² and was used to correct for library size and sequencing depth variations.

Detection of pause sites. For all expressed (c)TUs or RefSeq-TUs (exceeding 10 kbp in length with one unique TSS given all RefSeq annotated isoforms (UCSC RefSeq GRCh38)) i the pause site n^* was calculated for all bases m in a window of 350 bp downstream of the GRO-cap refined TSS and in a window from the TSS to the end of the first exon (excluding the last five bases) for RefSeq-TUs via maximizing the function

$$\rho_i = \max_m p_{im} \quad (7)$$

where ρ_i needed to exceed five times the median of the signal strength p_{im} for all nonnegative antisense bias corrected mNET-seq coverage values. In order to maximize the chances of finding the most likely pause site, two replicate tracks were constructed by taking the maximum of each nucleotide over the first and second replicates, respectively, regardless of treatment. We defined the pause site at the active site based on structural information¹⁵. The RNA-DNA hybrid within the paused Pol is in the tilted state that hinders nucleotide addition at the active site.

Thus, the subsequent nucleotide is not added yet. We defined the pause site to be the position in line with the “posttranslocated” RNA rather than with the ‘pre-translocated’ DNA. In conclusion, the pause site was calculated as $n^* = m^* + 1$, where m^* is the argument that maximizes ρ_i . For K562, this resulted in 10363 expressed, non-ambiguously classified RNAs (604 eRNA, 471 asRNA, 1314 sincRNA, 965 uaRNA, 445 conRNA, 209 lincRNA, and 6355 mRNA). For Raji, this resulted in 8145 expressed, non-ambiguously classified RNAs (501 eRNA, 318 asRNA, 500 sincRNA, 461 uaRNA, 253 conRNA, 145 lincRNA, and 5967 mRNA).

Molecular weight conversions. The known sequence and mixture of the utilized spike-ins allow to calculate a conversion factor to RNA amount per cell [cell^{-1}] given their molecular weight assuming perfect RNA extraction. The number of spike-in molecules per cell N [cell^{-1}] was calculated as

$$N = \frac{m}{Mn} N_A \quad (8)$$

with the mass m 25×10^{-9} [g] per spike-in, the number of cells n (3.8×10^7 for K562, 3.4×10^7 for Raji), the Avogadro constant N_A $6.02214085774 \times 10^{23}$ [mol^{-1}] and molar mass (molecular weight) of the spike-ins M [g mol^{-1}] calculated as

$$M = A_n \times 329.2 + (1 - \tau) \times U_n \times 306.2 + C_n \times 305.2 + G_n \times 345.2 + \tau \times U_n \times 322.26 + 159 + \tau \times U_n \times 322.26 + 159 \quad (9)$$

where A_n , U_n , C_n , and G_n are the number of each respective nucleotide within each spike-in polynucleotide. τ is set to 0.1 in case of a labeled spike-in and 0 otherwise ($\tau \cdot U_n$ corresponds to the number of 4sU nucleotides, 4sU $_n$). The addition of 159 to the molecular weight considers the molecular weight of a 5' triphosphate. Provided the above the conversion factor to RNA amount per cell κ [cell^{-1}] can be calculated as

$$\kappa = \text{mean} \left(\text{median}_i \left(\frac{tb_i}{L_i \times N} \right) \right) \quad (10)$$

for all labeled spike-in species i with length L_i . Note that imperfect RNA extraction efficiency would lead to an underestimation of cellular labeled RNA in comparison to the number of added spike-ins and thus to an underestimation of initiation frequencies. In case of a strong underestimation, however, the real initiation frequencies would lie above the pause-initiation limit, which is theoretically impossible. Thus, we assume this effect to be insignificant.

Estimation of productive initiation frequency I . The antisense bias corrected number of transcribed bases $tb_i^{\text{condition}}$ was calculated on all (c)TUs for treatment (42 °C, or CDK9 inhibition), and control (37 °C, or solvent) or expressed RefSeq-TUs (exceeding 10 kbp in length with one unique TSS given all RefSeq annotated isoforms (UCSC RefSeq GRCh38)). For each (c)TU or RefSeq-TU i the productive initiation frequency I_i [$\text{cell}^{-1} \text{min}^{-1}$], which corresponds to the pause release rate, was calculated as

$$I_i = \frac{1}{\kappa} \times \frac{tb_i^{\text{condition}}}{t \times L_i} \quad (11)$$

with labeling duration $t = 5$ [min] and length L_i . Note that for RefSeq-TUs, $tb_i^{\text{condition}}$ and L_i were restricted to regions of nonfirst constitutive exons (exonic bases common to all isoforms) located in the first 25 kbp. Given the 15- and 30-min heat shock treatment, we expect only the first 35 kbp to be significantly affected by changes in initiation frequency assuming an average elongation velocity of 2.4 kbp min^{-1} ²⁸. In addition, changes in splicing rate upon heat shock treatment^{77,78} should not influence constitutive exonic regions. This is to ensure, not to be biased by alternative splicing.

Estimation of pause duration d . For all cTUs for 15 and 30 min heat shock treated (42 °C), and control (37 °C) or expressed RefSeq-TUs (exceeding 10 kbp in length with one unique TSS given all RefSeq annotated isoforms (UCSC RefSeq GRCh38)) the pause duration d_i [min] was calculated as the residing time of the Pol in a window ± 100 bases m around the pause site (see above) as

$$d_i = \frac{\sum_{\pm 100} P_{im}}{I_i} \times \text{median}_i \left(\frac{v_i}{\sum_{\text{response window}} P_{im} \cdot v_i(t^* - t)} \right) \quad (12)$$

with pause release rate I_i and the number of polymerases p_{im} (antisense bias corrected mNET-seq coverage values) in a window ± 100 bases around the pause site. For pause sites less than 100 bp downstream of the TSS the first 200 bp of the cTU were considered. Note that the right part of the formula is restricted to mNET-seq instances above the 50% quantile for robustness and adjusts d_i to an absolute scale by comparing the heat shock response derived elongation velocities v_i with those derived from combining mNET-seq and TT-seq data in the response window [200, $v_i(t^* - t)$]. The productive initiation frequency represents the “true” initiation frequency if the fraction of Pol II terminating within the pause window is insignificant (unknown fraction of early termination). Note that the pause duration d obtained in this way reflects the effective pause between two initiation events that

successfully lead to productive elongation of a transcript and thus the relevant transcriptional outcome. Thus, our model is independent of the exact mechanism at the promoter-proximal pause site, may it be pausing or premature termination.

Calculation of response ratios (for calibration of d). For each condition j (control or heat shock 15 min) the antisense bias corrected number of transcribed bases tb_i^j was calculated on all expressed RefSeq-TUs i (exceeding 35 kbp in length with one unique TSS given all RefSeq annotated isoforms (UCSC RefSeq GRCh38)). Response ratios were calculated for a window from the TSS to 10 kbp downstream (excluding the first 200 bp) for each RefSeq-TU i as

$$r_i = 1 - \frac{tb_{i[0,2,10\text{kbp}]}^{\text{heat shock}}}{tb_{i[0,2,10\text{kbp}]}^{\text{control}}} \quad (13)$$

where negative values were set to 0 and values above 1 were set to 1.

Estimation of elongation velocity (for calibration of d). For each condition j (control or heat shock 15 min) the antisense bias corrected number of transcribed bases tb_i^j was calculated on all expressed RefSeq-TUs i (exceeding 35 kbp in length with one unique TSS given all RefSeq annotated isoforms (UCSC RefSeq GRCh38)), excluding the first 200 bp. All TUs were truncated by 5 kbp in length from the 3' end prior to calculation to avoid influence of some alterations in signal around the pA site after heat shock. For each TU i with $r_i > 0.1$ the elongation velocity v_i [kbp min^{-1}] was calculated as

$$v_i = \frac{tb_i^{\text{control}} - tb_i^{\text{heat shock}}}{tb_i^{\text{control}} \times \frac{r_i}{t^*} (t^* - t)} \quad (14)$$

with heat shock treatment duration $t^* = 15$ [min] and labeling duration $t = 5$ [min].

Pause-initiation limit. The previously derived inequality from²⁷

$$\frac{v}{I} \geq 50 \text{ [bp]} \quad (15)$$

states that new initiation events into productive elongation are limited by the velocity of the polymerase in the promoter-proximal region and that steric hindrance occurs at distance of less than 50 bp between the active sites of the initiating Pol II and the paused Pol II²⁸. Given the calculations of pause duration d and (productive) initiation frequency I above, we can reformulate this inequality to

$$\frac{200 \text{ [bp]}}{d \times I} \geq 50 \text{ [bp]} \quad (16)$$

with 200 [bp] being the above defined pause window.

Prediction of RNA secondary structure (MFE). The gene-wise mean minimum free energy (MFE) for a window of [−15, −65] bp upstream of the pause site was calculated from subsequent MFE estimates of 13-bp RNA fragments tiling the respective area using RNAfold from the ViennaRNA package⁷⁹.

HSF1 driven enhancers and promoters. cTUs were classified as HSF1 driven, if a HSF1 binding site was found in a window of 1500 bp upstream to 500 bp downstream of the TSS based on HSF1 binding events (peak calls) that were determined in heat shock conditions of cycling K562 cells (Vihervaara et al.³⁵, data availability: GSE43579).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. The complete set of mNET-seq, TT-seq, and RNA-seq sequencing data and processed files generated for this study was deposited in the GEO database (<http://www.ncbi.nlm.nih.gov/geo>) under accession code GSE123980. A list of all publicly available datasets and the corresponding accession codes used in this study is provided in Supplementary Table 5. The source data underlying Supplementary Fig. 3 (RT-qPCR, cell counts) are provided as a Source Data file.

Code availability

Computational analyses have been performed using R/Bioconductor. Custom scripts can be made available upon request. A detailed description of the bioinformatic workflow used to analyze TT-seq/RNA-seq data has been deposited in the bioRxiv preprint server⁸⁰.

Received: 21 January 2019 Accepted: 18 July 2019

Published online: 09 August 2019

References

- Malik, S. & Roeder, R. G. The metazoan mediator co-activator complex as an integrative hub for transcriptional regulation. *Nat. Rev. Genet.* **11**, 761–772 (2010).
- Coulon, A., Chow, C. C., Singer, R. H. & Larson, D. R. Eukaryotic transcriptional dynamics: from single molecules to cell populations. *Nat. Rev. Genet.* **14**, 572–584 (2013).
- Haberle, V. & Stark, A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat. Rev. Mol. Cell Biol.* **19**, 621–637 (2018).
- Gilmour, D. S. & Lis, J. T. RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in *Drosophila melanogaster* cells. *Mol. Cell Biol.* **6**, 3984–3989 (1986).
- Strobl, L. J. & Eick, D. Hold back of RNA polymerase II at the transcription start site mediates down-regulation of c-myc in vivo. *EMBO J.* **11**, 3307–3314 (1992).
- Bentley, D. L. & Groudine, M. A block to elongation is largely responsible for decreased transcription of c-myc in differentiated HL60 cells. *Nature* **321**, 702–706 (1986).
- Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13**, 720–731 (2012).
- Core, L. J., Waterfall, J. J., Lis, J. T. & Nascent, R. N. A. sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848 (2008).
- Nechaev, S. et al. Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in *Drosophila*. *Science* **327**, 335–338 (2010).
- Wada, T. et al. DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev.* **12**, 343–356 (1998).
- Yamaguchi, Y. et al. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* **97**, 41–51 (1999).
- Marshall, N. F. & Price, D. H. Control of formation of two distinct classes of RNA polymerase II elongation complexes. *Mol. Cell Biol.* **12**, 2078–2090 (1992).
- Marshall, N. F. & Price, D. H. Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *J. Biol. Chem.* **270**, 12335–12338 (1995).
- Vos, S. M. et al. Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. *Nature* **560**, 607–612 (2018).
- Vos, S. M., Farnung, L., Urlaub, H. & Cramer, P. Structure of paused transcription complex Pol II-DSIF-NELF. *Nature* **560**, 601–606 (2018).
- Henriques, T. et al. Widespread transcriptional pausing and elongation control at enhancers. *Genes Dev.* **32**, 26–41 (2018).
- Flynn, R. A., Almada, A. E., Zamudio, J. R. & Sharp, P. A. Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. *Proc. Natl. Acad. Sci. USA* **108**, 10460–10465 (2011).
- Sigova, A. A. et al. Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **110**, 2876–2881 (2013).
- Bunch, H. et al. RNA polymerase II promoter-proximal pausing in mammalian long non-coding genes. *Genomics* **108**, 64–77 (2016).
- Reppas, N. B., Wade, J. T., Church, G. M. & Struhl, K. The transition between transcriptional initiation and elongation in *E. coli* is highly variable and often rate limiting. *Mol. Cell* **24**, 747–757 (2006).
- Muse, G. W. et al. RNA polymerase is poised for activation across the genome. *Nat. Genet.* **39**, 1507–1511 (2007).
- Zeitlinger, J. et al. RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nat. Genet.* **39**, 1512–1516 (2007).
- Barski, A. et al. High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
- Kwak, H., Fuda, N. J., Core, L. J. & Lis, J. T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339**, 950–953 (2013).
- Churchman, L. S. & Weissman, J. S. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**, 368–373 (2011).
- Nojima, T. et al. Mammalian NET-Seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell* **161**, 526–540 (2015).
- Ehrensberger, A. H., Kelly, G. P. & Svejstrup, J. Q. Mechanistic interpretation of promoter-proximal peaks and RNAPII density maps. *Cell* **154**, 713–715 (2013).
- Gressel, S. et al. CDK9-dependent RNA polymerase II pausing controls transcription initiation. *Elife* **6**, <https://doi.org/10.7554/eLife.29736> (2017).
- Schwalb, B. et al. TT-seq maps the human transient transcriptome. *Science* **352**, 1225–1228 (2016).
- Shao, W. & Zeitlinger, J. Paused RNA polymerase II inhibits new transcriptional initiation. *Nat. Genet.* **49**, 1045–1051 (2017).
- Theodorakis, N. G., Zand, D. J., Kotzbauer, P. T., Williams, G. T. & Morimoto, R. I. Hemin-induced transcriptional activation of the HSP70 gene during erythroid maturation in K562 cells is due to a heat shock factor-mediated stress response. *Mol. Cell Biol.* **9**, 3166–3173 (1989).
- Sistonen, L., Sarge, K. D., Phillips, B., Abravaya, K. & Morimoto, R. I. Activation of heat shock factor 2 during hemin-induced differentiation of human erythroleukemia cells. *Mol. Cell Biol.* **12**, 4104–4111 (1992).
- Mathew, A., Mathur, S. K. & Morimoto, R. I. Heat shock response and protein degradation: regulation of HSF2 by the ubiquitin-proteasome pathway. *Mol. Cell Biol.* **18**, 5091–5098 (1998).
- Niskanen, E. A. et al. Global SUMOylation on active chromatin is an acute heat stress response restricting transcription. *Genome Biol.* **16**, 153 (2015).
- Vihervaara, A. et al. Transcriptional response to stress in the dynamic chromatin environment of cycling and mitotic cells. *Proc. Natl. Acad. Sci. USA* **110**, E3388–E3397 (2013).
- Vihervaara, A. et al. Transcriptional response to stress is pre-wired by promoter and enhancer architecture. *Nat. Commun.* **8**, 255 (2017).
- Vihervaara, A., Duarte, F. M. & Lis, J. T. Molecular mechanisms driving transcriptional stress responses. *Nat. Rev. Genet.* **19**, 385–397 (2018).
- Boehma, K., Saunders, A., Werner, J. & Lis, J. T. Transcription factor and polymerase recruitment, modification, and movement on dhsp70 in vivo in the minutes following heat shock. *Mol. Cell Biol.* **23**, 7628–7637 (2003).
- Ni, Z., Schwartz, B. E., Werner, J., Suarez, J. R. & Lis, J. T. Coordination of transcription, RNA processing, and surveillance by P-TEFb kinase on heat shock genes. *Mol. Cell* **13**, 55–65 (2004).
- Chakraborty, A. et al. Classical non-homologous end-joining pathway utilizes nascent RNA for error-free double-strand break repair of transcribed genes. *Nat. Commun.* **7**, 13049 (2016).
- Lis, J. T., Mason, P., Peng, J., Price, D. H. & Werner, J. P-TEFb kinase recruitment and function at heat shock loci. *Genes Dev.* **14**, 792–803 (2000).
- Zacher, B. et al. Accurate promoter and enhancer identification in 127 ENCODE and Roadmap epigenomics cell types and tissues by GenoSTAN. *PLoS ONE* **12**, e0169249 (2017).
- Core, L. J. et al. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat. Genet.* **46**, 1311–1320 (2014).
- Tome, J. M., Tippens, N. D. & Lis, J. T. Single-molecule nascent RNA sequencing identifies regulatory domain architecture at promoters and enhancers. *Nat. Genet.* **50**, 1533–1541 (2018).
- Schlackow, M. et al. Distinctive patterns of transcription and RNA processing for human lincRNAs. *Mol. Cell* **65**, 25–38 (2017).
- Li, B., Weber, J. A., Chen, Y., Greenleaf, A. L. & Gilmour, D. S. Analyses of promoter-proximal pausing by RNA polymerase II on the hsp70 heat shock gene promoter in a *Drosophila* nuclear extract. *Mol. Cell Biol.* **16**, 5433–5443 (1996).
- Li, J., Labbadia, J. & Morimoto, R. I. Rethinking HSF1 in stress, development, and organismal health. *Trends Cell Biol.* **27**, 895–905 (2017).
- Pope, S. D. & Medzhitov, R. Emerging principles of gene expression programs and their regulation. *Mol. Cell* **71**, 389–397 (2018).
- Mahat, D. B., Salamanca, H. H., Duarte, F. M., Danko, C. G. & Lis, J. T. Mammalian heat shock response and mechanisms underlying its genome-wide transcriptional regulation. *Mol. Cell* **62**, 63–78 (2016).
- Dubois, M. F. et al. Heat-shock inactivation of the TFIIH-associated kinase and change in the phosphorylation sites on the C-terminal domain of RNA polymerase II. *Nucleic Acids Res.* **25**, 694–700 (1997).
- Eghazi, E. et al. Heat-shock-specific phosphorylation and transcriptional activity of RNA polymerase II. *Exp. Cell Res.* **242**, 211–221 (1998).
- Schwartz, B. E., Larochelle, S., Suter, B. & Lis, J. T. Cdk7 is required for full activation of *Drosophila* heat shock genes and RNA polymerase II phosphorylation in vivo. *Mol. Cell Biol.* **23**, 6876–6886 (2003).
- Bunch, H. et al. TRIM28 regulates RNA polymerase II promoter-proximal pausing and pause release. *Nat. Struct. Mol. Biol.* **21**, 876–883 (2014).
- Adelman, K. et al. Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIIS. *Mol. Cell* **17**, 103–112 (2005).
- Roginski, R. S., Mohan Raj, B. K., Birditt, B. & Rowen, L. The human GRINL1A gene defines a complex transcription unit, an unusual form of gene organization in eukaryotes. *Genomics* **84**, 265–276 (2004).
- Cheng, B. et al. Functional association of Gdown1 with RNA polymerase II poised on human genes. *Mol. Cell* **45**, 38–50 (2012).
- Dubois, M. F. et al. Heat shock of HeLa cells inactivates a nuclear protein phosphatase specific for dephosphorylation of the C-terminal domain of RNA polymerase II. *Nucleic Acids Res.* **27**, 1338–1344 (1999).
- Strober, W. Trypan blue exclusion test of cell viability. *Curr. Protoc. Immunol.* A.3B.1–A.3B.2. <https://doi.org/10.1002/0471142735.ima03bs21> (2001).
- Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **132**, 365–386 (2000).
- Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-ΔΔC_T} method. *Methods* **25**, 402–408 (2001).
- Hafner, M. et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).

62. Lebedeva, S. et al. Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. *Mol. Cell* **43**, 340–352 (2011).
63. Farazi, T. A. et al. Identification of distinct miRNA target regulation between breast cancer molecular subtypes using AGO2-PAR-CLIP and patient datasets. *Genome Biol.* **15**, R9 (2014).
64. Saskia, G., Katja L. & Patrick C. Transient transcriptome sequencing: experimental protocol to monitor genome-wide RNA synthesis including enhancer transcription. <https://doi.org/10.17504/protocols.io.3vzgn76> (2019).
65. Chomczynski, P. & Mackey, K. Short technical reports. Modification of the TRI reagent procedure for isolation of RNA from polysaccharide- and proteoglycan-rich sources. *Biotechniques* **19**, 942–945 (1995).
66. Dölken, L. et al. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA* **14**, 1959–1972 (2008).
67. Dobin, A. & Gingeras, T. R. Mapping RNA-seq reads with STAR. *Curr. Protoc. Bioinform.* **51**, 11 14 11–19. <https://doi.org/10.1002/0471250953.bi1114s51> (2015).
68. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
69. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
70. Bullard, J. H., Purdom, E., Hansen, K. D. & Dudoit, S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinform.* **11**, 94 (2010).
71. Huber, W., Toedling, J. & Steinmetz, L. M. Transcript mapping with high-density oligonucleotide tiling arrays. *Bioinformatics* **22**, 1963–1970 (2006).
72. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
73. Nojima, T., Gomes, T., Carmo-Fonseca, M. & Proudfoot, N. J. Mammalian NET-seq analysis defines nascent RNA profiles and associated RNA processing genome-wide. *Nat. Protoc.* **11**, 413–428 (2016).
74. Conrad, T. & Ørom, U. A. Cellular fractionation and isolation of chromatin-associated RNA. *Methods Mol. Biol.* **1468**, 1–9 (2017).
75. Albright, L. M. & Slatko, B. E. Denaturing polyacrylamide gel electrophoresis. *Curr. Protoc. Nucleic Acid Chem.* A.3B.1–A.3B.5. <https://doi.org/10.1002/0471142700.nca03bs00> (2001).
76. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
77. Yost, H. J. & Lindquist, S. RNA splicing is interrupted by heat shock and is rescued by heat shock protein synthesis. *Cell* **45**, 185–193 (1986).
78. Shalgi, R., Hurt, J. A., Lindquist, S. & Burge, C. B. Widespread inhibition of posttranscriptional splicing shapes the cellular transcriptome following heat shock. *Cell Rep.* **7**, 1362–1370 (2014).
79. Lorenz, R. et al. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).
80. Gabriel, V., Leonhard, W., Patrick, C., Julien, G. & Björn S. Transient transcriptome sequencing: computational pipeline to quantify genome-wide RNA kinetic parameters and transcriptional enhancer activity. *bioRxiv*. <https://doi.org/10.1101/659912> (2019).
81. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
82. Lopez, M. S., Kliegman, J. I. & Shokat, K. M. The logic and design of analog-sensitive kinases and their small molecule inhibitors. *Methods Enzymol.* **548**, 189–213 (2014).

Acknowledgements

We would like to thank Kerstin Maier and Petra Rus (Max Planck Institute for Biophysical Chemistry) for sequencing. We would like to thank Anna Sawicka and Kristina Žumer (Max Planck Institute for Biophysical Chemistry) for sharing with us the spike-in plasmids. P.C. was supported by the Deutsche Forschungsgemeinschaft (SFB860, SPP1935), the European Research Council Advanced Investigator Grant TRANSRE-GULON (grant agreement no. 693023), and the Volkswagen Foundation.

Author contributions

S.G. conceived and carried out all experiments and contributed to bioinformatics analysis. B.S. designed and carried out all bioinformatics analysis. B.S. and P.C. designed and supervised research. S.G., B.S. and P.C. prepared the manuscript.

Additional information


Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-019-11536-8>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Peer review information: *Nature Communications* thanks John Lis, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019