Nascent RNA sequencing identifies a widespread sigma70-dependent pausing regulated by Gre factors in bacteria

Zhe Sun [1], Alexander V. Yakhnin [1], Peter C. FitzGerald [2], Carl E. McIntosh[2] & Mikhail Kashlev [1✉]

Promoter-proximal pausing regulates eukaryotic gene expression and serves as checkpoints to assemble elongation/splicing machinery. Little is known how broadly this type of pausing regulates transcription in bacteria. We apply nascent elongating transcript sequencing combined with RNase I footprinting for genome-wide analysis of σ^{70}-dependent transcription pauses in *Escherichia coli*. Retention of σ^{70} induces strong backtracked pauses at a 10−20-bp distance from many promoters. The pauses in the 10−15-bp register of the promoter are dictated by the canonical −10 element, 6−7 nt spacer and "YR$_{+1}$Y" motif centered at the transcription start site. The promoters for the pauses in the 16−20-bp register contain an additional −10-like sequence recognized by σ^{70}. Our in vitro analysis reveals that DNA scrunching is involved in these pauses relieved by Gre cleavage factors. The genes coding for transcription factors are enriched in these pauses, suggesting that σ^{70} and Gre proteins regulate transcription in response to changing environmental cues.

[1] RNA Biology Laboratory, National Cancer Institute, National Institutes of Health, Frederick, MD 21702, USA. [2] Genome Analysis Unit, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA. ✉email: kashlevm@mail.nih.gov

Transcription pausing is a fundamental mechanism shared by all three domains of life and is known to regulate gene expression, alternative splicing, co-transcriptional RNA processing, termination, and synchronization of transcription and translation[1–4]. In *E. coli*, a special pausing signal $G_{-10}Y_{-1}G_{+1}$ (Y_{-1} represents the pause site)[5–7] was proposed to slow down RNA polymerase (RNAP) near a translation start site, allowing coordination of RNAP movement with co-transcriptional translation. The elemental pause could be further stabilized by an RNA hairpin formed in the RNA exit channel of RNAP[8,9] or by RNAP backtracking[10,11]. During backtracking, RNAP moves backward along the DNA and the nascent RNA causing the extrusion of the RNA 3′ end into the RNAP secondary channel to induce a long pause or transcription arrest[10,12]. The backtracked pauses can be rescued by removing the extruded 3′ RNA end in a cleavage reaction stimulated by Gre cleavage factors[13–15]. In addition, transcription factors such as RfaH and RpoD (σ^{70}) have been shown to induce transcription pausing by interacting with RNAP and DNA[4].

The housekeeping initiation factor sigma70 (σ^{70}) recognizes the −10 and −35 elements in the promoter regions to form an open promoter complex (RPo) by unwinding the DNA duplex between the −10 element and the transcription start site (TSS)[16]. Normally, escape of RNAP from the promoter causes the release of σ^{70} in a stochastic manner[17]. However, in vivo chromatin immunoprecipitation followed by sequencing (ChIP-seq) and in vitro biochemical data show that σ^{70} is retained in RNAP at a significant distance from the promoter and the efficiency of the retention depends on the transcription unit[18–20]. The −10-like sequence in the initial transcribed region of the pR′ promoter of bacteriophage lambda has been shown to cause the retention of σ^{70} in the RNAP holoenzyme (Eσ^{70}) leading to transcription pausing[21]. In the σ^{70}-dependent pause, the DNA strands in the transcription bubble become scrunched inside RNAP and the strain accumulated during scrunching results in a backtracked σ^{70}-dependent pause state[22–24], which allows proper loading of the accessory antitermination bacteriophage λ Q protein. Elongation factors GreA and GreB release σ^{70}-dependent pauses in vitro[14,25,26] by stimulating the nascent RNA cleavage activity of backtracked RNAP. Although the σ^{70}-dependent pauses have been detected at several *E. coli* and phage promoters[27,28], their robustness, prevalence and their effect on gene expression in vivo remain largely unknown.

Nascent elongating transcript sequencing (NET-seq) has been developed to monitor the genome-wide transcription pausing at single nucleotide resolution in vivo[29]. In this study, we report a modified version of NET-seq combined with RNase I footprinting of the nascent transcripts (RNET-seq) for genome-wide identification of σ^{70}-dependent transcription pauses in *E. coli*. We found that a strikingly large number of *E. coli* genes contain strong σ^{70}-dependent pauses in the 5′ untranslated regions (UTRs) clustered at a distance of 10−20 bp from promoters. We determined the DNA signals essential for these pauses, identified the mechanism of their rescue by Gre factors, and proposed their role in repressing gene expression and in transcriptional responses to changing environmental cues.

Results

RNET-seq identifies σ^{70}-dependent transcription pausing in *E. coli*. In this work, we employed the RNET-seq technique for genome-wide identification of paused ternary elongation complexes (TECs) of RNAP containing the σ^{70} subunit isolated from wild-type (WT) and Δ*greAB E. coli* cells (Fig. 1a). Briefly, transcription-engaged RNAP was released from *E. coli* nucleoids by treatment with DNase I and RNase I followed by

immobilization on Ni^{2+}-NTA agarose beads through His-tag fused to σ^{70} (RpoD) or the β′ (RpoC) subunit (σ^{70} and β′ datasets). Treatment with RNase I degraded all transcripts except for the 3′-proximal fragments strongly protected by RNAP. A substantial fraction of the immobilized complexes was capable of [α-^{32}P] UTP incorporation and susceptible to the RNA cleavage stimulated by GreB (Fig. 1b and ref. [7]), indicating their engagement in active transcription. A strong positive correlation was observed between the biological replicates of RNET-seq when the normalized counts of reads in each gene were compared (Supplementary Fig. 1). The in vitro RNase I footprints of the regular paused TECs containing the σ^{70} subunit consisted of 16−17 nt of the 3′-proximal RNA (Supplementary Fig. 2). Similarly, the in vivo footprints by RNET-seq centered at 17−18 nt and 16−17 nt lengths in β′-WT and σ^{70}-WT datasets, respectively (Fig. 1c). We noted that the σ^{70}-WT data also contained short 6−15-nt RNAs derived from pausing close to promoters. The majority of short 6−11-nt RNAs could not be uniquely mapped to the reference *E. coli* genome and were discarded (Supplementary Fig. 3).

GreA and GreB proteins were previously identified as the major regulators of RNAP pausing and arrest close to promoters[25,30]. The σ^{70} data from Δ*greAB* cells showed a characteristic shift of the RNA length from 16–17 nt to 12–17 nt suggesting that, in the absence of Gre factors, σ^{70}-dependent pauses predominantly occurred in the 12–17-nt registers downstream from TSS (Fig. 1c, black and pink columns). We suggest that reactivation of backtracked pauses by Gre factors cleavage resulted in RNA extension to the 16–17-nt registers in WT cells. The σ^{70}-dependent pauses in Δ*greAB*, but not in WT cells, were also enriched in >17-nt reads suggesting that Gre factors efficiently suppressed RNAP backtracking caused by σ^{70}, and/or rescued the backtracked complexes[14] (Fig. 1c, pink column).

Proximity of σ^{70}-dependent transcription pauses to promoters. The RNET-seq peak representing a typical σ^{70}-dependent pause is shown in Fig. 2a. Only the peaks where the read counts are at least 20-fold over the median value of all RNA reads in a 51-bp window centered at the peaks are assigned as pause sites. In total, we identified 7412 and 3543 pauses recovered by σ^{70}- and β′-affinity pull-down in WT cells (σ^{70}-WT and β′-WT), respectively (Supplementary Data 1). The σ^{70}-WT library had lower background than the β′-WT of RNA reads in the 51-bp window, which resulted in a larger number of the pauses counted in σ^{70}-WT compared to β′-WT cells. This observation indicated that the majority of σ^{70} subunit was bound to RNAP within the narrow promoter-proximal regions of the genome. About 26% of the β′-WT pause sites were shared with the σ^{70}-WT pause sites and the fraction of shared pauses increased to 57% in Δ*greAB* cells (Fig. 2b). The total number of pauses was also 1.6–1.8-fold higher in Δ*greAB* cells: 12211 pauses in σ^{70}-Δ*greAB* cells and 6498 pauses in β′-Δ*greAB* cells (Supplementary Data 1). These data suggested that a substantial fraction of σ^{70}-dependent pauses at these sites was suppressed or released by Gre factors in WT cells. The Δ*greAB* cells had a larger number of pause sites and increased the normalized enrichment in untranslated and antisense regions compared to WT cells (Fig. 2c; Supplementary Fig. 4).

The majority of strong σ^{70}-dependent pauses in WT and Δ*greAB* cells was localized within ~50 bp distance downstream of the annotated TSS (Fig. 2d, top). We arbitrarily separated these pauses into G0, G1, and G2 groups located at −2 to 3, 10 to 20, and 31 to 39 bp distance from the closest TSS, respectively (Fig. 2d, bottom). Although these three groups were similarly populated in σ^{70}-WT cells, the G1 pauses (Supplementary Data 2) dominated in Δ*greAB* cells, suggesting that Gre factors primarily
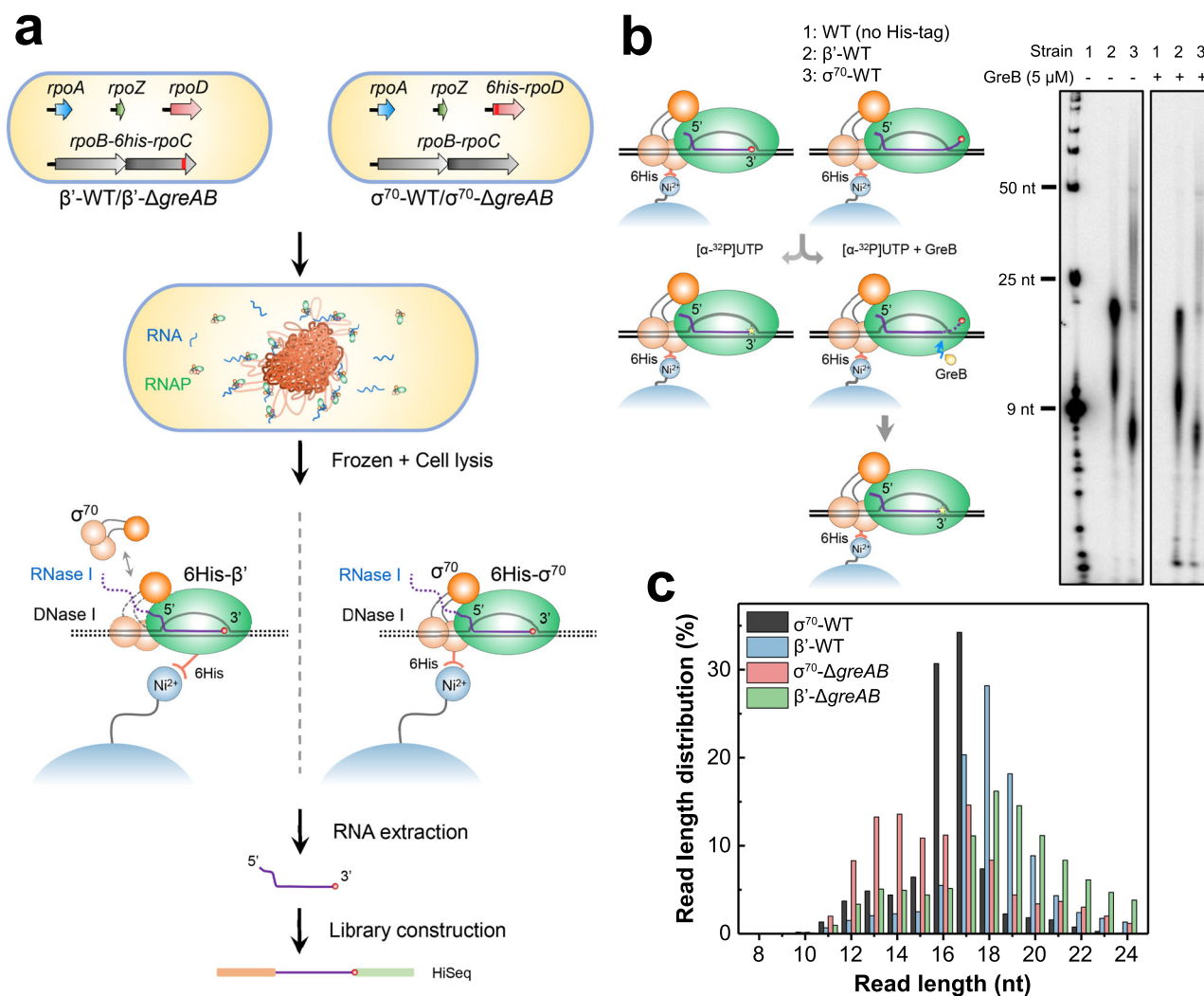
Fig. 1 RNET-seq identifies σ⁷⁰-dependent pauses and the corresponding translocation states of RNAP in WT and ΔgreAB cells. a Principles of RNET-seq. The σ⁷⁰ and β′ strains with 6His-tagged RpoD (σ⁷⁰) and RpoC (β′) subunits were used for purification of the intact paused TECs from bacterial nucleoids after treatment with nucleases (see details in Methods). The green oval represents the RNAP core enzyme. Three connected brown circles represent individual domains of the σ⁷⁰ subunit bound to the −10/−35 promoter elements or to RNAP core. β′-WT/β′-ΔgreAB, WT or ΔgreAB with His-tagged β′; σ⁷⁰-WT/σ⁷⁰-ΔgreAB, WT or ΔgreAB with His-tagged σ⁷⁰. **b** PAGE of the ³²P-RNA-labeled paused complexes and their sensitivity to GreB-stimulated cleavage. WT, *E. coli* W3110 strain lacking the His-tag in σ⁷⁰ and β′. The first lane, RNA ladder. Vertical bar, cleavage products. Similar quantity of paused TECs were used to allow direct comparison of the RNA yields between β′-WT and σ⁷⁰-WT cells. Data shown are representative of three independent experiments. **c** Histogram shows RNA length distributions (RNA footprints) for the uniquely mapped RNET-seq reads from the indicated strains. The length of protected RNA allowed determination of the translocation register of RNAP in β′-WT at each pause: 16−17-nt, 18-nt and >18-nt RNAs corresponded to the post-translocated, pre-translocated and backtracked states, respectively[51]. The average read lengths for σ⁷⁰-WT, β′-WT, σ⁷⁰-ΔgreAB and β′-ΔgreAB strains are 16.3-nt, 18.0-nt, 16.0-nt and 18.3-nt. The <16-nt RNAs were derived from RNAP pausing at a short <16-bp distance from promoters where the nascent transcripts were not yet accessible to the nucleases.

suppressed or released pausing at a short 10−20-bp distance from TSS. Heatmap analysis further revealed that G1 pauses were significantly enriched in σ⁷⁰-ΔgreAB compared to σ⁷⁰-WT cells (Fig. 2e). In contrast, G0 and G2 pauses predominantly observed in σ⁷⁰-WT cells were rarely as strong as G1 pauses (Fig. 2f). Notably, G0 pauses had their 5′ RNA ends residing upstream of the closest TSS, indicating that they originated from promoters located upstream from the nearest TSS. Most G0 and G2 pauses were substantially weaker than the G1 pauses in both, WT and ΔgreAB cells (Supplementary Data 1), and these pauses were not analyzed further.

Two categories of G1 pauses. As reported previously, a promoter-like −10 sequence located downstream from the

original promoter is essential for σ⁷⁰-dependent pausing[21,27,28]. To investigate whether the −10-like region (−10LR) was involved in G1 pauses in σ⁷⁰-ΔgreAB cells, we sorted these pauses based on their distance from the TSS and aligned them via centering at the corresponding TSS. A putative −10LR was identified for the pauses in the 16−20-nt, but not in the 10–15-nt G1 register (Supplementary Fig. 5). Information content (Ri) quantification of −10LR by a σ⁷⁰ model[31] showed an average Ri above 0 for pauses 16–20-nt from the TSS (Fig. 3a). Based on this difference, G1 pauses were subdivided into two categories: proximal G1p (10–15 nt) and distal G1d (16–20 nt). The G1d category showed significantly higher Ri of −10LR compared to all σ⁷⁰ promoters from RegulonDB[32] (Fig. 3b). The significantly shorter read length at G1p pauses compared to G1d and all other peaks indicated a
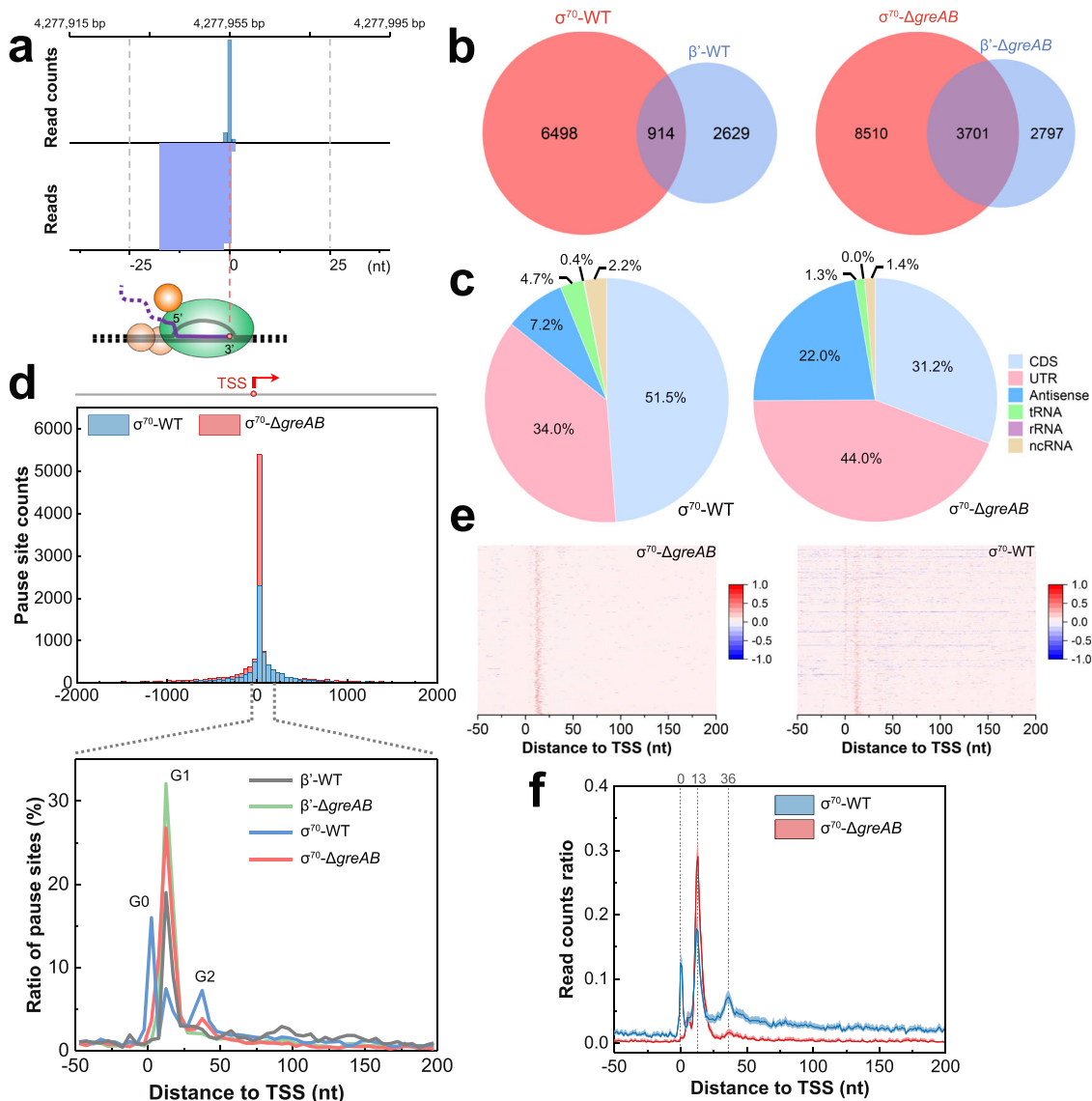
Fig. 2 Classification of σ[70]-dependent transcription pauses. a Example of σ[70]-dependent pause upstream of the *yjcE* coding sequence (CDS) identified by RNET-seq in the σ[70]-Δ*greAB* strain. The genomic coordinates for 3′ ends of all uniquely mapped RNA reads (bottom lane) were determined and the read count for each 3′ end position was calculated and plotted (top lane). The genomic positions where 3′ end/3′ end median (51-bp window) read counts ratio (pause score) was ≥ 20 and read counts/10[6] reads was ≥ 10 satisfied our stringent definition for a pause site. **b** Venn diagrams show the total and shared numbers of pauses identified in σ[70]-WT ($n = 7412$), β′-WT ($n = 3543$), σ[70]-Δ*greAB* ($n = 12211$) and β′-Δ*greAB* ($n = 6498$) strains. **c** Distribution of σ[70]-dependent pauses among CDS, UTR, Antisense, tRNA, rRNA and ncRNA regions in σ[70]-WT and σ[70]-Δ*greAB* strains. The "Antisense" pauses included those in CDS, tRNA, rRNA and ncRNA genes. **d** Distribution of pause sites in promoter-proximal regions. The TSS coordinates identified by dRNA-seq[64] were used to plot pause counts against the pause distance from the nearest TSS on the same DNA strand. The zero and positive coordinates correspond to the pauses overlapping the TSS or located downstream of the TSS, respectively. The upper panel shows the counts of pauses in 50-nt bins within −2000/ +2000-bp window centered at the TSS. The bottom panel shows the ratio obtained by dividing the count of pause sites in a 5-bp sliding window to the total count of pause sites in the −50/+200-bp register surrounding the TSS. Heatmap (**e**) and mean (**f**) of the read counts for σ[70]-Δ*greAB* G1 pause sites ($n = 3099$) in σ[70]-Δ*greAB* (left) and σ[70]-WT (right) strains. The pause sites were ranked based on the pause score (described in **a**). The counts of reads aligned to the sense and antisense strands in each coordinate were normalized to 0 to 1 and 0 to −1 by dividing the maximum read count in each −50/+200-bp region. The regions with multiple pause sites were counted only once (**e**). The dashed line and number on the top indicate the distance of the peak from the TSS. The line and the shadowed region represent the mean and 95% confidence interval for the read counts ratio (**f**).

close proximity of G1p pauses to promoters with their 5′ end residing directly at the TSS (Fig. 3c). Accordingly, the relatively long read length of the nascent RNA at G1d pauses suggested a high fraction of backtracked pausing (Fig. 3d and ref. [7]).

Figure 3e and 3f shows representative G1p and G1d pause sites identified by RNET-seq at *mraZ* (G1p) and *yieE* (G1d) promoters. An in vitro transcription assay confirmed the presence of pauses at the same distance from the TSS as the pauses that

were determined by our RNET-seq. These pauses were not observed in the presence of GreA and GreB proteins, indicating that G1p and G1d pauses included backtracked intermediates that were rescued by Gre factors (Fig. 3g, h, left). Pulling down the ^{32}P-RNA-labeled paused complexes by His-tagged σ[70] or by biotin group in template DNA confirmed the presence of a major fraction of σ[70] subunit in both paused complexes in vitro (Fig. 3g, h, right). The close similarity of the in vitro and in vivo results
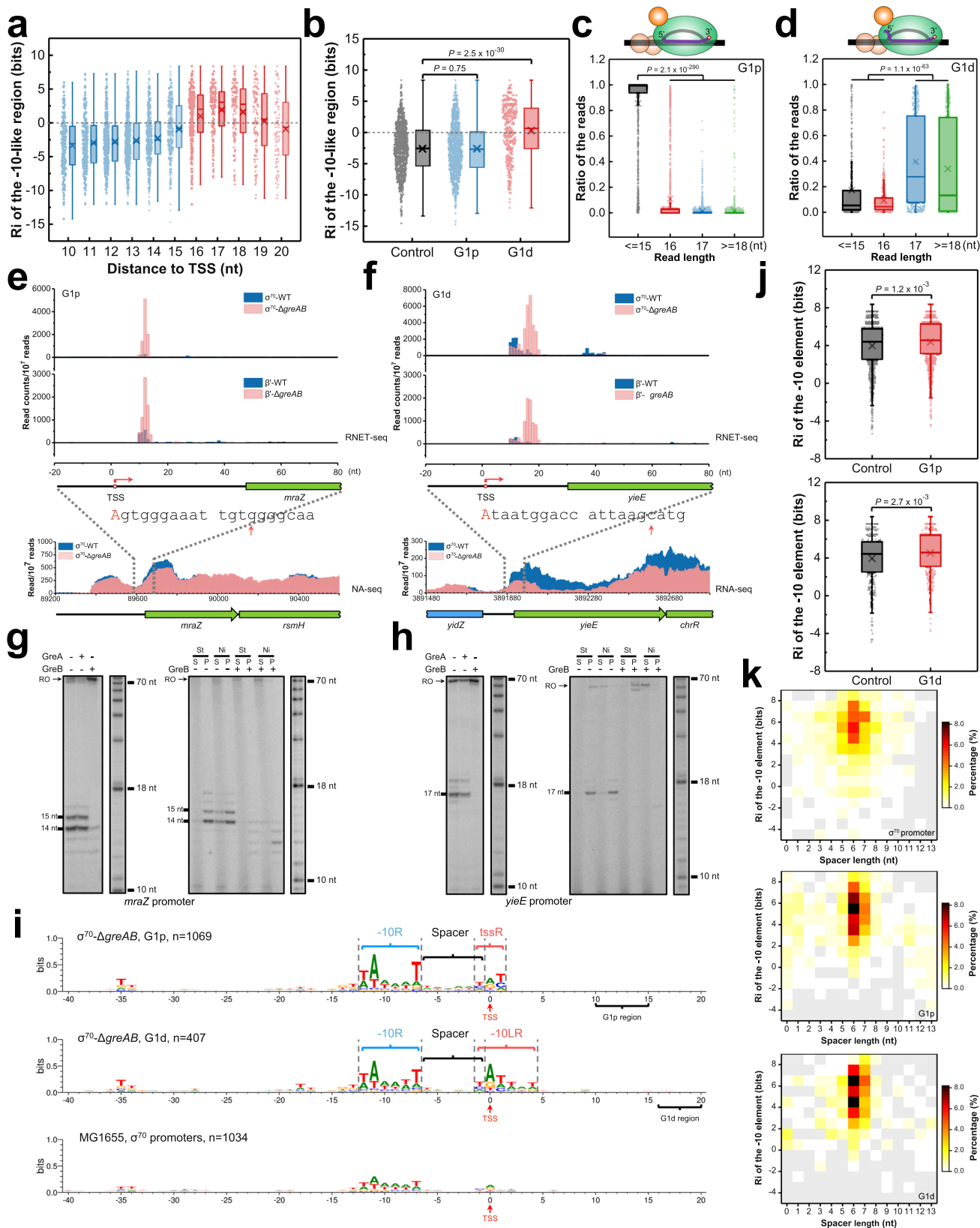
suggested that the σ70 subunit was involved in G1p and G1d pauses in promoter-proximal regions of many *E. coli* genes in vivo. Additionally, RNA-seq confirmed the 1.3- and 2.5-fold higher number of reads in a 200 bp region immediately downstream of the *mraZ* and *yieE* pause sites in WT compared with the Δ*greAB* cells (Fig. 3e, f, bottom). This enrichment indicated

the transcriptional upregulation of the corresponding genes caused by suppression of the G1 pausing by Gre factors in vivo.

An alignment of the G1p and G1d promoter sequences revealed several DNA motifs located immediately upstream from the G1 pauses, which were absent in the reference group of σ70-dependent promoters. The distinct promoter −10 element

Fig. 3 Statistical and in vitro biochemical analysis of G1 pauses. a Information content (Ri) for −10LR (−10-like region) encoded by all σ^{70}-$\Delta greAB$ G1 pauses as a function of its distance from the TSS. The second base in the −10-like hexamer marked the location of the −10LR. The highest Ri of the hexamers ranging from −1 to +2 was adopted and assigned to −10LR ($n = 3099$). **b** Boxplot compares the Ri of −10LR for proximal G1p and distal G1d pauses. All σ^{70} promoters from RegulonDB with a labeled −10 element were used as a control, $n = 950$ σ^{70} promoters. In this and all subsequent boxplots, the median (solid line), mean (cross), 25th and 75th percentiles are indicated, and the whiskers represent 1.5-fold interquartile range. **c, d** Read length distribution at G1p ($n = 1069$) and G1d ($n = 407$) pauses, respectively. Ratio of reads, number of reads with specific length(es)/number of total reads. The cartoon on the top depict the backtracked translocation states of G1d complexes based on a significant difference of their read lengths. Note, that the short ≤15-nt RNAs detected at most G1p pauses were due to the close proximity of G1p pauses to the TSS that precluded determination of the translocation state of G1p complexes by treatment with RNase I. **e, f** RNET-seq and RNA-seq profiles of two representative genomic regions containing G1p and G1d pauses identified by RNET-seq at *mraZ* and *yieE* promoters, respectively. The first 20 nt of *mraZ* and *yieE* transcripts are shown. The red capital letters and arrows indicate the TSS and the pause peaks from RNET-seq data. **g, h** in vitro validation of the σ^{70}-dependent G1p and G1d pauses at *mraZ* and *yieE* promoters, respectively. The left panel shows nascent RNA in the paused complexes obtained in the presence and absence of GreA or GreB. Immobilization on streptavidin beads through 5′-biotin DNA was used to confirm integrity of the RNA-labeled paused complexes (right panel). Eσ^{70} with His-tagged σ^{70} was used for the assay confirming the presence of σ^{70} in the paused complexes. RO, run-off transcripts; St, streptavidin; Ni, Ni^{2+}-NTA agarose; S, supernatant; P, pellet. The representative results are based on three independent experiments. **i** Sequence logo for σ^{70}-$\Delta greAB$ G1p and G1d promoters and for σ^{70} promoters from RegulonDB. The DNA sequences were aligned relative to the TSS. Only the strongest pause was used for analysis of the TSSs following multiple pause sites. Coordinate "0" represents TSS (commonly marked as the +1 site) in the sequence logo, otherwise the standard "+1" TSS nomenclature was used. −10R, −10 promoter element; tssR, region surrounding TSS; −10LR, −10-like region; spacer, spacing region between −10R and TSS. **j** Boxplot comparing Ri of the −10 elements for G1p (top) and G1d (bottom) promoters. G1p promoters, $n = 1069$ and G1d promoters, $n = 407$. −10R of the same numbers of randomly chosen promoters were used as a control. **k** Heatmap showing correlation between distribution of spacer length and information content (Ri) of the promoter −10 element for all σ^{70} promoters (top), promoters containing G1p (middle) and G1d (bottom) pauses. The two-tailed Mann-Whitney U-test was used for the statistical analysis shown above.

(−10R) for both the G1p and G1d promoters (Fig. 3i) indicated a more conserved −10R and/or more conserved 6-bp spacer length between −10R and TSS for the G1 promoters. A significantly higher Ri of the −10R element was observed at promoters located upstream of G1p and G1d pauses (Fig. 3j), as well as for the entire subset of σ^{70} promoters that followed by pause sites identified in this work (Supplementary Fig. 6). A heatmap showed a relatively broad spacer length distribution among all σ^{70}-dependent promoters in *E. coli* (Fig. 3k, top). In contrast, the G1p (63%) and G1d (66%) promoters had more uniform 6−7 nt spacer between the −10R and TSS, indicating that the narrow spacer length might contribute to the strength of G1 pauses (Fig. 3k, middle and bottom). The TSS region (tssR) of G1p promoters, consisted of three nucleotides centered at +1 TSS, was enriched with a "YR$_{+1}$Y" motif with a +1 purine (R) surrounded by two pyrimidines (Y) (Fig. 3i). Interestingly, the same "YR$_{+1}$Y" motif preceded by a 6-nt spacer was previously reported as a strong predictor for genome-wide TSS position and promoter strength[33]. A similar tssR motif was also identified in the reported σ^{70} promoters followed by σ^{70}-dependent pauses (Supplementary Fig. 7). Thus, promoter-proximal σ^{70}-dependent pausing appears to exhibit two distinct mechanisms involving binding of a σ^{70} to strong canonical −10R promoter element, optimal 6/7-bp spacer, and "YR$_{+1}$Y" tssR (G1p promoter), and those containing an additional −10LR sequence at a conserved 11-bp distance downstream from the −10R of the original promoter (G1d promoter). We noticed that the distance between −10R and −10LR sequences approximately corresponded to a single helical turn of B-DNA placing these elements on the same side of the DNA helix. This may facilitate a transition from G1p to G1d pause by the hopping of σ_2 domain, which is a modular domain of σ^{70} to bind −10 element (see Discussion for more details).

−10R, −10LR, tssR elements and spacer contribute to G1p and G1d pauses. Our in vitro testing of several G1p and G1d promoters showed that the pausing patterns and sensitivity to Gre factors closely matched the in vivo results. Briefly, the relative pause strength was largely reduced by GreA or GreB at a representative panel of G1 promoters that we tested in vitro (Supplementary Fig. 8, 9). The effect of GreB on these pauses was stronger than that of GreA (Fig. 4a, b). Several point mutations,

introduced to the G1p and G1 promoters that increased the Ri of their −10 element (−10R Ri+), significantly increased G1p, but not G1d pause strength, indicating that strong binding of σ^{70} to −10R was essential for the G1p pauses (Fig. 4c, d; Supplementary Fig. 10). On the other hand, mutations (−10LR Ri+/−) increasing or decreasing Ri of the distal −10LR of G1d promoters increased or decreased the G1d pause strength, respectively suggesting that the downstream −10LR was involved in G1d pausing (Fig. 4e; Supplementary Fig. 11c, d). The R$_{−1}$-to-Y and R$_{+2}$-to-Y mutations increasing tssR Ri of G1p promoters carrying the suboptimal R$_{−1}$R$_{+1}$Y$_{+2}$ and Y$_{−1}$R$_{+1}$R$_{+2}$ sequence moderately increased the pause strength of a subset of the pauses. Y-to-R mutations at "Y$_{−1}$R$_{+1}$Y$_{+2}$" tssR of G1p promoters that decreased their Ri significantly reduced G1p pausing indicating contribution to pause strength of the pyrimidine residues adjacent to TSS (Fig. 4f; Supplementary Fig. 11a, b). Although not all gain-of-function G1d promoter −10R and G1p promoter tssR mutations improved the pause strength (Fig. 4d, f), the statistical analysis of loss-of-function mutations strongly indicated that −10R and "Y$_{−1}$R$_{+1}$Y$_{+2}$" tssR of G1p promoters, as well as −10R and −10LR of G1d promoters were both essential for G1 pauses.

Finally, we tested the impact of the 6-bp −10R/TSS spacer length on G1 pauses using the *ileX* (G1p) and *focA* (G1d) promoters, both containing the suboptimal 8-bp spacers not typically found in G1 promoters. A 2-bp deletion reducing the *ileX* spacer to 6-bp length caused 8.1-fold increase of the G1p pause strength (Fig. 4g, lanes 1 and 2). Interestingly, a 2.4-fold increase was also observed for the wild-type *ileX* promoter when the regular dinucleotide A$_{+1}$U$_{+2}$ RNA primer corresponding to the native A$_{+1}$ of *ileX* tssR, was replaced with C$_{−1}$A$_{+1}$ primer to induce a 1-bp upstream shift of the TSS, which also shortened the −10R/TSS distance from 8 to 7-bp length (Fig. 4g, lanes 4 and 5; Supplementary Fig. 12). This finding suggested that the 6-bp distance between the 5′ RNA end and −10R, rather than the length of the DNA spacer per se, was crucial for G1p pauses. A similar result was obtained with the G1d *focA* promoter (Fig. 4h), pointing to a similar role of spacer in both types of G1 pauses. Shortening of the *yieE* promoter spacer from 7 to 6 bp moderately increased the pause strength (Supplementary Fig. 13) indicating that a 6-bp spacer length appeared to be optimal for the G1 pauses. Taken together, our mutational analysis confirmed that
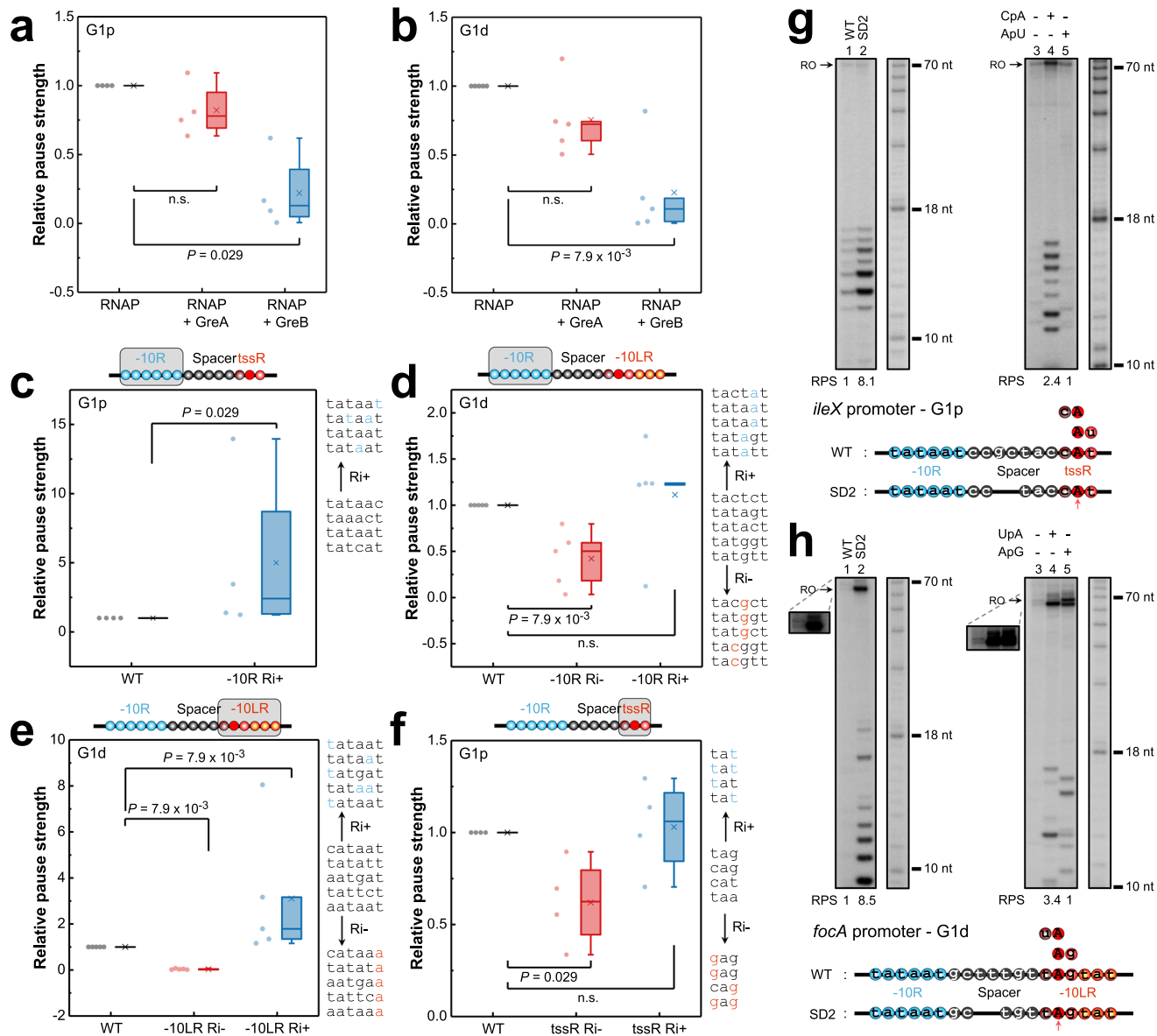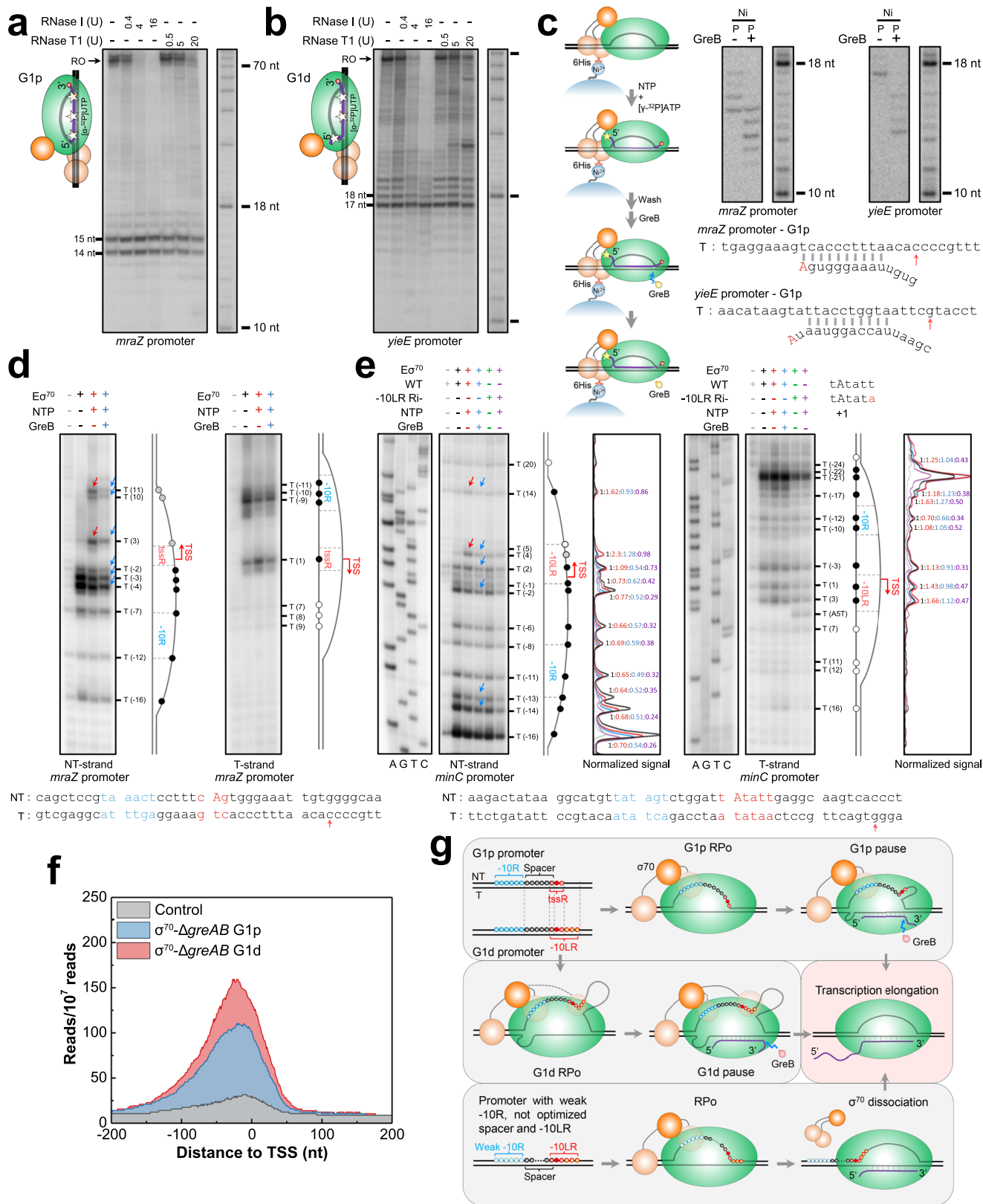
Fig. 4 −10R, spacer length and tssR/−10LR determine G1p and G1d pauses in vitro. Boxplots of pause strength for G1p (**a**) and G1d (**b**) pauses in the absence and presence of GreA or GreB. G1p (*exuR*, *mraZ*, *ileX* and *mocA*; n = 4 G1p promoters) and G1d (*yieE*, *minC*, *gadW*, *mrdB* and *artP*; n = 5 G1d promoters) promoters were used for the analysis (**a–f**). The pause strength was determined by dividing the signal intensity of run-off and paused RNA products to the signal intensity of paused RNA product in the gel for each in vitro template (Pause strength = Signal intensity[paused RNA]/(Signal intensity[paused RNA] + Signal intensity[run-off])). The pause strength in the absence of Gre factors was set to 1 (**a**, **b**). Boxplots show the effect on pause strength of −10R and tssR mutations in G1p promoters (**c**, **f**), and −10R and −10LR mutations in G1d promoters (**d**, **e**). Pause strength of the WT promoters was set to 1 (**c–f**). −10R (−10LR; tssR) Ri−/Ri+, mutated −10R (−10LR; tssR) with decreased or increased Ri are indicated. The gray rectangle in each cartoon represents the motif used for mutation analysis. The original and mutated (colored in blue or red) DNA sequences designed to increase (Ri+) or decrease (Ri−) Ri are shown on the right in gene order. Two-tailed Mann-Whitney *U*-test was used for statistical analysis of the data. Effect of the spacer length on G1p (**g**) and G1d (**h**) pauses. The in vitro transcription was initiated on the WT template or on the mutant template with the shortened DNA spacer (left); different dinucleotide RNA primers overlapping the tssR were employed to alter the position of the TSS (right). The inset shows the run-off transcripts with higher exposure to visualize the faint bands. Data represent three independent experiments. Structural elements of the WT and mutated promoters are shown on the bottom. Each circle represents a single nucleotide. Open blue circles, −10R; Dark red circle, overlapped nucleotide between spacer and tssR/−10LR; Open black and dark red circles, spacer; Red circles, tssR; Red and orange circles, −10LR; Filled red circle, TSS. Red arrows indicate TSS. WT, wild-type promoter; SD2, spacer with 2-nt deletion; RPS, relative pause strength. The analysis included the data from two or more independent experiments.

consensus −10R, 6-bp spacer, and "YR$_{+1}$Y" tssR, all known characteristic for strong *E. coli* promoters, were prerequisites for G1p pausing. In addition, the more distal G1d pauses required the −10LR located 11 bp downstream from the original −10R. This genome-wide result is consistent with the transcription pausing caused by binding of σ^{70} subunit to promoter-proximal

−10-like sequences[21,27,28] that was previously identified at several *E. coli* promoters in vitro.

G1 pauses involve backtracking and an extended transcription bubble. To address the structural foundation of G1 pauses, we

a *mraZ* promoter

b *yieE* promoter

c

mraZ promoter - G1p

T : tgaggaaagtcacccctttaacaccccgtttt

Aguggggaaauugug

yieE promoter - G1p

T : aacataagtattacctggtaattcgtacct

Auaauggaccauuaagc

d

NT-strand *mraZ* promoter

T-strand *mraZ* promoter

NT: cagctccgta aactcctttc Agtgggaaat tgtgggggcaa
T : gtcgaggcat ttgaggaaag tcacccttta acaccccgtt

e

NT-strand *minC* promoter

Normalized signal

T-strand *minC* promoter

Normalized signal

NT: aagactataa ggcatgttat agtctggatt Atattgaggc aagtcaccct
T : ttctgatatt ccgtacaata tcagacctaa tataactccg ttcagtggga

f

g

probed conformational changes in the RNA–DNA hybrid and transcription bubble in TECs at several G1p and G1d pauses identified in vivo and confirmed in vitro. The complete resistance to RNase T1 and RNase I of the nascent RNA at G1p*mraZ* and G1d*yieE* pauses (in UTRs of the *mraZ* and *yieE* genes) consisting of 14−15-nt and 17-nt transcripts, respectively (Fig. 5a, b), and

the high sensitivity of these complexes to GreB-induced transcript cleavage (Fig. 3g, h), indicated the presence of backtracked pauses at both G1 pause sites. Treatment with GreB generated cleavage products shortened by 4−5-nt at the 3′ end, confirming 4−5-bp backtracking of RNAP at the G1p and G1d pauses (Fig. 5c; Supplementary Fig. 14). As reported before, backtracking at

Fig. 5 In vitro analysis of G1p/G1d pauses and the corresponding open promoter complexes. Protection of the nascent RNA by $E\sigma^{70}$ (6His-σ^{70}) holoenzyme from digestion by RNases I and T1 at G1p (**a**) and G1d (**b**) pauses. In the regular (non-paused) elongation complex, RNAP protects 14 nt (RNase T1) and 17−18 nt (RNase I) of the 3′ RNA from nuclease digestion[65]. The cartoons on the left show the proposed alternative translocation states of the RNA in the paused complex. The stars indicate the RNA positions labeled by [α-^{32}P] UMP. Data shown represent two independent experiments. **c** GreB-induced transcript cleavage of nascent RNA at G1p (*mraZ*) and G1d (*yieE*) pauses. The workflow for the experiment is shown on the left. Ni, Ni^{2+}-NTA beads; P, pellet. Results represent two independent experiments. The template strand sequences of the *mraZ* and *yieE* promoters and backtracked RNAs at the pause sites are shown at the bottom. Red arrows indicate the pausing peaks identified by RNET-seq. **d** Permanganate footprints of the non-template and template strands of the transcription bubble at the *mraZ* (G1p) promoter. The positions of all T residues in the bubble are indicated. The diagrams on the right show the transcription bubble at the *mraZ* promoter during G1p pausing. Black filled circles, T residues sensitive to KMnO$_4$ in the absence and presence of NTP; gray filled circles, permanganate-sensitive T residues only in the presence of NTP; white filled circle, T residues resistant to permanganate. **e** Permanganate footprints of the transcription bubble at the *minC* (G1d) promoter. Both DNA strands of the *mraZ* and *minC* promoters including the −10R (blue), tssR/−10LR (red) elements and TSS (red capital) are shown at the bottom, and the corresponding G1p and G1d pause sites are marked by red arrows. **f** Profiles of median ChIP-seq reads coverage at G1p, G1d and control promoters based on the heatmaps (Supplementary Fig. 16). Permanganate footprinting results are representative of three independent experiments. **g** Model depicting the structural properties of σ^{70}-dependent G1p and G1d pauses. The interaction of σ^{70} domains with the promoter elements, the DNA scrunching and the corresponding changes in the RNA register at G1p and G1d pauses are indicated.

≥3-bp distance increases sensitivity to GreB and makes these pauses more resistant to GreA[13]. In contrast, backtracking at 1–2-bp distance makes these pauses more susceptible to GreA[34]. The substantially lower sensitivity of G1p and G1d pauses to GreA compared to GreB (Fig. 4a, b) confirmed backtracking of more than 2 bp at the G1 pauses.

Potassium permanganate footprinting, which probes unpaired T residues in DNA, showed relatively normal size (~17-nt) and location of the transcription bubble in the RNAP-promoter open complex (RPo) at the *mraZ* promoter, which codes for a G1p pause (Fig. 5d, −NTP, black lane). In contrast, the TEC at the G1p^{mraZ} pause showed an unusually long ~27-nt bubble, substantially larger than the bubble detected in the regular TEC carrying the similar length of nascent RNA that was obtained from T7A1 promoter containing no G1 pause (Fig. 5d, +NTP, red lane; Fig. 5g; Supplementary Fig. 15). Strikingly, the corresponding paused TEC at the G1d^{minC} promoter exhibited an even larger (>30-nt) transcription bubble compared to the size detected at G1p pause sites (Fig. 5e, red arrows; Fig. 5g). A point mutation introduced to the −10LR of *minC* eliminated the G1d^{minC} pause also reduced the size of the bubble to the scale typically observed in the RPo (Fig. 5e, purple lane). Thus, the extended bubble appeared to be a hallmark of the G1d promoters making them different from the regular and the G1p promoters. Cleavage of the nascent RNA at the G1p^{mraZ} and G1d^{minC} sites by GreB rescued these pauses (Fig. 3g; Supplementary Fig. 9b). However, treatment with GreB reduced, but did not completely eliminate the bubble at these pause sites and promoter region (Fig. 5d, e, blue arrows), suggesting that these promoters contained a large fraction of RNAP capable of forming a RPo-like promoter complex, which was trapped in the catalytically inactive state[30,35]. Our analysis of the published ChIP-seq/σ^{70} data[36] confirmed a high enrichment of RNAP holoenzyme in a 400-nt window centered at the TSS of the G1p and G1d promoters compared to the promoters lacking G1 pauses (Fig. 5f; Supplementary Fig. 16).

σ^{70}-induced pausing controls the expression of regulator genes. Although regulation of σ^{70}-dependent pauses by Gre factors has been well documented in vitro[14,27,28], their biological role and impact on genome-wide transcription levels warranted further investigation. Our data showed that G1 pauses were significantly increased in cells lacking Gre factors. ~70% of all G1 peaks from RNAP (G1p, 1128/(1128 + 424); G1d, 366/(366 + 158)) identified in ΔgreAB cells had the matching strong σ^{70} peaks that accumulated at G1 pauses (Fig. 6a). Not all pausing peaks identified by σ^{70}-affinity were also identified by β′-affinity. A

substantial fraction of σ^{70} may have dissociated from Eσ^{70} during promoter escape or was lost during purification of the complexes by β′-affinity. The RNA-seq also showed that genes containing G1 pauses were expressed at a significantly higher level compared to the randomly selected genes (Fig. 6b), which was consistent with the canonical −10 element of the strong G1 promoters (Fig. 3i). Gene ontology (GO) analysis[37] showed that *E. coli* genes containing the G1 pauses were enriched among genes coding for the general and gene-specific transcription regulators (Fig. 6c). Most importantly, our RNA-seq analysis of transcription levels in σ^{70}-WT and σ^{70}-ΔgreAB cells revealed that genes harboring G1 pauses were consistently downregulated in the σ^{70}-ΔgreAB compared to σ^{70}-WT cells, and this downregulation was especially pronounced in genes containing the strong G1 pauses (Fig. 6d). Our analysis of the published RNA-seq data revealed that the transcription of the *greA* and *greB* genes were regulated in an opposite manner under each stress condition, causing induction of one but repression of the other *gre* gene (Supplementary Fig. 17). In turn, the G1 pauses are released by either GreA or GreB depending on the type of stress and backtracking distance of the corresponding pause. Thus, our results provide strong evidence that the highly dynamic G1 pauses with the rapidly exchanging backtracked states are involved in a global regulation of promoter escape and in the local transcriptional networks governed by specialized transcription regulators (Fig. 6e).

Discussion

Promoter-proximal pausing is broadly employed for regulation of genes in metazoans[1,2,38]. However, only a limited number of bacterial and bacteriophage promoters have been shown to be regulated by promoter-proximal pausing in vitro, and the protein factors involved in vivo remain unknown. Here, by using σ^{70} subunit-based RNET-seq with a single nucleotide resolution, we identified the genome-wide pause sites caused by the σ^{70} subunit in *E. coli*. Further analysis showed that these pause sites are largely enriched in the promoter regions and regulated by Gre factors. We characterized two distinct mechanisms for promoter-proximal pausing that act in vivo consecutively at 10–15-bp (G1p) and 16–20-bp (G1d) distances from the TSS. The G1p promoters consist of the canonical −10R, a 6–7 nt −10R to TSS spacer, and a "YR$_{+1}$Y" tssR. All of these features were previously shown to determine the high strength of *E. coli* promoters[33,39–42]. Although the strong σ^{70} binding seemed to facilitate rapid and stable recruitment of RNAP in vivo, it also hindered promoter escape due to the strong anchoring of σ^{70} to the canonical promoter elements, ultimately leading to DNA scrunching and
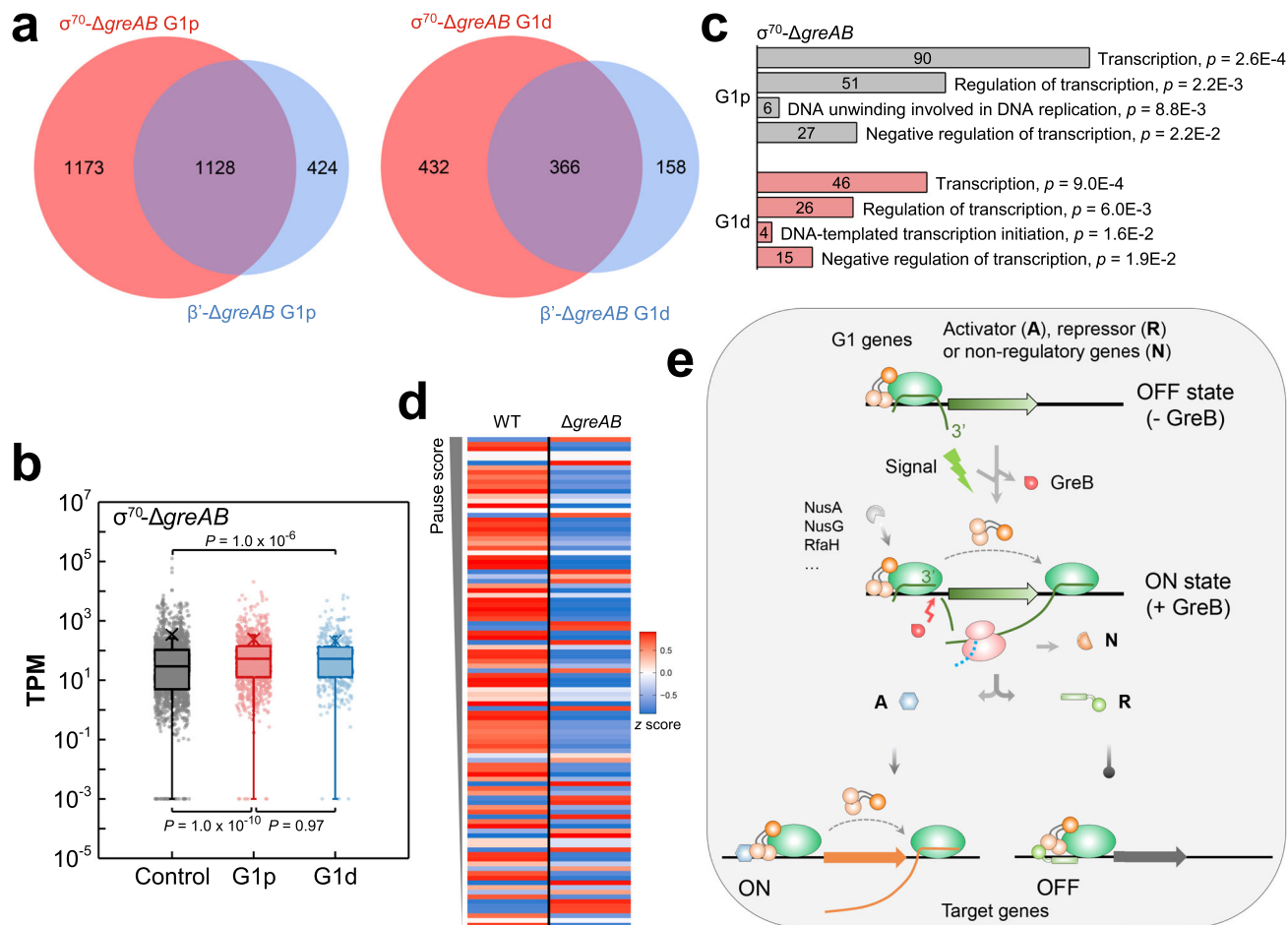
Fig. 6 Role of σ70-dependent pauses in transcription regulation. a Venn diagrams of all G1p (left) and G1d (right) pauses from ΔgreAB cells identified in this work. **b** Boxplot of transcripts per million (TPM) for the genes with and without G1p or G1d pauses. Control genes, $n = 1641$; genes with G1p pauses, $n = 863$ and genes with G1d pauses, $n = 386$. P value was calculated by two-tailed Mann-Whitney U-test. **c** Gene Ontology (GO) analysis of genes with G1p and G1d pauses in the corresponding promoter-proximal regions. All significantly enriched gene categories are listed. The number of genes in each category is shown inside the bars. **d** Heatmap shows the transcription pattern of genes containing G1 pauses in σ70-WT and σ70-ΔgreAB datasets. The G1 genes whose start codon is ≤ 20 bp upstream and ≤ 100 bp downstream from σ70-ΔgreAB G1 pause sites (pause score ≥ 1000, $n = 104$) are shown. Data from three biological replicates are presented. WT, σ70-WT strain; ΔgreAB, σ70-ΔgreAB strain. **e** Schematic illustration of the mechanism for σ70-induced promoter-proximal pausing, its suppression by GreB, and the impact of the pausing on global regulation of transcription. σ70 induces strong backtracked G1 pauses to inhibit gene transcription by hindering RNAP elongation. The GreB expression is increased under certain environmental stresses to relieve the G1 pauses. The release of G1 pauses increases the corresponding genes transcription, especially the genes coding for transcription regulators. The transcription regulators further up- or down-regulate transcription of target genes to response to environmental perturbations.

RNAP inactivation by backtracking as depicted in Fig. 5g. Promoter clearance is generally considered as a hallmark of transition from transcription initiation to elongation. Transcription initiation is a multistep process in which formation of the open promoter complex is following by multiple cycles of abortive transcription, generating short 2–15-nt RNAs that are rapidly released from RNAP[43]. The transition to productive elongation occurs at a ~9–11-nt distance from promoters[44]. The G1p paused complexes primarily occur in the 10–15-nt register of promoters rather than the complexes engaged in abortive transcription. These pauses, but not the abortive complexes were efficiently rescued by GreB to increase the run-off products (Supplementary Fig. 18). These data indicate that the σ70-dependent G1p pausing derives from the early elongation complexes that remain bound to promoters.

The G1d pauses shared a similar promoter-like structure as the G1p pauses but contained an additional −10LR element that causes retention of σ70 after RNAP escapes from the promoter. This mechanism is reminiscent of the pauses reported in vitro at the λpR' and lac promoters[21,27,28], which also have a canonical −10 region, a 6-nt spacer between −10 region and TSS and a −10-like region, to which σ70 binds to induce RNAP pausing at a 16−17-bp distance from the corresponding TSS. These similarities strongly indicate that the σ70-dependent pauses identified in vitro at the λ bacteriophage pR' promoter have the same mechanism as the σ70-dependent G1d pauses described in our work. We noticed that a large number of G1 promoters contained both G1p and G1d pauses, suggesting that the G1p promoters may increase the local concentration of σ70 near the promoter DNA to facilitate hopping of the σ2 domain from the original −10 element to the secondary −10LR sequence located nearby. Indeed, the −10LR of G1d promoters appeared to be positioned at ~11 bp downstream from the original −10R on the same face of the DNA helix, which may further promote σ2 domain hopping to generate similar contacts with DNA in RPo and G1d-paused TEC as reported at λpR' promoter[45]. In this model, the other domains of the holoenzyme may remain bound to the original promoter elements. The dsDNA constrained by the σ2 domain may help maintain the interaction between σ4 and the −35 promoter element to further stabilize the paused elongation

complex. Although elemental pauses have been reported as precursors for the longer hairpin-dependent and backtracked pauses in *E. coli*[4,8], we did not identify the putative elemental pause motifs near the G1 pauses, indicating that they seem not to be essential for σ[70]-dependent pausing.

The metadata analyses (Fig. 5f; Supplementary Fig. 16) showed that the G1 promoters effectively recruit RNAP, but strongly limit its escape to productive elongation. Holding RNAP at the promoter should block access of other RNAP molecules to the corresponding gene[35,46], thus, turning RNAP itself into a general transcription repressor. The G1 pausing may represent yet another layer of gene repression in addition to the well-known mechanisms of promoter occlusion by repressors that block open complex formation[39]. In addition, the G1 pauses may expedite a transcriptional response to changing environmental cues after being released by Gre factors. Indeed, transcription of the *greB* gene appeared to be induced under different stress conditions (Supplementary Fig. 17) supporting its crucial role in stress responses. This mechanism seems to be similar to the robust promoter-proximal pausing of RNA polymerase II and its rescue by TFIIS for rapid response to external signals in eukaryotes[1,2,47]. We found that G1 pauses are enriched in the genes coding for transcription regulators (Fig. 6c), ultimately establishing the G1 pause-containing genes as key nodes involved in regulation of cellular responses to environmental perturbations.

Binding sites for σ[70] on RNAP core overlap with those for the general Nus factors (NusA, NusG and RfaH) known to synchronize transcription and translation, control pausing during elongation and processivity of RNAP[48–51] (see Fig. 6e). The G1 pauses may serve as a checkpoint enabling a temporal assembly of these factors at the promoter to guarantee the subsequent proper readout and regulation by the downstream elongation and termination signals. This notion is consistent with a negative correlation between the binding pattern of σ[70] and the binding patterns of NusA and NusG observed by ChIP-seq analysis of promoter-proximal regions[52]. The G1 pausing may also stabilize binding of σ[70] to RNAP and make transcription of the target genes, such as ncRNA and antisense RNA genes, independent of regulation by Nus and Rho factors[18,53]. In addition, the G1 pausing could increase retention of σ[70] in the elongation complexes at a large distance from the original promoter to additionally reinforce pausing caused by σ[70]-mediated recognition of the −10-like sequences[18].

Further analysis is required to investigate the role of the robust σ[70]-dependent pausing in transcription elongation at a large distance from promoters including transcription terminators[54]. The high evolutionary conservation of σ[70] suggests that this pausing mechanism is likely shared by other bacteria. RNA polymerase II initiation factors TFIIB and TFIIE[55–57], possessing homology with bacterial σ factors, are the likely candidates to regulate promoter-proximal pausing in eukaryotes.

Methods

Bacterial strains and growth conditions. *E. coli* strains β′-WT (W3110 *rpoC*-6×His::*kan*) and β′-Δ*greAB* (W3110 *rpoC*-6×His::*kan greA*::*tet greB*::*amp*) were engineered as was previously described[7]. σ[70]-WT (W3110 6×His-*rpoD*) strain was constructed using a CRISPR-Cas9 system. For the His-tagging, a homologous recombination DNA with His-tag sequence (5′-catcaccatcaccatcac-3′) was inserted 3′ of the G residue of the start codon (AT**G**) of *rpoD* and the ~1.0 kb surrounding DNA was amplified by overlap PCR and cloned into plasmid pTargeT. After electroporation, the tagged strain was identified by PCR and confirmed by Sanger sequencing. The *greA* and *greB* genes were disrupted by P1 transduction from strain β′-Δ*greAB* to obtain the σ[70]-Δ*greAB* (W3110 6×His::*rpoD greA*::*tet greB*::*amp*) strain. The primers used are shown in Supplementary Data 3. All *E. coli* strains were grown in LB medium (tryptone 10 g l[−1], yeast extract 5 g l[−1], NaCl 10 g l[−1]) or on LB plate containing 50 µg ml[−1] kanamycin, 40 µg ml[−1] spectinomycin, 50 µg ml[−1] ampicillin or 12.5 µg ml[−1] tetracycline when appropriate.

RNET-seq and data analysis. Cell collection, lysis and elongation complexes pulldown. An overnight cell culture was diluted in 100 ml LB medium (OD$_{600}$ = 0.02) and cultured at 37 °C to reach a mid-log phase (OD$_{600}$ = 0.5). To stabilize binding of σ[70] to RNAP core during TEC purification, low ionic strength conditions (described below) were used throughout the purification protocol. Namely, the cell culture was combined with an equal volume of frozen 2 × crush buffer (20 mM Tris-HCl pH 7.8, 10 mM ethylenediaminetetraacetic acid (EDTA), 100 mM NaCl, 1 M Urea, 25 mM NaN$_3$, 2 mM β-mercaptoethanol, 10% ethanol, 0.4% NP40, 1 mM PMSF) and the cells were collected by centrifugation (18000 g, 15 min, 4 °C), instantly frozen in liquid nitrogen and placed on ice. The cells were resuspended and lysed by 120 kU Ready-Lyse lysozyme (Lucigen), 400 U RNase I (Invitrogen) and 40 U alkaline phosphatase (NEB) at room temperature for 10 min. The chromosomal DNA was pelleted and treated with 300 U RNase I, 6 U Turbo DNase (Invitrogen) and 100 U DNase I (Roche) by vortexing at room temperature for 10 min. After centrifugation (18,000 g, 3 min, 4 °C), the supernatant containing the solubilized TECs (~700 µl) was incubated with 200 µl of Ni^{2+}-NTA beads for 1 h at 4 °C with continuous shaking (1000 rpm). The beads were washed 4 times with 1 ml of the wash buffer (20 mM Tris-HCl pH 7.8, 1 M betaine, 5% glycerol, 2 mM β-mercaptoethanol, 2.5 mM imidazole) and 3 times by 1 ml pre-elution buffer (20 mM Tris-HCl pH 7.8, 40 mM KCl, 0.3 mM MgCl, 5% glycerol, 2 mM β-mercaptoethanol, 2.5 mM imidazole). The TECs immobilized on the beads were digested once again with 100 U RNase I, 2 U Turbo DNase and 40 U DNase I in 150 µl pre-elution buffer containing 200 µg ml[−1] bovine serum albumin for 30 min at room temperature with continuous shaking (600 rpm). The beads were washed 4 times with the wash buffer and loaded onto 0.5 ml Ultrafree-MC centrifugal filters (Millipore). The immobilized material was eluted with the wash buffer containing 0.3 M imidazole. The nucleic acids in the eluates were extracted once with 400 µl phenol:chloroform:isoamylalcohol (PCI; 25:24:1) and once with 300 µl chloroform. The top water phase was collected and mixed with 3 volumes (~1200 µl) of isopropanol. After precipitation at −80 °C for 30 min and centrifugation, the nucleic acids pellet was washed by 180 µl of 80% ethanol and air-dried. The pellet was dissolved in 12 µl nuclease-free water. The DNA was removed by 2 U Turbo DNase and 10 U DNase I at 37 °C for 15 min. The residual RNA was extracted by PCI, precipitated by isopropanol and solubilized in 10 µl nuclease-free water.

Barcode ligation and reverse transcription. The RNA was ligated to 10.7 pmol barcode DNA linker using 200 U T4 RNA ligase 2 (NEB) overnight at 16 °C. The ligation product was extracted by chloroform, precipitated by isopropanol and solubilized in 10 µl nuclease-free water. Reverse transcription was performed using the RNA-DNA chimera and 3 µM phosphorylated reverse transcription primer in 1 × PrimeScript buffer containing 0.5 mM dNTPs, 5 mM DTT, 0.6 U µl[−1] SuperaseIn RNase inhibitor (Invitrogen) and 10 U µl[−1] PrimeScript Reverse Transcriptase (Takara) at 48 °C for 30 min. After 2 U RNase H (NEB) treatment for 15 min at 37 °C, the reaction mixture was separated by 10% Urea-TBE PAGE. The cDNA products at 75−100-bp range were excised from the gel and extracted with nuclease-free water for 10 min at 70 °C. The gel chunks were removed by filtering and the cDNA was precipitated by 3 volumes of isopropanol at −80 °C for 30 min and dissolved in 4 µl nuclease-free water.

Circularization, library preparation and Illumina sequencing. The resulting cDNA was circularized by 40 U ssDNA ligase (Lucigen) at 60 °C for 4 h. The circularized DNA was subjected to PCR to generate a sequencing library using Illumina index primers and PrimeSTAR Max DNA polymerase (Takara). The PCR product was loaded and electrophoresed by 8% TBE PAGE. The DNA product excised from the gel was extracted overnight by 680 µl DNA soaking buffer (0.3 M NaCl, 10 mM of Tris-HCl pH 8.0, 0.97 mM EDTA) at room temperature. The DNA library was precipitated by isopropanol, washed once by cold 80% ethanol, air dried and dissolved in 8 µl 10 mM Tris-HCl pH 8.0. The concentration of the library was determined by an Agilent 2100 bioanalyzer. Illumina sequencing was performed by the NIH Intramural Sequencing Center. The DNA libraries were quantified by qPCR, pooled and loaded on an Illumina HiSeq 2500 using 2 × 50 bp paired-end sequencing in rapid run mode.

Data analysis. After a quality check, the primer sequence was trimmed from the raw R1 reads by Cutadapt[58] and PCR duplicates were removed using Clumpify from the BBMap suite based on the random barcode. The random barcode was further removed and the reads were aligned to the *E. coli* genome NC_000913.2 using Bowtie[59]. After disregarding the multi-mapped reads, some strong false-positive pausing peaks appeared in tRNA and rRNA genes, and other repetitive sequences in the genome. These peaks were excluded from the further analysis. The 5′ end coordinates of all uniquely aligned R1 reads, which correspond to 3′ end of RNA, were recorded by BEDTools[60] and the total read counts at each coordinate were determined. The coordinate was picked up and defined as a transcription pause site when its read counts was at least 20-fold of the median read counts in a surrounding 51-nt window size and not less than 10 per million reads.

DNA templates and in vitro transcription. The wild-type promoters from the −80 to +60 region relative to the TSS used for in vitro transcription, were amplified by PCR using genome as template and cloned into T-Vector pMD19 (Simple, Takara). Primers containing the mutations were used to PCR the whole derived pMD19 plasmid constructed above. The DNA product was self-ligated using T4 DNA ligase (Invitrogen) and transformed to DH5α competent cells. Mutations were confirmed by Sanger sequencing and the plasmid was used to

amplify DNA template for in vitro transcription. When appropriate, a 5′-biotin-labeled primer was used to amplify DNA template with biotin labeling at the 5′ end of non-template strand. Primers used to amplify DNA templates are listed in Supplementary Data 3. Single round in vitro transcription reactions were performed in transcription buffer (40 mM Tris-HCl pH 8.0, 1 mM dithiothreitol, 0.1 mg ml^{-1} BSA, 10 mM MgCl$_2$, 50 mM KCl) in two steps. First, 20 nM linear DNA template and 50 nM Eσ^{70} were mixed and incubated at 37 °C for 10 min to form the open complex. When indicated, 200 nM GreA or 50 nM GreB was added in this step. Then 20 μM GTP, UTP, CTP, 2 μM ATP and 5 μCi [γ-^{32}P] ATP (PerkinElmer) were used to start the reaction for 10 min. In the second step, the reaction mixture was chased with the addition of 20 μM ATP and 10 μg ml^{-1} rifampicin for 3 min. The reaction was terminated by adding the same volume of 2 × stop buffer (10 M Urea, 250 mM EDTA pH 8.0, 0.05% xylene cyanol and bromphenol blue) and analyzed on 23% (10:1, acrylamide:bisacrylamide) polyacrylamide gel with 7 M urea. All procedures of in vitro transcription, the following RNase I, RNase T1 cleavage and GreB stimulated cleavage assays were performed at 37 °C unless indicated otherwise.

For testing RNAP activity before RNET-seq, 10 μl Ni^{2+}-NTA beads with ECs were washed three times by 200 μl pre-elution buffer. Then 10 mM MgCl$_2$ and 10 μCi [α-^{32}P] UTP (PerkinElmer) were added to the beads to elongate the nascent RNAs for 10 min. After washing the beads three times by wash buffer, the ECs were eluted by 10 μl wash buffer containing 0.3 M imidazole. For pull-down experiments, 5′-end biotin labeled DNA template and reconstituted Eσ^{70} (6His-σ^{70}) were used. The same in vitro transcription was done as mentioned above on 8 μl streptavidin and Ni^{2+}-NTA beads. After reaction and spinning down the beads, the top solution was collected ("supernatant" fraction) and the bottom beads were immediately washed three times to stop the reaction. The transcription products were released by heating the beads resuspended by the same volume of stop buffer at 95 °C for 5 min ("pellet" fraction). To initiate transcription by dinucleotide, 200 μM CpA, ApU, UpA, or ApG (TriLink) were added during open complex formation. Then 20 μM NTPs, 2 μCi [α-^{32}P] UTP and 10 μg ml^{-1} rifampicin were added and incubated for 3 min before stopping the reaction.

RNase I and RNase T1 footprinting of the nascent RNA. In an 8 μl reaction mixture, 20 nM DNA template and 50 nM reconstituted Eσ^{70} (6His-σ^{70} or 6His-β′) was incubated for 10 min on 8 μl Ni^{2+}-NTA beads. Then 20 μM GTP, ATP, 2 μM UTP and 3 μCi [α-^{32}P] UTP were added to initiate the reaction at the rrnB P1 promoter for 10 min. An additional 20 μM CTP was used for the mraZ and yieE promoters. After chasing the reaction by 20 μM UTP and 10 μg ml^{-1} rifampicin for 3 min, the beads were washed twice and treated by the indicated amount of RNase I (Invitrogen) or RNase T1 (ThermoFisher) for 10 min at 24 °C. The beads were washed two times and extracted with 3 μl PCI to terminate the reaction.

GreB cleavage assay. The same reaction on Ni^{2+}-NTA beads that was used for the RNase footprinting was pre-incubated to form RPo. Transcription was initiated by adding 20 μM GTP, UTP, CTP, 2 μM ATP and 5 μCi [γ-^{32}P] ATP for 10 min. The reaction was chased with 20 μM ATP for 3 min. After washing two times, 50 nM GreB was added for 10 min to induce cleavage of the transcripts. The beads were washed twice to stop the reaction and the products were denatured at 95 °C for 5 min.

Potassium permanganate DNA footprinting. DNA was labeled by [γ-^{32}P] ATP individually at the 5′ end of the template or the non-template strands. The labeled DNA (~12,000 cpm) and 150 nM Eσ^{70} were used to form the paused TECs. The sample was mixed with equal volume of 20 mM KMnO$_4$ by vortexing for 15 s and quenched by 1.3 M β-mercaptoethanol. After adding 80 μg salmon sperm DNA (Invitrogen) and nuclease-free water to a total volume of 100 μl, the DNA fragments were extracted by PCI and precipitated by adding 1/10 volume of sodium acetate and 2.5 volumes of ethanol for 1 h at −20 °C. The pellet was resuspended in 10% (v/v) piperidine and treated for 15 min at 90 °C. The DNA fragments were re-precipitated and washed twice with 70% ethanol. The DNA pellet was dissolved in 20 μl nuclease-free water, dried by vacuuming and dissolved in the loading buffer (95% formamide, 20 mM EDTA pH 8.0, 0.2% SDS, 0.05% xylene cyanol and bromphenol blue). The sequencing ladders were generated by a Thermo Sequenase Cycle Sequencing Kit (ThermoFisher). The resultant DNA products were analyzed by 10% (19:1, acrylamide:bisacrylamide) PAGE containing 7.5 M urea.

RNA-seq and data analysis. To extract total RNA for RNA-seq, 8 ml E. coli cells grown to mid-log phase (OD$_{600}$ = 0.5) were spun down, resuspended in 800 μl TRIzol (Invitrogen) and incubated for 4 min at 95 °C. The total RNA was purified by 400 μl PCI extraction and 200 μl chloroform extraction. After centrifugation, an equal volume of isopropanol was added to the top water phase to precipitate the RNA. The genomic DNA was digested with 50 U DNase I for 30 min at room temperature. The RNA was purified by RNeasy Mini Kit (Qiagen) and its concentration was quantified by Agilent 2100 bioanalyzer. The libraries were constructed using TruSeq Stranded Total RNA Library Prep Kit (Illumina) and applied to MiSeq using 2 × 150 bp paired-end sequencing at the Center for Cancer Research Sequencing Facility. The reads that passed quality control and filtering of the raw data were aligned to the E. coli genome NC_000913.2 using STAR[61]. The raw counts of the aligned reads for each gene were

calculated by HTseq[62]. Fold changes of genes transcription between different samples were calculated by DESeq2[63].

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

References
1. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13**, 720–731 (2012).
2. Core, L. & Adelman, K. Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev.* **33**, 960–982 (2019).
3. Mayer, A., Landry, H. M. & Churchman, L. S. Pause & go: from the discovery of RNA polymerase pausing to its functional implications. *Curr. Opin. Cell Biol.* **46**, 72–80 (2017).
4. Kang, J. Y., Mishanina, T. V., Landick, R. & Darst, S. A. Mechanisms of transcriptional pausing in bacteria. *J. Mol. Biol.* **431**, 4007–4029 (2019).
5. Vvedenskaya, I. O. et al. Interactions between RNA polymerase and the "core recognition element" counteract pausing. *Science* **344**, 1285–1289 (2014).
6. Larson, M. H. et al. A pause sequence enriched at translation start sites drives transcription dynamics in vivo. *Science* **344**, 1042–1047 (2014).
7. Imashimizu, M. et al. Visualizing translocation dynamics and nascent transcript errors in paused RNA polymerases in vivo. *Genome Biol.* **16**, 98 (2015).
8. Zhang, J. W. & Landick, R. A two-way street: regulatory interplay between RNA polymerase and nascent RNA structure. *Trends Biochem. Sci.* **41**, 293–310 (2016).
9. Weixlbaumer, A., Leon, K., Landick, R. & Darst, S. A. Structural basis of transcriptional pausing in bacteria. *Cell* **152**, 431–441 (2013).
10. Komissarova, N. & Kashlev, M. Transcriptional arrest: *Escherichia coli* RNA polymerase translocates backward, leaving the 3′ end of the RNA intact and extruded. *Proc. Natl Acad. Sci. USA* **94**, 1755–1760 (1997).
11. Saba, J. et al. The elemental mechanism of transcriptional pausing. *Elife* **8**, e40981 (2019).
12. Artsimovitch, I. & Landick, R. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc. Natl Acad. Sci. USA* **97**, 7090–7095 (2000).
13. Borukhov, S., Sagitov, V. & Goldfarb, A. Transcript cleavage factors from *Escherichia Coli*. *Cell* **72**, 459–466 (1993).
14. Marr, M. T. & Roberts, J. W. Function of transcription cleavage factors GreA and GreB at a regulatory pause site. *Mol. Cell* **6**, 1275–1285 (2000).
15. Abdelkareem, M. et al. Structural basis of transcription: RNA polymerase backtracking and its reactivation. *Mol. Cell* **75**, 298–309 (2019).
16. Paget, M. S. & Helmann, J. D. The σ^{70} family of sigma factors. *Genome Biol.* **4**, 203 (2003).
17. Mooney, R. A., Darst, S. A. & Landick, R. Sigma and RNA polymerase: an on-again, off-again relationship? *Mol. Cell* **20**, 335–345 (2005).
18. Deighan, P., Pukhrambam, C., Nickels, B. E. & Hochschild, A. Initial transcribed region sequences influence the composition and functional properties of the bacterial elongation complex. *Genes Dev.* **25**, 77–88 (2011).
19. Harden, T. T. et al. Bacterial RNA polymerase can retain σ^{70} throughout transcription. *Proc. Natl Acad. Sci. USA* **113**, 602–607 (2016).
20. Bar-Nahum, G. & Nudler, E. Isolation and characterization of σ^{70}-retaining transcription elongation complexes from *Escherichia coli*. *Cell* **106**, 443–451 (2001).
21. Ring, B. Z., Yarnell, W. S. & Roberts, J. W. Function of E. coli RNA polymerase σ factor σ^{70} in promoter-proximal pausing. *Cell* **86**, 485–493 (1996).

22. Kainz, M. & Roberts, J. Structure of transcription elongation complexes in vivo. *Science* **255**, 838–841 (1992).

23. Strobel, E. J. & Roberts, J. W. Two transcription pause elements underlie a σ70-dependent pause cycle. *Proc. Natl Acad. Sci. USA* **112**, E4374–E4380 (2015).

24. Zhilina, E., Esyunina, D., Brodolin, K. & Kulbachinskiy, A. Structural transitions in the transcription elongation complexes of bacterial RNA polymerase during σ-dependent pausing. *Nucleic Acids Res.* **40**, 3078–3091 (2012).

25. Stepanova, E., Wang, M., Severinov, K. & Borukhov, S. Early transcriptional arrest at *Escherichia coli rplN* and *ompX* promoters. *J. Biol. Chem.* **284**, 35702–35713 (2009).

26. Perdue, S. A. & Roberts, J. W. A backtrack-inducing sequence is an essential component of *Escherichia coli* σ70-dependent promoter-proximal pausing. *Mol. Microbiol.* **78**, 636–650 (2010).

27. Nickels, B. E., Mukhopadhyay, J., Garrity, S. J., Ebright, R. H. & Hochschild, A. The σ70 subunit of RNA polymerase mediates a promoter-proximal pause at the *lac* promoter. *Nat. Struct. Mol. Biol.* **11**, 544–550 (2004).

28. Brodolin, K., Zenkin, N., Mustaev, A., Mamaeva, D. & Heumann, H. The σ70 subunit of RNA polymerase induces *lacUV5* promoter-proximal pausing of transcription. *Nat. Struct. Mol. Biol.* **11**, 551–557 (2004).

29. Churchman, L. S. & Weissman, J. S. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**, 368–373 (2011).

30. Hatoum, A. & Roberts, J. Prevalence of RNA polymerase stalling at *Escherichia coli* promoters after open complex formation. *Mol. Microbiol.* **68**, 17–28 (2008).

31. Shultzaberger, R. K., Chen, Z., Lewis, K. A. & Schneider, T. D. Anatomy of *Escherichia coli* σ70 promoters. *Nucleic Acids Res.* **35**, 771–788 (2007).

32. Santos-Zavaleta, A. et al. RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in *E. coli* K-12. *Nucleic Acids Res.* **47**, D212–D220 (2019).

33. Vvedenskaya, I. O. et al. Massively systematic transcript end readout, "MASTER": transcription start site selection, transcriptional slippage, and transcript yields. *Mol. Cell* **60**, 953–965 (2015).

34. Feng, G. H., Lee, D. N., Wang, D., Chan, C. L. & Landick, R. GreA-induced transcript cleavage in transcription complexes containing *Escherichia coli* RNA polymerase is controlled by multiple factors, including nascent transcript location and structure. *J. Biol. Chem.* **269**, 22282–22294 (1994).

35. Susa, M., Kubori, T. & Shimamoto, N. A pathway branching in transcription initiation in *Escherichia coli*. *Mol. Microbiol.* **59**, 1807–1817 (2006).

36. Singh, S. S. et al. Widespread suppression of intragenic transcription initiation by H-NS. *Genes Dev.* **28**, 214–219 (2014).

37. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).

38. Chen, F. X., Smith, E. R. & Shilatifard, A. Born to run: control of transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **19**, 464–478 (2018).

39. Browning, D. F. & Busby, S. J. Local and global regulation of transcription initiation in bacteria. *Nat. Rev. Microbiol.* **14**, 638–650 (2016).

40. Kiryu, H., Oshima, T. & Asai, K. Extracting relations between promoter sequences and their strengths from microarray data. *Bioinformatics* **21**, 1062–1068 (2005).

41. Hook-Barnard, I. G. & Hinton, D. M. Transcription initiation by mix and match elements: flexibility for polymerase binding to bacterial promoters. *Gene Regul. Syst. Bio.* **1**, 275–293 (2007).

42. Henderson, K. L. et al. Mechanism of transcription initiation and promoter escape by *E. coli* RNA polymerase. *Proc. Natl Acad. Sci. USA* **114**, E3032–E3040 (2017).

43. Wade, J. T. & Struhl, K. The transition from transcriptional initiation to elongation. *Curr. Opin. Genet. Dev.* **18**, 130–136 (2008).

44. Revyakin, A., Liu, C., Ebright, R. H. & Strick, T. R. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* **314**, 1139–1143 (2006).

45. Shi, J. et al. Structural basis of Q-dependent transcription antitermination. *Nat. Commun.* **10**, 2925 (2019).

46. 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).

47. Sheridan, R. M., Fong, N., D'Alessandro, A. & Bentley, D. L. Widespread backtracking by RNA Pol II is a major effector of gene activation, 5′ pause release, termination, and transcription elongation rate. *Mol. Cell* **73**, 107–118 (2019).

48. Shankar, S., Hatoum, A. & Roberts, J. W. A transcription antiterminator constructs a NusA-dependent shield to the emerging transcript. *Mol. Cell* **27**, 914–927 (2007).

49. Kang, J. Y. et al. Structural basis for transcript elongation control by NusG family universal regulators. *Cell* **173**, 1650–1662 (2018).

50. Artsimovitch, I. & Landick, R. The transcriptional regulator RfaH stimulates RNA chain synthesis after recruitment to elongation complexes by the exposed nontemplate DNA strand. *Cell* **109**, 193–203 (2002).

51. Yakhnin, A. V. et al. NusG controls transcription pausing and RNA polymerase translocation throughout the Bacillus subtilis genome. *Proc. Natl Acad. Sci. USA* **117**, 21628–21636 (2020).

52. Mooney, R. A. et al. Regulator trafficking on bacterial transcription units in vivo. *Mol. Cell* **33**, 97–108 (2009).

53. Petushkov, I., Esyunina, D. & Kulbachinskiy, A. Possible roles of σ-dependent RNA polymerase pausing in transcription regulation. *RNA Biol.* **14**, 1678–1682 (2017).

54. Ray-Soni, A., Bellecourt, M. J. & Landick, R. Mechanisms of bacterial transcription termination: all good things must end. *Annu. Rev. Biochem.* **85**, 319–347 (2016).

55. Ohkuma, Y. et al. Structural motifs and potential σ homologies in the large subunit of human general transcription factor TFIIE. *Nature* **354**, 398–401 (1991).

56. Burton, S. P. & Burton, Z. F. The σ enigma: bacterial σ factors, archaeal TFB and eukaryotic TFIIB are homologs. *Transcription* **5**, e967599 (2014).

57. Sainsbury, S., Bernecky, C. & Cramer, P. Structural basis of transcription initiation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16**, 129–143 (2015).

58. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).

59. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).

60. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

61. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

62. Anders, S., Pyl, P. T. & Huber, W. HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

63. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

64. Thomason, M. K. et al. Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in *Escherichia coli*. *J. Bacteriol.* **197**, 18–28 (2015).

65. Komissarova, N. & Kashlev, M. Functional topography of nascent RNA in elongation intermediates of RNA polymerase. *Proc. Natl Acad. Sci. USA* **95**, 14699–14704 (1998).

Acknowledgements

Author contributions

Z.S. and M.K. conceived and designed the project. A.Y. optimized the RNET-seq workflow. Z.S. performed the RNET-seq, RNA-seq and biochemical experiments. Z.S., C.M. and P.F. wrote and executed the custom scripts. Z.S. and M.K. analyzed the data. Z.S. and M.K. wrote the manuscript.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-21150-2.

Correspondence and requests for materials should be addressed to M.K.

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

Reprints and permission information is available at http://www.nature.com/reprints