nature structural & molecular biology

RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes

Kira Glover-Cutter[1,3], Soojin Kim[1,3], Joaquin Espinosa[2] & David L Bentley[1]

We investigated co-transcriptional recruitment of pre-mRNA processing factors to human genes. Capping factors associate with paused RNA polymerase II (pol II) at the 5′ ends of quiescent genes. They also track throughout actively transcribed genes and accumulate with paused polymerase in the 3′ flanking region. The 3′ processing factors cleavage stimulation factor and cleavage polyadenylation specificity factor are maximally recruited 0.5–1.5 kilobases downstream of poly(A) sites where they coincide with capping factors, Spt5, and Ser2-hyperphosphorylated, paused pol II. 3′ end processing factors also localize at transcription start sites, and this early recruitment is enhanced after polymerase arrest with the elongation factor DRB. These results suggest that promoters may help specify recruitment of 3′ end processing factors. We propose a dual-pausing model wherein elongation arrests near the transcription start site and in the 3′ flank to allow co-transcriptional processing by factors recruited to the pol II ternary complex.

An intricate web of interactions between the RNA pol II elongation complex and pre-mRNA processing factors ensures efficient and accurate mRNA biogenesis by coordinating maturation with synthesis of the transcript[1–4]. Co-transcriptional engagement of processing factors involves recognition of consensus sites in the nascent pre-mRNA and protein-protein interactions with components of the pol II ternary complex, including the C-terminal domain (CTD)[2,5,6]. The pol II CTD comprises heptad repeats (YSPTSPS) that are reversibly phosphorylated at the Ser2 and Ser5 positions as polymerase traverses a gene[5,7], and these modifications have been proposed to regulate recruitment of pre-mRNA processing factors. The CTD is essential for efficient pre-mRNA processing in mammalian systems[2–4] but has only a minor role in budding yeast[8]; therefore, recruitment mechanisms for processing factors almost certainly differ between these species. In contrast to the situation in yeast, relatively little is known about the timing and order of recruitment of different processing factors during the pol II transcription cycle in metazoan cells, although splicing factors and 3′ end processing factors have been localized at active transcription sites[9–12].

There are two human enzymes for maturation of 5′ ends: human capping enzyme (HCE), a bifunctional triphosphatase-guanylyltransferase, and RNA 7-methyltransferase (MT)[13]. Capped pol II transcripts as short as 25–40 bases are made in transcription extracts and isolated nuclei, indicating that capping is a very early co-transcriptional event[14–17]. Both HCE and MT can bind directly to pol II that is phosphorylated on the CTD[18], but *in vitro* studies indicate that these interactions with actively transcribing pol II are weak[16,17], and their *in vivo* significance remains unknown.

Co-transcriptional recruitment of capping enzymes may be coordinated with promoter-proximal pausing[15–17,19], a common feature of pol II transcription of mammalian but not yeast genes[20,21]. In support of this idea, pol II pauses at the 5′ end of the *DHFR* gene, and HCE localizes in the same region[22]. Promoter-proximal pausing is mediated by negative elongation factor (NELF) and DRB sensitivity-inducing factor (DSIF) comprising Spt4 and Spt5. DRB arrests pol II elongation by inhibiting Cdk9, which phosphorylates the pol II CTD, DSIF and NELF[20,21]. Spt5 binds directly to HCE and stimulates capping[23]. These observations led to the hypotheses that promoter-proximal pausing facilitates capping[17,19] or, alternatively, that recruitment of capping enzyme facilitates release from the pause[24]. These possibilities have yet to be investigated in detail *in vivo* by localizing capping enzymes relative to pol II and Spt5.

Three different processing machineries carry out 3′ end maturation of polyadenylated mRNAs, nonadenylated histone mRNAs and U small nuclear RNAs. Cleavage and polyadenylation require cleavage stimulation factor (CstF) and cleavage polyadenylation specificity factor (CPSF)[25], which both share common subunits with the histone 3′ end processing complex, heat-labile factor (HLF)[26,27]. U snRNA 3′ ends are processed by a distinct CTD binding complex[28]. CstF binds to the pol II CTD through its 50-kDa subunit and to G/U-rich sequences adjacent to the poly(A) site through its 64-kDa subunit. CstF77 bridges the other two subunits and contacts CPSF160, which binds the AAUAAA sequence[25]. CPSF73 is an endonuclease that cleaves poly(A) sites[29] and histone 3′ ends[26]. CPSF can bind to the transcription factor TFIID and to the body of pol II, supporting the theory that CPSF binds first to the promoter via TFIID and is then transferred to

[1]Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, UCHSC, MS8101, PO Box 6511, Aurora, Colorado 80045, USA. [2]Department of Molecular Cellular and Developmental Biology, University of Colorado, Boulder, Colorado 80309, USA. [3]These authors contributed equally to this work. Correspondence should be addressed to D.L.B. (david.bentley@UCHSC.edu).

Received 19 August; accepted 21 November; published online 23 December 2007; doi:10.1038/nsmb1352

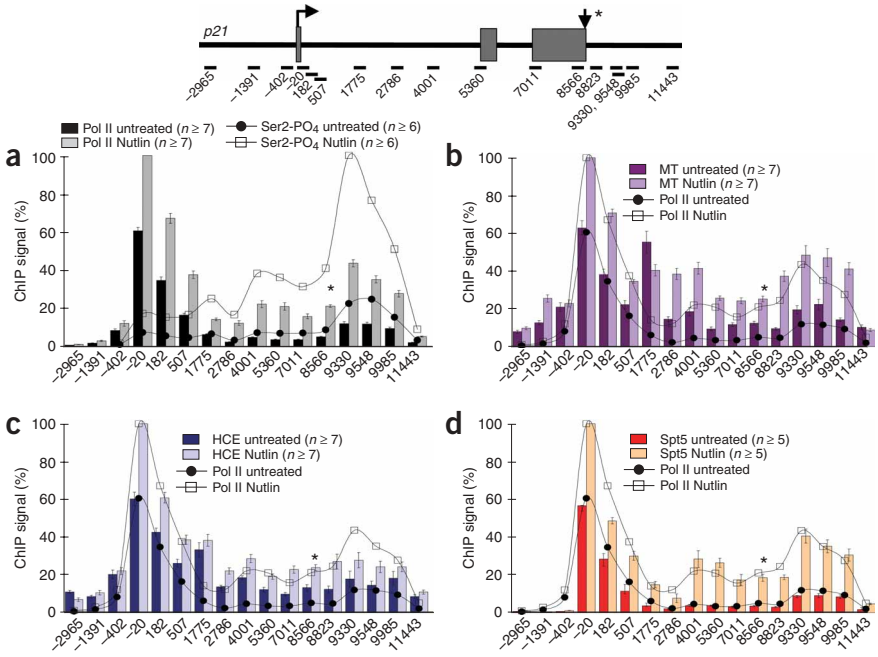

Figure 1 Capping enzymes localize with paused pol II at 5′ and 3′ ends and throughout the *p21* gene. Top, diagram of *p21* with the center of real-time PCR products marked relative to the transcription start site as in ref. 36. (**a**–**d**) Relative ChIP signals in HCT116 cells plotted for total pol II and CTD Ser2-PO_4 (**a**), cap methyltransferase (MT) (**b**), human capping enzyme (HCE) (**c**) and Spt5 (**d**) before (untreated, dark bars) and after activation by Nutlin-3a (pale bars). Down arrow and * indicate poly(A) signal at +8570. Flanking primers –2965 and +11443 serve as intergenic background controls. Pol II traces from **a** are included in **b**–**d**. Values are normalized to the maximal value in the Nutlin-3a data set. For reasons we do not understand, the density of capping enzymes relative to pol II and Spt5 was reproducibly higher at amplicon +1775 in intron 1 than at other positions. Mean values from *n* PCR reactions with s.e.m. are shown.

pol II[30,31]. During elongation CPSF has been proposed to associate with pol II in a manner that excludes CstF, and, later, CstF is recruited at a pause site after transcription of the AAUAAA element[30,32]. Pausing has been detected downstream of the β-actin gene (*ACTB*), where it has been linked to transcription termination[33]. It is not clear whether the phenomenon of 3′ pausing is general among mammalian genes or how it is related to recruitment of 3′ end processing factors *in vivo*. Whereas cleavage/polyadenylation factors have been detected on metazoan genes[10–12], the 5′–3′ distribution and order of recruitment of CPSF and CstF on genes *in vivo* have yet to be determined.

The basis for specific recruitment of different mRNA 3′ end processing complexes to different genes is a major unresolved question. Two possible solutions are (i) that signals in the nascent RNA recruit the appropriate processing factors and (ii) that elements in the promoter dictate recruitment of appropriate factors[31,34,35], perhaps by specifying different modifications of the pol II CTD. It is not clear whether 3′ end processing factors localize to transcription start sites *in vivo*, but if they do, it would lend support to the notion that promoters help to specify their recruitment. In this report we localized pre-mRNA 5′ and 3′ end formation factors relative to pol II on human genes to determine where on the gene, and when in the transcription cycle, they are recruited.

RESULTS

Capping enzymes colocalize with pol II paused at promoters

To investigate co-transcriptional recruitment of capping factors, we mapped HCE, MT and pol II by high-resolution chromatin immunoprecipitation (ChIP) on the human *p21* (*CDKN1A*) gene before and after transcriptional activation. Protein occupancy was assayed at 17 positions within *p21* and its flanking sequences, in HCT116 cells where the gene can be activated by p53 (ref. 36) (**Fig. 1**, top diagram). Each position corresponds to a 50- to 125-base pair (bp) amplicon that was quantified by real-time PCR using SybrGreen fluorescence (see Methods). Before transcriptional activation, under conditions wherein *p21* mRNA is very low by real-time RT-PCR, a high density of pol II is loaded at the *p21* transcription start site[36] (amplicon –20, **Fig. 1a**) where it is precisely colocalized with MT and HCE (**Fig. 1b**,**c**). Spt5 also colocalized with HCE, MT and pol II near the transcription start site of the uninduced *p21* (**Fig. 1d**) and *c-fos* (*FOS*) genes (**Supplementary Fig. 1** online). Together these results indicate that before activation, the *p21* and *c-fos* genes are primed at their 5′ ends with paused pol II transcription complexes that are associated with Spt5 and capping factors.

Capping enzymes and paused pol II at the 5′ and 3′ ends of *p21*

When *p21* transcription was induced by activating p53 with the HDM2 inhibitor (a p53 binding protein), Nutlin-3a, pol II and Spt5 occupancy increased at the transcription start site and throughout the length of the gene but not in the intergenic control regions (**Fig. 1**). Unexpectedly, we consistently observed coincident peaks of pol II and Spt5 density at amplicons +9330, +9548, and +9985 located 1–1.5 kilobases (kb) downstream of the poly(A) site at +8570 (see **Fig. 1a**,**d**). Pol II and Spt5 occupancy fell off between +9985 and +11443 as transcription terminated. Spt5 (DSIF) therefore seems to be a constitutive component of pol II transcription complexes at initiation, elongation and termination phases of the transcription cycle.

In summary, ChIP revealed a distribution of pol II and Spt5 with twin peaks at the *p21* transcription start site and well downstream of the poly(A) site. This distinctive distribution was specific to pol II and Spt5 and was not observed for TFIIB or trimethylated histone H3K4, which show peaks near the 5′ end but not in the 3′ flanking region (**Fig. 2a** and **Supplementary Fig. 2a** online). The peak of pol II downstream of *p21* coincided with a valley in histone H3 occupancy, consistent with histone displacement associated with high pol II density[37] (**Fig. 2a**). We interpret this 3′ peak of pol II that precedes termination as the result of pausing. Notably, the highest level of CTD Ser2 phosphorylated pol II on *p21* occurred at this 3′ pause site located about 1 kb downstream of the poly(A) site (**Fig. 1a**).

When *p21* was activated, HCE and MT occupancy increased, with pol II not only at the 5′ end but also at positions within the gene (**Fig. 1b**,**c**). Consistent with early capping of nascent transcripts, the cap-binding complex (CBC) was also present at the 5′ end of activated *p21* (**Supplementary Fig. 2b**), similar to previous results at *c-fos*[9].

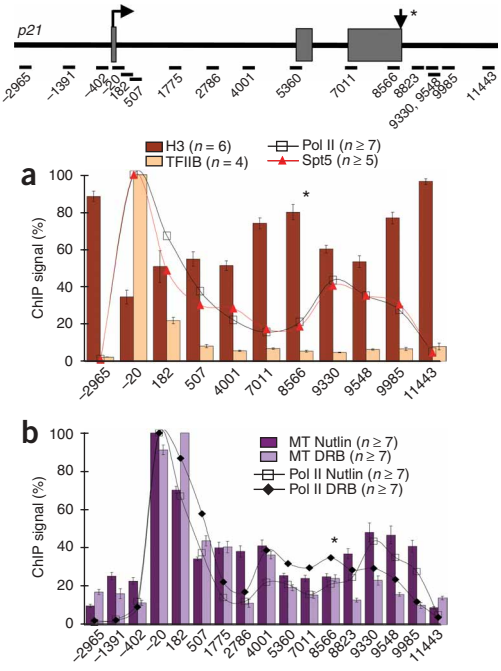

Figure 2 Pol II pausing and histone localization at 5′ and 3′ ends of *p21*. (**a**) Pol II, Spt5, histone H3 and TFIIB localization by ChIP on the Nutlin-3a activated *p21*. Displacement of H3 occurs at sites of pol II accumulation at the 5′ and 3′ ends. Note that unlike pol II and Spt5, TFIIB peaks near the 5′ end but not at the 3′ end. (**b**) DRB inhibits pausing and accumulation of MT downstream of the *p21* poly(A) site. ChIP of pol II and MT after induction by Nutlin-3a or DRB normalized to the maximum value in each data set to emphasize differences in 5′–3′ distribution. Mean values from *n* PCR reactions with s.e.m. are shown.

HCE and MT levels, normalized to pol II, remained relatively high throughout the gene and 3′ flank (**Supplementary Table 1a** online), demonstrating that their association with pol II complexes persists during elongation and termination. Although methylation is delayed relative to guanylation in coupled transcription and capping reactions *in vitro*[16,17], we did not observe any delay in recruitment of MT relative to HCE. Notably, MT, and to a lesser extent HCE, also accumulated with paused pol II and Spt5 downstream of the poly(A) site (**Fig. 1b**,**c**).

Unlike most genes, *p21* is activated by the elongation inhibitor DRB (**Fig. 2b**), resulting in transcription by pol II that is hypophosphorylated on CTD Ser2 (**Supplementary Table 1a**)[36]. We asked whether DRB activation altered the association of capping factors with *p21*. Relative to Nutlin-3a, DRB suppressed the 3′ pause and this loss of pol II from the 3′ flanking region correlated with reduced MT (**Fig. 2b**) and Spt5 (**Supplementary Fig. 2c**) occupancy. Despite this inhibition of pausing, DRB did not prevent termination, as shown by the loss of pol II between positions +9330 and +11443 (**Fig. 2b**). In summary, the experiments in **Figures 1** and **2** show that capping factors track with pol II and Spt5 in transcription complexes that are paused at 5′ and 3′ ends as well as in those that are elongating through the gene. The capping factors therefore seem to remain in association with pol II long after addition of the cap, which is expected to occur within the first 50 bases[15].

5′ and 3′ pol II pausing with capping factors on multiple genes

To determine whether the twin peaks of pol II and capping factors at the 5′ and 3′ ends of *p21* are common to other genes, we also analyzed *MYC* and *GAPDH* (**Fig. 3**). The maximum occupancy of capping factors on *MYC* and *GAPDH* coincided with the peaks of pol II and Spt5 at the transcription start sites, as previously reported for *DHFR*[22], but they were also detected at low levels within the genes (**Fig. 3b**,**d**). Downstream of the poly(A) sites, MT density increased coincident with paused pol II that is highly phosphorylated on CTD Ser2 residues (**Fig. 3a**,**c**). Similar results were obtained in HCT116 and HeLa cells (**Supplementary Fig. 3a** online). Likewise, increased HCE density was detected downstream of *MYC* but not *GAPDH* (**Fig. 3b**,**d**). The absence of HCE from the 3′ end of *GAPDH* may be due to a gene-specific effect of the sequence of the poly(A) site or 3′ flanking region. Together these results suggest that a quite general feature of human genes is that elongation is punctuated by dual pauses at the transcription

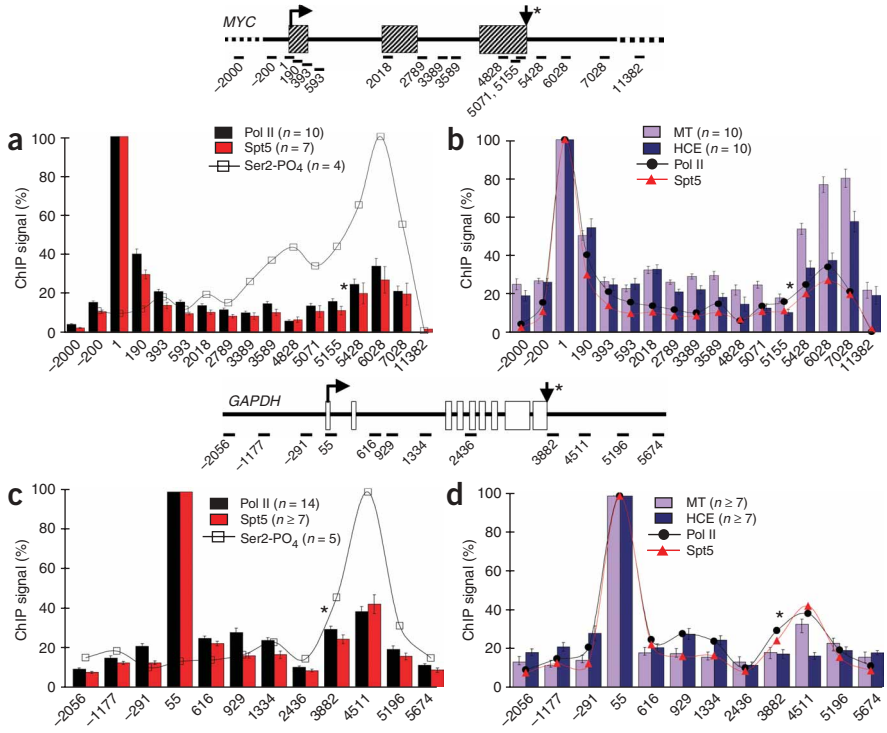

Figure 3 Capping enzymes track with pol II on the *MYC* and *GAPDH* genes. Positions of real-time PCR products are indicated relative to the *MYC* P2 start site. (**a**) Relative ChIP signals are shown for total pol II, CTD Ser2-PO_4 and Spt5. (**b**) MT and HCE on *MYC* in HCT116 cells. Traces of pol II and Spt5 from **a** serve as reference. (**c**) Total pol II, CTD Ser2-PO_4 and Spt5 on *GAPDH*. (**d**) MT and HCE on *GAPDH*. Traces of pol II and Spt5 from **c** serve as reference. * and down arrow indicate poly(A) signals at +5166 relative to the *MYC* P2 transcription start site, and +3830 for *GAPDH*. Mean values from *n* PCR reactions with standard errors of mean (s.e.m.) are shown.

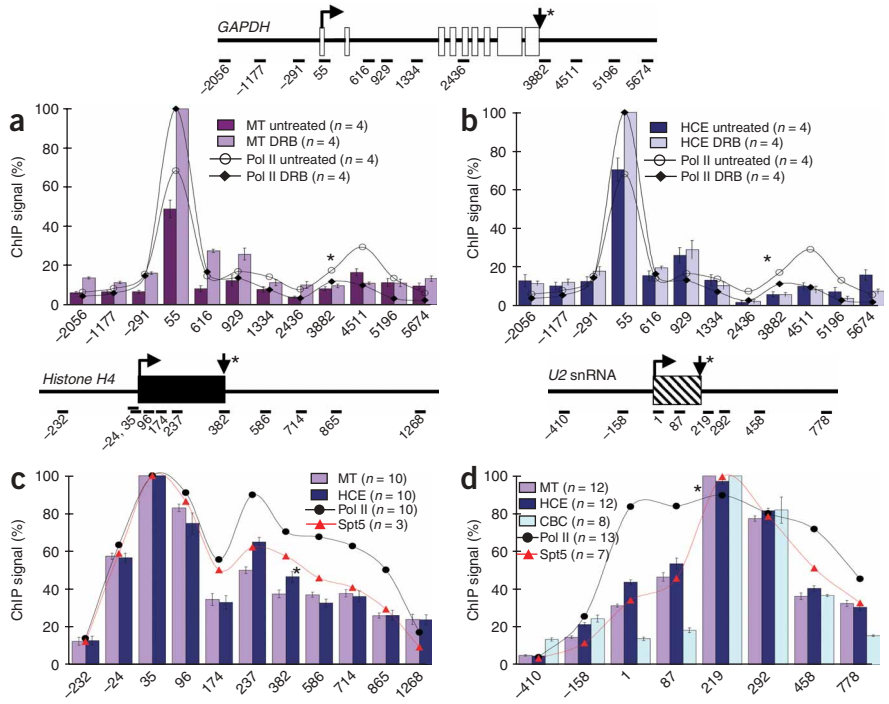

Figure 4 Localizaton of capping factors on various human genes. (**a**,**b**) Association of capping factors with DRB-arrested pol II complexes on *GAPDH*. Positions of real-time PCR products are indicated. Relative ChIP signals are shown for pol II, MT and HCE on the *GAPDH* gene in untreated and DRB treated HCT116 cells. (**c**,**d**) Localization of capping factors, pol II and Spt5 on the *histone H4* (*H4/d*) and *U2* snRNA genes. (**c**) Pol II, Spt5 MT and HCE on *H4/d* (NCBI accession number X60483). (**d**) Pol II, Spt5, MT, HCE and CBC on *U2* snRNA gene (NCBI accession number U57614). CBC data are the sum of four determinations each with anti-CBC20 and 80. Mature RNA 3′ ends (down arrow and *) map to +387 for *H4/d* and +190 for *U2* snRNA. Mean values from *n* PCR reactions with standard errors of mean (s.e.m.) are shown.

start site and in the 3′ flanking region, where pol II accumulates in association with Spt5 and one or both capping factors.

The colocalization of HCE and MT with paused pol II at 5′ ends of uninduced *p21* and *c-fos* (**Fig. 1** and **Supplementary Fig. 1**) indicates that they may associate with polymerases that are not engaged in active elongation. We investigated this idea further by asking whether capping factors localize with pol II that has been arrested by DRB. As expected, inhibition of Cdk9 with DRB increased pol II occupancy at the transcription start site and decreased the CTD phospho-Ser2 signal at the 3′ end of *GAPDH* (data not shown), but it did not entirely eliminate pol II from positions near the 3′ end of the gene (**Fig. 4a**), consistent with previous results[38]. Notably, both capping factors accumulated with DRB-arrested pol II at the transcription start site, suggesting that pol II complexes that are not actively transcribing are still associated with capping factors (**Fig. 4a**,**b**).

Capping factors and Spt5 at the histone H4 and U2 snRNA genes

Histone mRNAs and U snRNAs are among the shortest pol II transcripts, and it is not known whether they are capped co-transcriptionally or post-transcriptionally. We found that capping factors and Spt5 were co-transcriptionally recruited to the *histone H4/d* and *U2* snRNA genes at levels comparable to those for poly(A)$^+$ mRNA coding genes after normalizing to pol II (**Supplementary Table 1b**). On *H4/d*, HCE, MT and Spt5 peaked near the 5′ end and persisted with pol II through the open reading frame and into the 3′ flanking region (**Fig. 4c**), similar to what is seen for poly(A)$^+$ mRNA genes.

On *U2*, in contrast to other genes, the 5′–3′ profiles of capping factors and Spt5 did not overlap precisely with pol II. Pol II density remained approximately constant across the *U2* gene cluster with no evidence of promoter-proximal pausing (**Fig. 4d**, +1, +87, +219). In contrast, HCE, MT, Spt5 and CBC levels rose between the 5′ and 3′

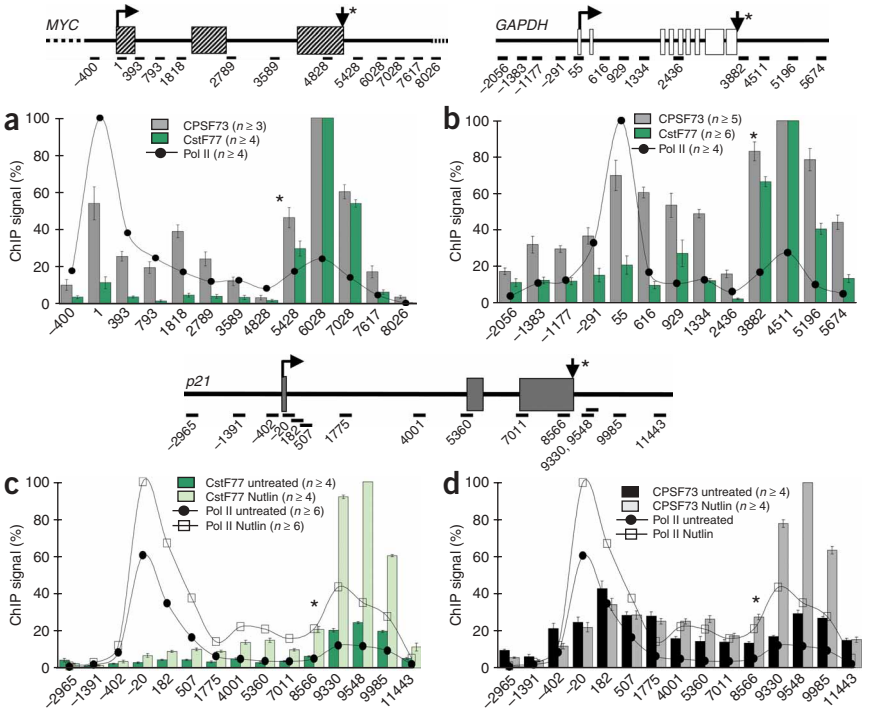

Figure 5 Localization of 3′ end processing factors at 5′ and 3′ ends of *p21*, *MYC* and *GAPDH*. Positions of real-time PCR products on each gene are indicated. (**a**,**b**) Relative ChIP signals in HCT116 cells are shown for CPSF73 and CstF77 on *MYC* (**a**) and *GAPDH* (**b**). (**c**,**d**) CstF77 (**c**) and CPSF73 (**d**) on untreated and Nutlin-3a activated *p21* normalized to the maximum value for the Nutlin-3a data set. * and down arrow indicate poly(A) sites. Mean values from *n* PCR reactions with standard errors of mean (s.e.m.) are shown.

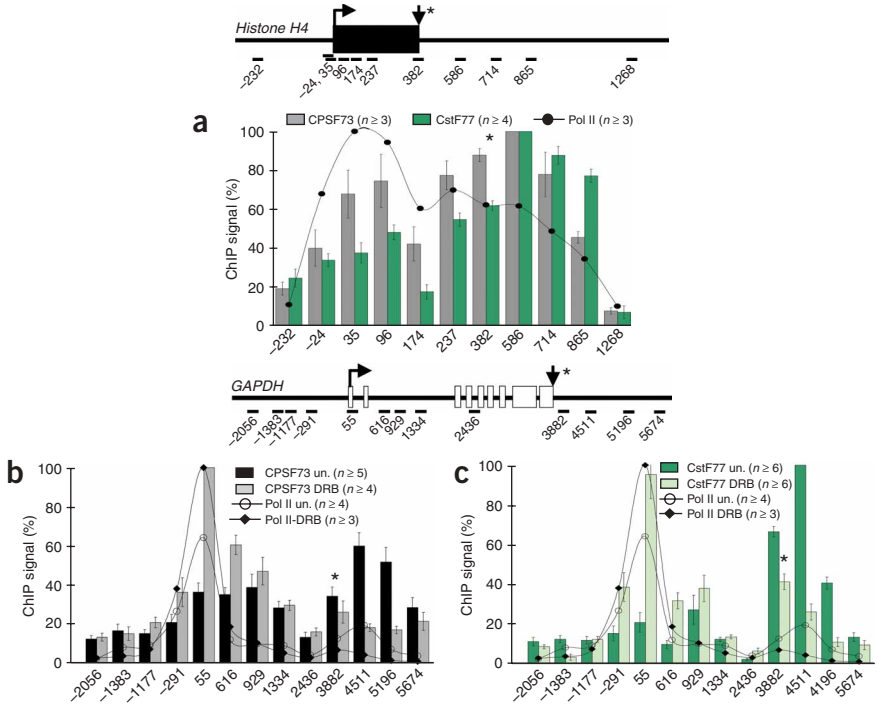

Figure 6 Co-transcriptional recruitment of 3′ end processing factors to a histone gene. (**a**) ChIP of CPSF73 and CstF77 subunits of HLF and pol II on the histone H4/d gene *HIST1H4K* in HCT116 cells. Down arrow and * indicates mRNA 3′ end. (**b**,**c**) DRB enhances recruitment of 3′ end processing factors at the 5′ end of GAPDH. Relative ChIP signals are shown for CPSF73 (**b**) and CstF77 (**c**) on *GAPDH*, both untreated (un.) and after DRB treatment. Pol II and CstF77 profiles for untreated cells are the same as **Fig. 5b**. The untreated CPSF73 data are a subset of those in **Figure 5b**. Data in **b** are normalized to the maximum value for the DRB sample at +55, and those in **c** are normalized to the maximum value for the untreated sample at +4511. Mean values from *n* PCR reactions with standard errors of mean (s.e.m.) are shown.

ends of the gene (**Fig. 4d**), peaking at +219, close to the 3′ box. In summary, recruitment of Spt5, capping factors and CBC to *U2* is delayed relative to pol II. These results are consistent with co-transcriptional capping of *U2* snRNA and furthermore suggest that capping factors and Spt5 are coordinately recruited[23] without detectable promoter-proximal pausing. The apparent lack of pausing is consistent with our finding that recruitment of NELF-A, relative to pol II, is approximately five times less at *U2* than at *H4/d* (data not shown).

Colocalization of 3′ processing factors and pol II at 5′ and 3′ ends

To investigate the timing and order of cleavage/polyadenylation factor recruitment, we localized CstF and CPSF by ChIP with antibodies to CstF77 and CPSF73. Peaks of CstF and CPSF recruitment exactly coincided with one another and with the accumulation of paused pol II 0.5–1.5 kb downstream of the *MYC* and *GAPDH* poly(A) sites (**Fig. 5a**,**b**). These factors were not detected on the *U2* snRNA gene (**Supplementary Fig. 3b**). High levels of CPSF73 were also recruited independently of CstF77 near the transcription initiation sites and within *MYC* and *GAPDH* (**Fig. 5a**,**b**). These results show that CPSF recruitment starts early in the transcription cycle, probably close to the time of initiation. Most CstF77 recruitment, by contrast, occurs late, after synthesis of the poly(A) site.

We also examined recruitment of CstF and CPSF to *p21* before and after transcriptional activation by Nutlin-3a (**Fig. 5c**,**d**). After activation, CstF77 was elevated three- to four-fold in the 3′ flank where recruitment was maximal. Low levels were also detectable within the gene and at the transcription start site (**Fig. 5c**). Unlike CstF77, CPSF73 and CstF64 accumulated near the start site of the uninduced *p21* gene. Transcription activation had little effect on CPSF73 and CstF64 at the start site, but enhanced their recruitment downstream of the poly(A) site, coincident with paused pol II (**Fig. 5d**, **Supplementary Fig. 4a** online). Notably, the binding of CPSF73 and CstF64 near the 5′ end of *p21* resembles that of TBP, which is also mostly recruited before activation[36]. In summary, the results in **Figure 5** and **Supplementary Figure 4a** suggest that CPSF73 and CstF64 can be recruited to a 5′ end independently of CstF77 (see Discussion). Similarly, the CPSF73 and CstF77 subunits of the histone 3′ end processing complex (HLF) were recruited near the 5′ end of *H4/d* and persisted throughout its length, with maximal accumulation about 200 bases downstream of the 3′ cleavage site (**Fig. 6a**). Sm proteins, presumably associated with small nuclear ribonucleoproteins such as U7, were also localized throughout the *H4/d* gene (**Supplementary Fig. 4b**).

Inhibitors of Cdk9 cause elongational arrest and can disrupt cleavage and polyadenylation[38,39], but it is not clear how they affect recruitment of 3′ end processing factors. We investigated how DRB affects pol II, CstF and CPSF localization. DRB enhanced the 5′ pause and suppressed the 3′ pause on *GAPDH* (**Fig. 6b**), as we observed on *p21* (**Fig. 2b**). DRB reduced CstF and CPSF levels downstream of the *GAPDH*, *p21* and *MYC* poly(A) sites (**Fig. 6b**,**c**, **Supplementary Fig. 5a**,**b** online, **Supplementary Table 2** online). It also reduced HLF levels downstream of *H4/d* (**Supplementary Fig. 5c**,**d**). Most of this suppression can be attributed to reduced pol II occupancy, but we did detect an additional reduction in CstF77 and CPSF73 relative to pol II, downstream of *p21*, *MYC* and *H4/d* (**Supplementary Table 2**). The most remarkable effect of DRB is that it elevated CPSF73 and CstF77 binding to the 5′ ends of *GAPDH*, *p21* and *MYC* relative to pol II (**Fig. 6b**,**c**, **Supplementary Table 2**). This result therefore implicates Cdk9 phosphorylation in controlling the recruitment of 3′ end processing factors at 5′ ends.

DISCUSSION

Colocalization of capping factors with pol II at 5′ and 3′ ends

Pre-mRNA processing factors can, in principle, be recruited to transcribed genes in two ways that are not mutually exclusive: (i) by binding to processing signals in nascent pre-mRNA and (ii) by binding to a protein 'landing pad' in the pol II elongation complex. Our ChIP analysis shows extensive colocalization of 5′ and 3′ end processing factors with pol II even at positions quite distant from where capping and cleavage/polyadenylation take place. Capping enzymes were situated not only at the 5′ end where capping is thought to occur[15], but also throughout the gene and in 3′ flanking regions more than a kilobase downstream of poly(A) sites (**Figs. 1** and **3**). In other words, capping factors seem to linger on the pol II 'landing pad'

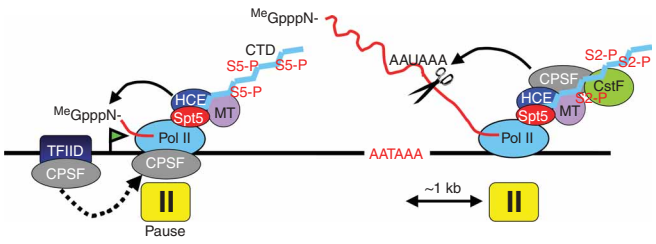

Figure 7 Dual-pausing model for co-transcriptional 5′ and 3′ end maturation of polyadenylated mRNAs. This model is based on data reported here and in refs. 30 and 31.

long after addition of the cap. This association could be due to stable binding initiated at the 5′ end or to dynamic exchange throughout the length of the gene. The function of these factors within genes and in their 3′ flanking regions is unclear, but they are appropriately situated to influence elongation, termination and 3′ end maturation. There are several precedents for viral and cellular capping enzymes influencing transcription initiation, elongation and termination[24,40,41]. It is also of interest that HCE promoted poly(A) site cleavage in a transcription-coupled *in vitro* system[42]. The recruitment patterns we observed for human capping enzymes differ from their yeast counterparts. Yeast Ceg1 and Abd1 localize at 5′ ends, and Abd1 is also found within genes, but neither has been detected in 3′ flanking regions[7,43].

Human capping factors colocalized with pol II poised at promoters, before transcriptional activation and after inhibiting elongation with DRB (**Figs. 1b**,**c** and **4a**,**b**). Association of capping factors with pol II is therefore not sufficient to trigger release from a 5′ pause. Conversely, our results suggest that capping factors might collaborate with negative elongation factors, Spt4/5 and NELF, in mediating promoter-proximal pausing. Pol II pausing may not be a universal prerequisite for co-transcriptional capping, however, because capping factors and CBC were efficiently recruited to the *U2* snRNA gene (**Fig. 4d**, **Supplementary Table 1b**), where we did not detect pausing. Maximal capping factor and CBC density on *U2* mapped close to the 3′ box, coincident with a peak of Spt5 (**Fig. 4d**). This result is therefore consistent with the idea that Spt5 facilitates recruitment of capping factors[23,24]. Such a mechanism could also aid capping factor recruitment before the CTD has been extensively phosphorylated, such as at the 5′ ends of uninduced genes (**Fig. 1**, **Supplementary Fig. 1**).

Recruitment of 3′ processing factors at 3′ and 5′ ends

Major peaks of accumulation of CstF and CPSF coincide with pol II paused 0.5–1.5 kb downstream of poly(A) sites (**Fig. 5**). Similarly, HLF occupancy peaked downstream of H4/d (**Fig. 6a**). This 3′ recruitment probably involves recognition of the processing sites in the nascent transcript. 3′ end processing factors also localized far upstream of their cognate processing sites at positions within genes and at 5′ ends. This 5′ recruitment, in anticipation of transcription of processing sites, points to an additional recruitment mechanism distinct from RNA recognition: namely protein-protein binding to pol II transcription complexes. 3′ end processing factors therefore seem to differ in this regard from splicing factors, whose recruitment seems to require transcription of consensus splice sites[44].

CPSF73 and CstF64 localize close to transcription start sites, suggesting that their initial recruitment is linked to initiation **Fig. 5a**, **Supplementary Fig. 4a**, ref. 10). In contrast, high levels of CstF77 were only detected downstream of the *MYC*, *GAPDH* and *p21* poly(A) sites (**Figs. 5a**,**b**), suggesting that it is recruited later. On *p21*, most CPSF73 and CstF64 recruitment preceded transcriptional activation, whereas most CstF77 only arrived after the gene was activated (**Figs. 5c**,**d**, **Supplementary Fig. 4a**). The low abundance of CstF77 at the *p21* transcription start site is reproducibly elevated in response to activation by Nutlin-3a or DRB (**Fig. 5c**, **Supplementary Fig. 5b**). These results suggest the possibility that the trimeric CstF complex may assemble during transcription with recruitment of CstF64 before CstF77.

When elongation was inhibited by DRB, the normal order of recruitment of cleavage/polyadenylation factors was disrupted. In DRB, CstF77 was recruited early, along with CPSF73, near the 5′ end (**Fig. 6b**,**c**). DRB-arrested pol II complexes at 5′ ends may therefore adopt a conformation that stabilizes interactions with both these 3′ end processing factors. DRB could also enhance the accumulation of 3′ processing factors at 5′ ends by influencing the proposed handoff of CPSF from TFIID to pol II (**Fig. 7**)[31].

Two lines of evidence suggest that promoter elements may specify the recruitment of 3′ end processing factors to 5′ ends independent of the RNA signals that they recognize. First, a U snRNA promoter, which is recognized by a distinct TBP-containing complex, is important for U snRNA 3′ end processing[23,24] by the CTD-bound integrator complex[36]. Second, it has been suggested[31] that CPSF binds initially with TFIID to the promoter of mRNA genes and then is handed off to pol II, with which it travels into the gene (**Fig. 7**). Our results strongly support this model by showing that CPSF73 is recruited close to transcription start sites and persists throughout the length of genes (**Figs. 5** and **6**). 3′ end processing factors also localized near the 5′ end of *H4/d* (**Fig. 6a**, **Supplementary Table 2**). Together our results imply that promoter elements could have a general role in specifying the recruitment of 3′ end processing factors.

Pausing, termination and recruitment of 3′ end processing factors

Transcriptional pausing has been implicated both in enhancing cleavage and polyadenylation and as a prelude to termination[33,45,46], but the relationship between pausing, 3′ end formation and termination is still unclear. Interestingly, Spt5, which has negative elongation properties, colocalized with pol II at 3′ pause sites and during subsequent termination (**Figs. 1** and **3**, **Supplementary Fig. 2c**). A major conclusion of this study is that cleavage/polyadenylation factors are maximally recruited at the 3′ pause 0.5–1.5 kb downstream of poly(A) sites. This downstream pause detected by ChIP may correspond to the pause in live cells detected by fluorescence recovery after photobleaching in the 3′ portion of a reporter gene[47]. The 3′ pause was normally followed by termination (**Figs. 1** and **3**, **Supplementary Fig. 2c**), but termination could be uncoupled from pausing by DRB (**Figs. 2b** and **6b**).

The 3′ pause coincides with maximal CTD Ser2 phosphorylation (**Figs. 1a** and **3a**,**c**) and is suppressed by DRB, an inhibitor of Ser2 phosphorylation (**Figs. 2b** and **6b**, **Supplementary Table 1a**). In contrast, the 5′ pause occurs when pol II is hypophosphorylated on Ser2 and is enhanced by DRB. CTD Ser2 phosphorylation is usually regarded as a signature of elongating pol II within a gene[5,7]; however, our results point to a role for this modification well downstream of the gene where 3′ processing factors are recruited. How DRB inhibits cleavage and polyadenylation[38,39] is not completely resolved, but the process correlates with both reduced recruitment of CstF and CPSF in 3′ flanking regions and inhibition of 3′ pausing (**Fig. 6b**,**c**, **Supplementary Table 2**).

The torpedo model for coupling termination with 3′ end processing stipulates that RNA cleavage at the poly(A) site must precede

termination to permit entry of a 5′–3′ RNA exonuclease that degrades the transcript and facilitates polymerase release[48]. Whether poly(A) site cleavage normally occurs co-transcriptionally or post-transcriptionally is unclear[1,49]. Our demonstration that the endonuclease CPSF73 localizes on pol II complexes paused downstream of the poly(A) site indicates that the necessary factors are in place to carry out co-transcriptional cleavage before termination, as required by the torpedo model.

A dual-pausing model of co-transcriptional pre-mRNA processing

We propose a dual-pausing model for coordination of pol II transcription with pre-mRNA processing in metazoans (**Fig. 7**). In this model, 5′ and 3′ transcriptional pause sites mark where assembly of functional capping and cleavage/polyadenylation complexes occurs, and probably also where 5′ and 3′ end maturation of the nascent transcript takes place. HCE and MT were both recruited at 5′ pause sites (**Figs. 1** and **3**). This observation therefore supports *in vitro* data suggesting a link between co-transcriptional capping and promoter-proximal pausing[17,24]. We also propose that co-transcriptional poly(A) site cleavage takes place in the context of a pol II complex at the 3′ pause site where maximal recruitment of CstF and CPSF occurs (**Fig. 5**), approximately 1 kb downstream of the poly(A) site. Localization of the active pol II–3′ end processing complex well downstream of the poly(A) consensus sequence agrees well with the cleavage of *Chironomus* BR1 transcripts after 600 bases of downstream sequence has been transcribed[1]. This model is also supported by the fact that efficient poly(A) site cleavage requires an intact RNA tether linking the poly(A) site with the downstream polymerase[50,51].

The colocalization of capping and cleavage/polyadenylation factors with paused pol II at both ends of human genes is quite distinct from the recruitment patterns of homologous processing factors in budding yeast. We speculate that during evolution, co-transcriptional recruitment of processing factors has adapted to altered patterns of pol II transcription, including the punctuation of the transcription cycle in metazoans by prominent pause sites.

METHODS

Antibodies. Rabbit anti-pan CTD against pol II, anti-CstF77, anti-histone H3 and anti -H3K4me3 have been described (see **Supplementary Methods** online). Rabbit antibodies were raised against His-tagged mouse capping enzyme MCE (residues 211–597), mouse MT (residues 141–465), human Spt5 (residues 703–1087) and full-length human GST-TFIIB. Rabbit anti-human CPSF73 was raised against the N-terminal peptide NH_2-MSAIPAEESDQLLIRPLGAGQES-COOH and affinity purified. Rabbit anti-CTD Ser2-PO_4 was raised against a phosphorylated peptide with two heptad repeats. Anti-Sm, anti-CstF64 and anti-CBC were gifts from T. Blumenthal and I. Mattaj.

ChIP. HCT116 human colon cancer cells (p53 positive) were treated with DRB (50 μM, Sigma) or Nutlin-3a (10 μM, Calbiochem) for 8 h before being subjected to ChIP as described[36] (**Supplementary Methods**). In all cases, data from side-by-side cultures with and without DRB are shown. High resolution was achieved by shearing chromatin into approximately 250-bp fragments.

Real-time PCR. PCR reactions were performed with SybrGreen using the Roche LC-480 apparatus (Roche Applied Science) as described in **Supplementary Methods**. Primer sequences for *p21* have been described[36], and additional primers are listed in **Supplementary Table 3** online. To generate 5′–3′ profiles of occupancy, ChIP values were normalized relative to the amplicon with the highest fluorescence value for each gene. Where untreated and Nutlin-3a- or DRB-treated samples were compared, the highest DNA value between the two sets of data was used as the normalization point. Note that where two peaks of approximately equal ChIP signals were detected, neither is exactly 100% after normalization and taking the mean of multiple experiments. Points were then averaged and the s.e.m. calculated for each primer set. Each PCR determination was made on an independent plate relative to a standard curve generated on the same plate. *n* values refer to the number of PCR determinations from at least three independent immunoprecipitations.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

This work was supported by US National Institutes of Health (NIH) grants GM063873 and GM58613 to D.L.B. and CA117907 to J.E. K.G.-C. was supported by NIH fellowship 5F31 GM072099 and S.K. by F32 GM076951. We thank I. Mattaj (EMBL, Heidelberg) and T. Blumenthal (University of Colorado, Boulder) for antibodies. We also thank T. Blumenthal, J. Jaehning, R. Davis, T. Evans, N. Gomes, G. Bjerke, G. Bilousova, B. Erickson and members of the Bentley and Espinosa labs for their help.

Published online at http://www.nature.com/nsmb/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Bauren, G., Belikov, S. & Wieslander, L. Transcriptional termination in the Balbiani ring 1 gene is closely coupled to 3′-end formation and excision of the 3′-terminal intron. *Genes Dev.* **12**, 2759–2769 (1998).
2. Bentley, D.L. Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. *Curr. Opin. Cell Biol.* **17**, 251–256 (2005).
3. Maniatis, T. & Reed, R. An extensive network of coupling among gene expression machines. *Nature* **416**, 499–506 (2002).
4. Hirose, Y. & Manley, J.L. RNA polymerase II and the integration of nuclear events. *Genes Dev.* **14**, 1415–1429 (2000).
5. Phatnani, H.P. & Greenleaf, A.L. Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.* **20**, 2922–2936 (2006).
6. Greenleaf, A.L. Positive patches and negative noodles: linking RNA processing to transcription? *Trends Biochem. Sci.* **18**, 117–119 (1993).
7. Komarnitsky, P., Cho, E.J. & Buratowski, S. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev.* **14**, 2452–2460 (2000).
8. Licatalosi, D.D. *et al.* Functional interaction of yeast pre-mRNA 3′ end processing factors with RNA polymerase II. *Mol. Cell* **9**, 1101–1111 (2002).
9. Listerman, I., Sapra, A.K. & Neugebauer, K.M. Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nat. Struct. Mol. Biol.* **13**, 815–822 (2006).
10. Swinburne, I.A., Meyer, C.A., Liu, X.S., Silver, P.A. & Brodsky, A.S. Genomic localization of RNA binding proteins reveals links between pre-mRNA processing and transcription. *Genome Res.* **16**, 912–921 (2006).
11. Venkataraman, K., Brown, K.M. & Gilmartin, G.M. Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. *Genes Dev.* **19**, 1315–1327 (2005).
12. Zhang, Z. & Gilmour, D.S. Pcf11 is a termination factor in *Drosophila* that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. *Mol. Cell* **21**, 65–74 (2006).
13. Furuichi, Y. & Shatkin, A.J. Viral and cellular mRNA capping: past and prospects. *Adv. Virus Res.* **55**, 135–184 (2000).
14. Coppola, J.A., Field, A.S. & Luse, D.S. Promoter-proximal pausing by RNA polymerase II *in vitro*: transcripts shorter than 20 nucleotides are not capped. *Proc. Natl. Acad. Sci. USA* **80**, 1251–1255 (1983).
15. Rasmussen, E.B. & Lis, J.T. *In vivo* transcriptional pausing and cap formation on 3 *Drosophila* heat-shock genes. *Proc. Natl. Acad. Sci. USA* **90**, 7923–7927 (1993).
16. Moteki, S. & Price, D. Functional coupling of capping and transcription of mRNA. *Mol. Cell* **10**, 599–609 (2002).
17. Chiu, Y.L. *et al.* Tat stimulates cotranscriptional capping of HIV mRNA. *Mol. Cell* **10**, 585–597 (2002).
18. Shuman, S. Origins of mRNA identity: capping enzymes bind to the phosphorylated C-terminal domain of RNA polymerase II. *Proc. Natl. Acad. Sci. USA* **94**, 12758–12760 (1997).
19. Pei, Y. & Shuman, S. Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5. *J. Biol. Chem.* **277**, 19639–19648 (2002).
20. Peterlin, B.M. & Price, D.H. Controlling the elongation phase of transcription with P-TEFb. *Mol. Cell* **23**, 297–305 (2006).
21. Saunders, A., Core, L.J. & Lis, J.T. Breaking barriers to transcription elongation. *Nat. Rev. Mol. Cell Biol.* **7**, 557–567 (2006).
22. Cheng, C. & Sharp, P.A. RNA polymerase II accumulation in the promoter-proximal region of the dihydrofolate reductase and γ-actin genes. *Mol. Cell. Biol.* **23**, 1961–1967 (2003).
23. Wen, Y. & Shatkin, A.J. Transcription elongation factor hSPT5 stimulates mRNA capping. *Genes Dev.* **13**, 1774–1779 (1999).
24. Mandal, S.S. *et al.* Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. *Proc. Natl. Acad. Sci. USA* **101**, 7572–7577 (2004).

25. Zhao, J., Hyman, L. & Moore, C. Formation of mRNA 3′ ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol. Mol. Biol. Rev.* **63**, 405–445 (1999).
26. Dominski, Z., Yang, X.C. & Marzluff, W.F. The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing. *Cell* **123**, 37–48 (2005).
27. Kolev, N.G. & Steitz, J.A. Symplekin and multiple other polyadenylation factors participate in 3′-end maturation of histone mRNAs. *Genes Dev.* **19**, 2583–2592 (2005).
28. Baillat, D. *et al.* Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* **123**, 265–276 (2005).
29. Mandel, C.R. *et al.* Polyadenylation factor CPSF-73 is the pre-mRNA 3′-end-processing endonuclease. *Nature* **444**, 953–956 (2006).
30. Nag, A., Narsinh, K. & Martinson, H.G. The poly(A)-dependent transcriptional pause is mediated by CPSF acting on the body of the polymerase. *Nat. Struct. Mol. Biol.* **14**, 662–669 (2007).
31. Dantonel, J.C., Murthy, K.G., Manley, J.L. & Tora, L. Transcription factor TFIID recruits factor CPSF for formation of 3′ end of mRNA. *Nature* **389**, 399–402 (1997).
32. Nag, A., Narsinh, K., Kazerouninia, A. & Martinson, H.G. The conserved AAUAAA hexamer of the poly(A) signal can act alone to trigger a stable decrease in RNA polymerase II transcription velocity. *RNA* **12**, 1534–1544 (2006).
33. Gromak, N., West, S. & Proudfoot, N.J. Pause sites promote transcriptional termination of mammalian RNA polymerase II. *Mol. Cell. Biol.* **26**, 3986–3896 (2006).
34. Neuman de Vegvar, H., Lund, E. & Dahlberg, J.E. 3′ end formation of U1 snRNA precursors is coupled to transcription from snRNA promoters. *Cell* **47**, 259–266 (1986).
35. Hernandez, N. & Weiner, A.M. Formation of the 3′ end of U1 snRNA requires compatible snRNA promoter elements. *Cell* **47**, 249–258 (1986).
36. Gomes, N.P. *et al.* Gene-specific requirement for P-TEFb activity and RNA polymerase II phosphorylation within the p53 transcriptional program. *Genes Dev.* **20**, 601–612 (2006).
37. Workman, J.L. Nucleosome displacement in transcription. *Genes Dev.* **20**, 2009–2017 (2006).
38. 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).
39. Bird, G., Zorio, D.A. & Bentley, D.L. RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3′-end formation. *Mol. Cell. Biol.* **24**, 8963–8969 (2004).
40. Vos, J.C., Sasker, M. & Stunnenberg, H.G. Vaccinia virus capping enzyme is a transcription initiation factor. *EMBO J.* **10**, 2553–2558 (1991).
41. Hagler, J., Luo, Y. & Shuman, S. Factor-dependent transcription termination by vaccinia RNA polymerase. Kinetic coupling and requirement for ATP hydrolysis. *J. Biol. Chem.* **269**, 10050–10060 (1994).
42. Adamson, T.E., Shutt, D.C. & Price, D.H. Functional coupling of cleavage and polyadenylation with transcription of mRNA. *J. Biol. Chem.* **280**, 32262–32271 (2005).
43. Schroeder, S., Schwer, B., Shuman, S. & Bentley, D. Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes Dev.* **14**, 2435–2440 (2000).
44. Lacadie, S.A. & Rosbash, M. Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5′ss base pairing in yeast. *Mol. Cell* **19**, 65–75 (2005).
45. Yonaha, M. & Proudfoot, N.J. Specific transcriptional pausing activates polyadenylation in a coupled *in vitro* system. *Mol. Cell* **3**, 593–600 (1999).
46. Kaneko, S., Rozenblatt-Rosen, O., Meyerson, M. & Manley, J.L. The multifunctional protein p54nrb/PSF recruits the exonuclease XRN2 to facilitate pre-mRNA 3′ processing and transcription termination. *Genes Dev.* **21**, 1779–1789 (2007).
47. Darzacq, X. *et al. In vivo* dynamics of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **14**, 796–806 (2007).
48. Rosonina, E., Kaneko, S. & Manley, J.L. Terminating the transcript: breaking up is hard to do. *Genes Dev.* **20**, 1050–1056 (2006).
49. Osheim, Y.N., Proudfoot, N.J. & Beyer, A.L. EM visualization of transcription by RNA polymerase II: downstream termination requires a poly(A) signal but not transcript cleavage. *Mol. Cell* **3**, 379–387 (1999).
50. Rigo, F., Kazerouninia, A., Nag, A. & Martinson, H.G. The RNA tether from the poly(A) signal to the polymerase mediates coupling of transcription to cleavage and polyadenylation. *Mol. Cell* **20**, 733–745 (2005).
51. Bird, G., Fong, N., Gatlin, J.C., Farabaugh, S. & Bentley, D.L. Ribozyme cleavage reveals connections between mRNA release from the site of transcription and pre-mRNA processing. *Mol. Cell* **20**, 747–758 (2005).