

# AT-rich sequence elements promote nascent transcript cleavage leading to RNA polymerase II termination

Eleanor White, Kinga Kamieniarz-Gdula, Michael J. Dye\* and Nick J. Proudfoot\*

Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK

Received August 3, 2012; Revised November 27, 2012; Accepted November 29, 2012

## ABSTRACT

**RNA Polymerase II (Pol II) termination is dependent on RNA processing signals as well as specific terminator elements located downstream of the poly(A) site. One of the two major terminator classes described so far is the Co-Transcriptional Cleavage (CoTC) element. We show that homopolymer A/T tracts within the human  $\beta$ -globin CoTC-mediated terminator element play a critical role in Pol II termination. These short A/T tracts, dispersed within seemingly random sequences, are strong terminator elements, and bioinformatics analysis confirms the presence of such sequences in 70% of the putative terminator regions (PTRs) genome-wide.**

## INTRODUCTION

Transcription termination is of vital importance to the cell. It prevents interference with transcription of downstream genes (1), and is thought to allow recycling of RNA polymerase II (Pol II) for new rounds of transcription. Indeed, it has been demonstrated that termination increases transcription re-initiation by enhancing recruitment of Pol II and associated general transcription factors to the promoter (2). The structural basis of this enhancement process likely relates to the dynamic interactions between promoter and terminator regions of Pol II-transcribed genes (3,4). Transcriptional termination has also been demonstrated to elevate mRNA and protein levels by enhancing splicing and preventing degradation of the nascent transcript (5). In recent years, it has been shown that with some genes, transcriptional read-through due to inefficient termination may generate antisense RNAs that can induce heterochromatin formation, which in turn restricts gene expression (6,7).

Although for some mammalian genes, efficient termination requires solely the transcription of a functional polyA signal (PAS), others possess additional termination elements, located downstream of the PAS. Two major terminator classes, pause sites and Co-Transcriptional Cleavage (CoTC) elements, have been described. With pause element terminators, PAS cleavage creates an entry site for the 5'-3' exonuclease, Xrn2, to degrade the 3' nascent RNA leading to termination. It is thought that pause elements enhance the termination process by slowing down transcription elongation, which allows Xrn2 more time to degrade the nascent RNA, effectively 'catching up' with transcribing Pol II (8,9). CoTC-mediated termination similarly requires Pol II to become termination competent by transcribing a functional PAS. However, in this case, the cleavage event that creates an entry site for Xrn2 occurs at the CoTC element rather than the upstream PAS. Degradation of the 3' nascent RNA leads to release of Pol II from the chromatin template with the associated pre-mRNA. Subsequently, PAS cleavage leads to release of the mRNA from Pol II (10).

This study describes a detailed analysis of the minimal sequence elements required to mediate CoTC-dependent Pol II termination. Strikingly, dispersed tracts of AT-rich sequence appear to provide optimal sequence parameters to allow efficient CoTC termination.

## MATERIALS AND METHODS

### Primers

See Supplementary Table S1 for all primer sequences used in this study.

### Cell culture analysis

Subconfluent, 10-cm-diameter plates of HeLa cells were transfected with 10  $\mu$ g of  $\beta$ -globin and 1  $\mu$ g of pTat plasmids using Lipofectamine 2000 (Invitrogen) following

\*To whom correspondence should be addressed. Tel: +44 1865 275 568; Email: michael.dye@path.ox.ac.uk  
Correspondence may also be addressed to Nick J. Proudfoot. Tel: +44 1865 275 566; Fax: +44 1865 275 556; Email: nicholas.proudfoot@path.ox.ac.uk

the manufacturer's guidelines. Plasmid, and stable cell line, construction is described in Supplementary Materials and Methods. At 24 or 48 hr after transfection, cells were harvested for RNA or protein isolation respectively. The RNAi procedure was carried out as described (11) (Supplementary Materials and Methods).

### RNA analysis

One microgram of RNA was reverse transcribed with Superscript III Reverse Transcriptase (Stratagene) followed by quantitative PCR with a QuantiTect SYBR green kit (Qiagen) on a Corbett Rotor Gene 3000 machine using primers listed (Supplementary Table S1). The RPA protocol has been described (12) as has the HIV-LTR riboprobe (13). The NRO protocol and M13 probes B3 and B4 were described in (14), P and U3 in (15), and A in (16). The nuclear fractionation protocol has been described in (17). Northern Blot and hscRACE analysis are described in Supplementary Materials and Methods.

### Protein analysis

Twenty microgram of HeLa protein lysate was separated on an SDS-polyacrylamide gel and transferred to Hybond-C extra nitrocellulose membrane using a semi-dry Transblot apparatus. Membranes were probed with anti-human  $\beta$ -globin (Santa Cruz Biotechnology) at 1:1000, and anti-actin (Sigma) at 1:1000. Secondary antibodies were anti-mouse (Sigma) at 1:2000 or anti-rabbit (Sigma) at 1:2000. Signals were detected with an ECL kit (GE Healthcare) and quantitated using ImageQuant software.

### Computational analysis

Motif discovery was performed using MEME [Multiple Em for Motif Elicitation; (18)]. For the analysis in Supplementary Figure S3, p $\beta$ T<sup>+</sup>, p $\beta$ Ta->t, p $\beta$ Tcon, T8 and T9 were used as input sequences, and p $\beta$ Ta->c, p $\beta$ Trev and p $\beta$ a30 as negative sequences (additional parameters: -nmotifs 3 -minw 6 -maxw 300 -revcomp). PTR definition and pattern searching was performed using custom R scripts and Bioconductor Biostrings and BSgenome packages (19), and based on RefGene hg18 annotation obtained from UCSC genome browser (20).

## RESULTS

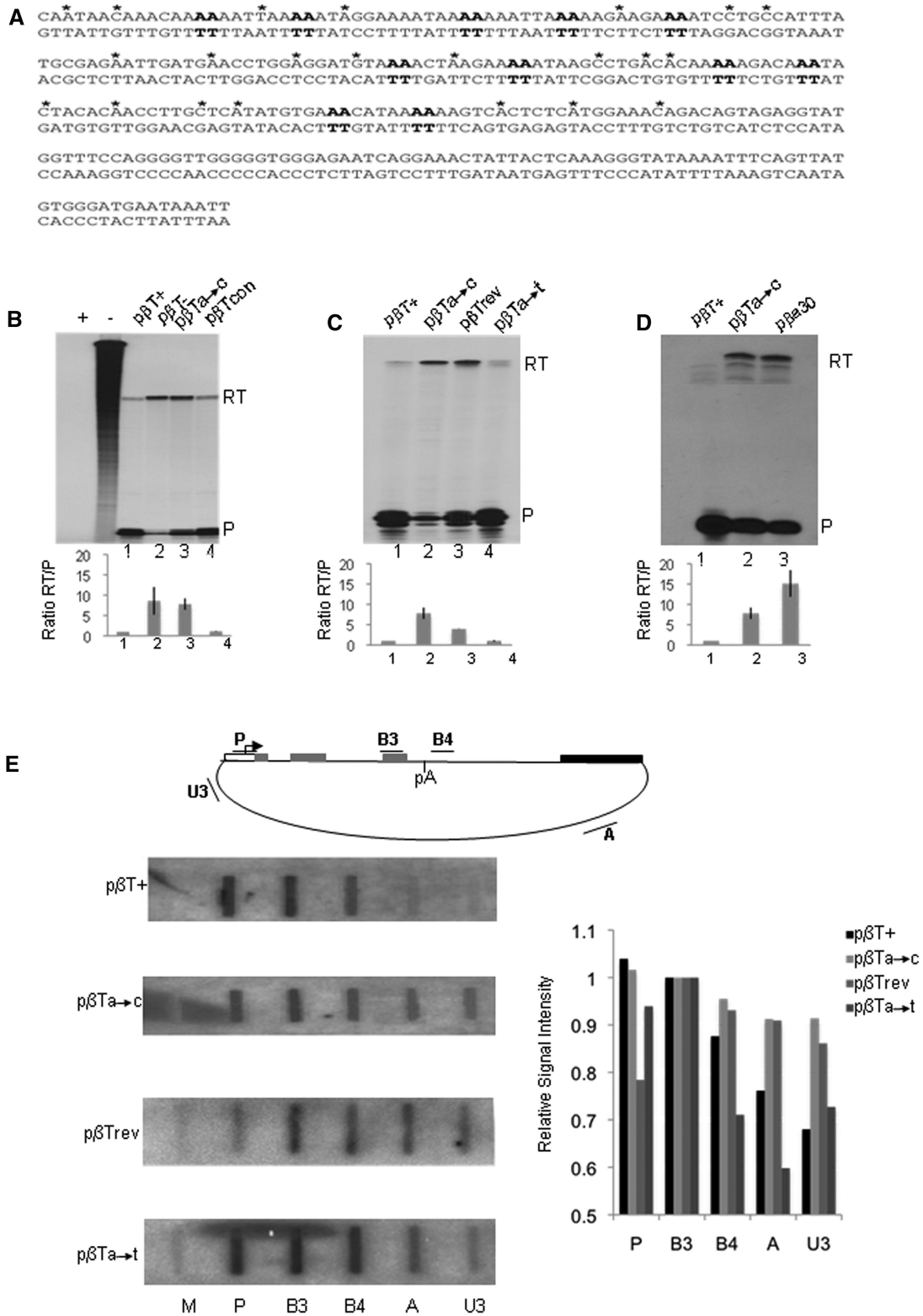
### Dispersed oligoA or T tracts promote transcriptional termination

To investigate the sequence dependence of termination elements for Pol II-transcribed genes in mammals, we have studied the well-characterized termination element of the human  $\beta$ -globin gene. This comprises an approximately 850 base pair (bp) AT-rich sequence element located 1 kb downstream of the PAS in the  $\beta$ -globin 3' flanking region. This element has been divided into three parts ( $\approx$ 300 bp), each capable of directing independent termination (21). Here we analyse the most potent 3' positioned terminator element (referred to as the Terminator).

To define a minimal element required for efficient transcription termination, the Terminator was subdivided into four sections labelled A, B, C and D. These were cloned into a  $\beta$ -globin construct, p $\beta$  $\Delta$ 5–10 (21) (p $\beta$ T<sup>-</sup>), which contains the  $\beta$ -globin gene driven by the HIV-LTR and 240 bp 3' flanking sequence beyond the PAS, which lacks terminator activity. p $\beta$ T<sup>+</sup> contains the  $\beta$ -globin gene, 240 bp of post PAS and the whole Terminator element. An additional construct p $\beta$  $\Delta$ 8–10 was used containing the  $\beta$ -globin gene with approximately 800 bp of 3' flanking region sequence, which lacks terminator activity. HeLa cells were transiently transfected with constructs p $\beta$ A, p $\beta$ AB and p $\beta$ ABC, as well as p $\beta$ T<sup>+</sup> as a positive control and p $\beta$ T<sup>-</sup> and p $\beta$  $\Delta$ 8–10, as negative controls (Supplementary Figure S1). In each case, a plasmid expressing the transactivator Tat (pTat) was co-transfected to enable efficient Pol II initiation/elongation from the HIV-LTR (22). The transcription termination capacity of each sequence was measured by an RNase protection read-through assay using nuclear RNA (RPA-RT) (13). A uniformly radiolabelled antisense RNA probe that spans the HIV-LTR yields a short protected RNA (P, 85 nt), derived from the 5'-end of mRNA as well as a larger read-through product (RT, 240 nt) from transcripts reading through the gene and around the plasmid backbone, back into the HIV-LTR (Supplementary Figure S1). The RT product is relatively unstable and nuclear restricted (23). Consequently, RNA levels are determined by a combination of RNA synthesis and degradation. However, the ratio of RT/P reflects transcription termination efficiency. As demonstrated (Supplementary Figure S1), the RT/P ratio decreases when the Terminator element increases in length, suggesting that the entire Terminator is required for efficient termination. Note that there is also a spacing effect, as p $\beta$  $\Delta$ 8–10 generates less read-through than p $\beta$ T<sup>-</sup>.

Inspection of the Terminator sequence revealed a number of oligoA tracts dispersed throughout the Terminator, so we next examined their possible role in directing mammalian Pol II termination. Two constructs were engineered: p $\beta$ Ta->c, in which A tracts were disrupted by substitution of 11 double A with double C residues (Figure 1A), and p $\beta$ Tcon, where 22 point mutations were introduced into the Terminator at locations other than the A tracts. Using the RPA-RT assay, we find that disruption of the oligoA tracts in p $\beta$ Ta->c causes a significant decrease in Terminator efficiency (Figure 1B, lane 3), whereas, with the control construct (p $\beta$ Tcon), termination efficiency was unaffected (Figure 1B, lane 4). These results were repeated multiple times and the RT/P ratios are presented graphically (Figure 1B). This analysis shows that A tracts in the Terminator sequence are required for efficient Pol II termination. However, this approach does not exclude the possibility that a low C content is required.

We next tested whether the Terminator efficiency is dependent on oligoA sequence or simply AT richness. Two additional constructs were made: p $\beta$ Trev has the Terminator element in reverse orientation, while p $\beta$ Ta->t has the 11 double As mutated to double Ts in the oligo A tracts. We show that upon reversal of the



**Figure 1.** Dispersed oligoA and T tracts promote transcriptional termination. (A) The human  $\beta$ -globin Terminator sequence with the location of AA to CC or TT doublet mutations in  $p\beta Ta \rightarrow c$  and  $p\beta Ta \rightarrow t$ , respectively, highlighted in bold. \* denotes point mutations (A  $\rightarrow$  N, C  $\rightarrow$  N, T  $\rightarrow$  N) in  $p\beta Tcon$ , which mutates non-oligoA tracts. (B) RNase protection read-through assay (RPA-RT). Nuclear transcripts were hybridized to a radiolabelled antisense RNA probe (HIV-LTR) spanning the transcription start site of the  $\beta$ -globin constructs. Band P (85 nt) results from hybridization of correctly initiated transcripts while band RT (240 nt) derives from read-through transcripts. RNA was from HeLa cells transiently transfected with  $p\beta Tc \rightarrow a$  and  $p\beta Tcon$  constructs.  $p\beta T+$  was used as a positive control and  $p\beta T-$  as a negative control. Radioactive signals were quantitated using PhosphoImager, the RT/P ratio calculated and corrected to the  $p\beta T+$  positive control, set at 1. Quantitation of these data is

(continued)

Terminator element (p $\beta$ Trev), the RT/P ratio increases, which is indicative of a transcription termination defect (Figure 1C, lane 3). However, with p $\beta$ Ta->t, the RT/P ratio remains relatively unaffected, indicating that mutation of A to T in the oligoA tracts maintains termination activity (Figure 1C, lane 4). Again, these results were repeated multiple times and quantitated graphically (Figure 1C). Next the possibility that an A-rich sequence may be sufficient to promote Pol II termination was considered. A new construct, p $\beta$ a30, in which the Terminator was substituted by 30 A residues, was analysed by RPA-RT assay (Figure 1D). As shown, the ratio of RT/P for p $\beta$ a30 (lane 3) is relatively high indicative of low termination efficiency (as in p $\beta$ Ta->c; lane 2). This result shows that a continuous sequence of 30 A residues is insufficient to terminate transcription. In effect, these Terminator mutants show that dispersed AT tracts are required for CoTC-mediated termination, suggesting that the strength of the RNA:DNA hybrid in the Terminator is critical for activity. Furthermore, the AT tracts do not appear to be sufficient for efficient termination, which could imply that they must be in the correct context to be functional.

These results were confirmed by a more direct termination assay. Nuclear run on (NRO) analysis was performed on nuclei isolated from HeLa cells transiently transfected with p $\beta$ T+, p $\beta$ Ta->c, p $\beta$ Trev or p $\beta$ Ta->t constructs and pTat. NRO analysis measures nascent transcripts generated by actively transcribing Pol II. Nascent transcripts were detected by single-stranded DNA probes homologous to  $\beta$ -globin gene sequence in the transfected plasmid, including the promoter (P),  $\beta$ -globin exon 3 (B3) and sequence immediately downstream of the PAS (B4). Antisense probes for the vector backbone (A), and for U3 of the HIV-LTR (U3), were used to detect transcriptional read-through. Nascent transcripts were  $\alpha^{32}$ P-UTP labelled, partially hydrolyzed and then hybridized to single-stranded DNA probes (Figure 1E). As expected, NRO signals in the p $\beta$ T+ experiment were relatively high over probes P, B3 and B4 and low over probes A and U3, indicating efficient termination. In contrast, near equal NRO signals were detected throughout the plasmid in p $\beta$ Ta->c. The increased signals over A and U3 show that mutation of the Terminator A tracts (from A to C) blocks termination resulting in Pol II transcriptional read-through. A similar profile was obtained for p $\beta$ Trev, confirming that the Terminator does not function in the

reverse orientation. Interestingly, with p $\beta$ Ta->t, signals over probes P, B3 and B4 were high but significantly lower over probes A and U3 (Figure 1E), confirming that termination is largely maintained when oligoA tracts are changed to T residues.

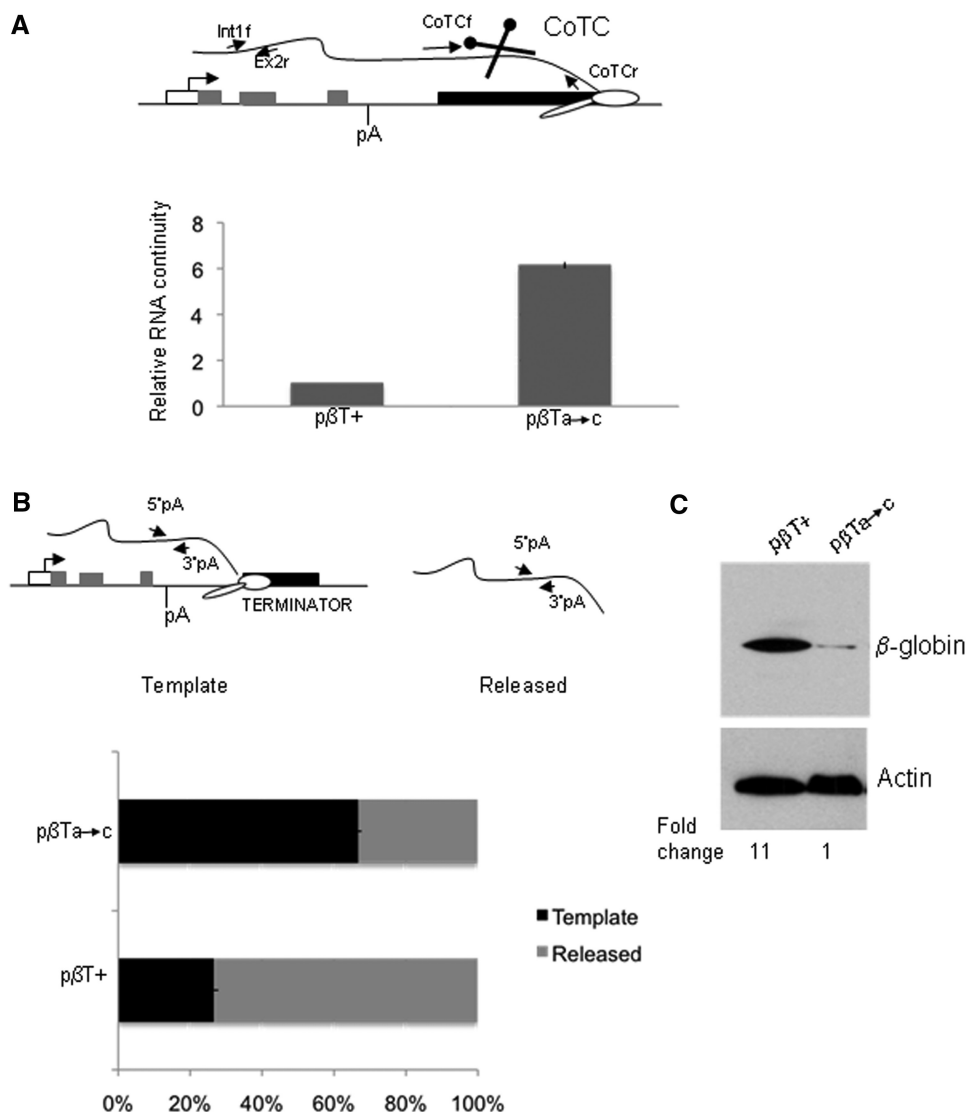
### OligoA tracts are required for the CoTC activity of the human $\beta$ -globin terminator

It has previously been shown that the transcriptional termination capacity of the  $\beta$ -globin terminator is dependent on CoTC of the nascent Terminator transcript (21). To test whether the doublet mutations of the Terminator A tracts introduced in p $\beta$ Ta->c affect CoTC activity, a quantitative reverse transcription PCR (qRT-PCR) assay was carried out. The relative level of qRT-PCR product obtained over the Terminator region gives an indication of the extent of transcript cleavage. This value was normalized to nascent transcription over the body of the gene (Figure 2A). As indicated, disruption of Terminator A tracts causes transcript cleavage to decrease by approximately 6-fold (higher RNA continuity). This experiment demonstrates that CoTC activity is significantly reduced by disruption of the homopolymer tracts.

West *et al.* (2008) (10) reported that termination by CoTC cleavage results in release of Pol II and associated pre-mRNA from the chromatin template into the nucleoplasm. Effectively, CoTC mediates termination prior to pre-mRNA 3'-end processing and enhances gene expression, resulting in elevated mRNA and protein levels (5). To further analyse the effect of p $\beta$ Ta->c mutation on CoTC activity and termination, we measured the distribution of pre-mRNA from p $\beta$ T+ and p $\beta$ Ta->c in chromatin and nucleoplasm fractions. Transfected HeLa cell nuclei were lysed and centrifuged to pellet pre-mRNA still linked to chromatin (template fraction; 24,25). The remaining nuclear supernatant is termed the released fraction. RNA isolated from each fraction was then subjected to qRT-PCR analysis to measure transcript levels not yet cleaved at the  $\beta$ -globin PAS (see Figure 2B for primer positions). The level of qRT-PCR product obtained from the template and released fraction is expressed as a percentage of the total PCR product in both fractions. As demonstrated in p $\beta$ T+, 22% uncleaved pre-mRNA was present in the template fraction and the remaining 78% was in the released fraction (Figure 2B). These results confirm that CoTC-mediated termination promotes Pol II release from chromatin prior to pre-mRNA 3'-end processing. In contrast, p $\beta$ Ta->c generated increased levels of uncleaved pre-mRNA in the template fraction (62%) (Figure 2B). This shows that disruption of the oligoA tracts reduces release of pre-mRNA into the nucleoplasm. Since CoTC-mediated release of pre-mRNA into the nucleoplasm enhances gene expression (5), we also examined whether mutating the oligoA tracts of the human  $\beta$ -globin gene Terminator influenced the levels of  $\beta$ -globin protein produced. Western Blot analysis (Figure 2C) revealed that with p $\beta$ Ta->c, dramatically less (ten-fold) protein is synthesised as compared with p $\beta$ T+. These results confirm that oligoA tracts in the Terminator sequence are required for efficient transcriptional termination.

#### Figure 1. Continued

presented graphically as a ratio of RT/P. Error bars denote standard error of mean. (C) RPA-RT analysis of p $\beta$ Trev and p $\beta$ Ta->t as in (B). p $\beta$ T+ and p $\beta$ Ta->c were used as positive and negative controls. (D) RPA-RT analysis of p $\beta$ a30 RNA as in (B). (E) Top panel: Diagram of the  $\beta$ -globin plasmid. The start site of transcription (arrow) and regions of complementarity of antisense ssDNA M13 probes are indicated (bold). Bottom left panel: NRO analyses of nascent transcripts of p $\beta$ T+, p $\beta$ Ta->c, p $\beta$ Trev and p $\beta$ Ta->t constructs. NRO probes are indicated below their respective positions on the filters. M represents empty M13 vector and is a background control. Bottom right panel: NRO signals were quantified by PhosphoImager and normalized to probe B3, taken as 1.

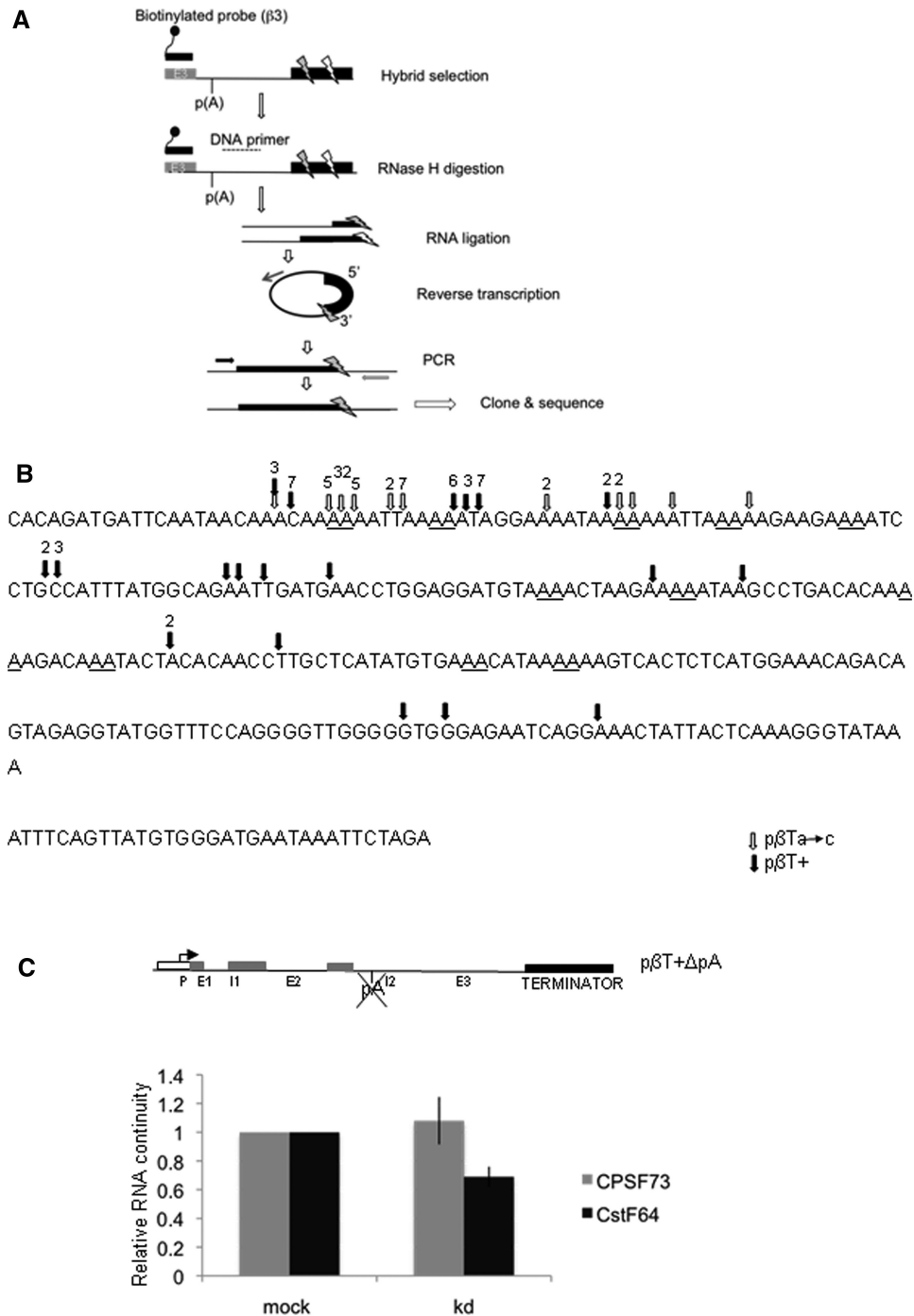


**Figure 2.** OligoA tracts define the CoTC activity of the human  $\beta$ -globin Terminator. (A) Top panel: Diagram showing the position of the PCR amplicons, and primers used for RT-PCR. PCR product detected over the Terminator region (CoTC) is indicative of the level of transcripts not cleaved over that region. Lower panel: qRT-PCR analysis of total RNA isolated from HeLa cells transiently transfected with p $\beta$ T+ or p $\beta$ Ta->c. The ratio of Terminator to intron1/exon2 RNA from p $\beta$ Ta->c is normalized to p $\beta$ T+ (referred to as Relative RNA continuity). Error bars denote standard deviation. (B) Top panel: Diagram showing the position of the PCR amplicons, and primers used for RT-PCR. In the nuclear fractionation procedure, pre-mRNA (curved black line), attached to the DNA template (black line) via Pol II (empty icon), are found in the template fraction whereas nucleoplasmic pre-mRNA, not attached to the template, is found in the released fraction. Bottom panel: qRT-PCR analysis of p $\beta$ T+ or p $\beta$ Ta->c RNA. Levels of pre-mRNA detected in template and released fractions for p $\beta$ T+ and p $\beta$ Ta->c are expressed as percentage of total pre-mRNA in the sample. Error bars denote standard deviation. (C) Western blot analysis of HeLa cells transfected with p $\beta$ T+ or p $\beta$ Ta->c.  $\beta$ -globin and actin proteins are indicated. Level of  $\beta$ -globin from p $\beta$ T+ is normalized to p $\beta$ Ta->c.

### OligoA mutations change the distribution of cleavage sites over the human $\beta$ -globin terminator transcript

To determine the location and distribution of CoTC cleavage sites in the wild-type Terminator transcript, and to analyse the effect of homopolymer tract disruption on the distribution of these sites, we next performed hybrid selection circular (hsc)RACE as described in Figure 3A (26). Nuclear RNA isolated from cells transfected with p $\beta$ T+ or p $\beta$ Ta->c was subjected to hscRACE, and the 3'-ends of these selected RNAs were identified by

sequencing (Figure 3B). The various species detected are likely to correspond to a combination of primary sites of transcript cleavage, degradation intermediates or possibly positions of Pol II pausing. The ends identified in the wild-type Terminator were scattered throughout the Terminator region, with a hotspot of activity found at the 5'-end of the Terminator. The RNA ends from the mutant Terminator (p $\beta$ Ta->c) displayed a strikingly different distribution. Rather than being spread out across the region, they are concentrated at the 5'-end of the Terminator. It is evident from comparing the ends



**Figure 3.** OligoA mutations change CoTC cleavage distribution. (A) The hscRACE procedure. Terminator element (black box), exon 3 (grey box) and cleavage sites (lightning bolts) are shown. Biotinylated RNA probe,  $\beta 3$  (tailed box), was hybridized to nuclear RNA isolated from HeLa cells transfected with  $\beta T+$  or  $\beta T\alpha \rightarrow c$  and then isolated by magnetic selection. Selected RNA is subjected to oligonucleotide (dotted line)-directed RNase H digestion releasing RNA from beads. This is ligated and reverse transcribed across the ligated junction. cDNA is then amplified by PCR using gene-specific primers. The PCR products are analysed on a gel, cloned and sequenced. (B) Terminator DNA sequence showing the positions of CoTC-derived RNA 3'-ends in  $\beta T+$  (black filled arrows) and  $\beta T\alpha \rightarrow c$  (empty arrows). Positions of AA to CC mutations in  $\beta T\alpha \rightarrow c$  are underlined. The numbers above each arrow indicate the number of clones identified for each site. The hscRACE experiment was performed three and seven times for  $\beta T+$  and  $\beta T\alpha \rightarrow c$ , respectively, in order to obtain fifty positive colonies for each construct. (C) Effect of siRNA-mediated knock down of CPSF73 and CstF64.  $\beta T+\Delta pA$  has a mutated PAS (AATAAA  $\rightarrow$  GAATTC), known to inactivate  $\beta$ -globin mRNA polyadenylation (Dye and Proudfoot, 1999). This was transfected into CPSF73 and CstF64 siRNA-treated and mock-treated cells. Relative RNA continuity (based on RT/PCR analysis as in Figure 2A) was assessed and is presented graphically. Error bars denote standard deviation.

produced by  $\beta T^+$  and  $\beta Ta \rightarrow c$  that mutation of the A tracts prevents cleavage in the 3' portion of the Terminator transcript (Figure 3B). Indeed, the 5'-end of the Terminator is the most A-rich region of the Terminator, even with the mutated terminator sequence. These results reinforce the view that the A tracts are important in termination.

We then considered the possibility that RNA cleavage might be mediated by the 3'-end cleavage/polyA factor complex. In particular, since many mRNA encoding genes possess multiple PAS (27), we reasoned that CoTC elements might be equivalent to distal PAS. However, our hscRACE data (Figure 3B) did not identify CoTC 3'-ends with added oligoA or polyA tracts. We further tested for the potential role of 3'-end cleavage factors CPSF-73 and CstF-64 in CoTC cleavage. The  $\beta T^+$  construct was modified by mutation of the major  $\beta$ -globin PAS, and, following transfection into HeLa cells treated with siRNAs against CPSF-73 or CstF-64 mRNA, was subjected to qRT/PCR analysis to measure CoTC cleavage (Figure 3C). Notably, knock down of CPSF73 and CstF64 by RNAi reduces  $\beta$ -actin PAS cleavage by 2- and 1.3-fold, respectively (Supplementary Figure S2), but does not significantly diminish  $\beta$ -globin CoTC cleavage activity. Unexpectedly, CstF64 knock down appears to enhance CoTC activity, suggesting that there is a competition effect where at weaker or ill-defined PAS, the CoTC activity predominates. These results confirm that 3'-end cleavage/polyA factors are not associated with CoTC cleavage.

#### Analysis of stable $\beta$ -globin integrants

The effect of oligoA tract disruption on transcriptional termination has, so far, only been analysed in a plasmid-based system. Thus, we elected to confirm that oligoA mutations affect transcription termination when placed in a chromosomal context. We used the HEK293 cells and the Flp-In T-Rex recombinase system to create stable cell lines. Two cell lines were made that contain the  $\beta$ -globin gene driven by the HIV-LTR, 240 bp of post PAS sequence and either the wild-type Terminator or the mutant Terminator from the  $\beta Ta \rightarrow c$  construct [referred to as wild-type (wt) or mutant (mut) stable cell lines, respectively]. These stable cell lines were also transfected with pTat to allow expression from the HIV-LTR.

To confirm that the  $\beta$ -globin gene is expressed in the stable cell lines, we analysed RNA isolated from the wild-type and mutant stable cell lines before and after transcriptional activation with Tat, and also the parental HEK293 cell line, by Northern Blot.  $\beta$ -globin mRNA was detected, confirming that the  $\beta$ -globin gene is expressed in both the wild-type and mutant stable cell lines (Figure 4A, lanes 2 & 4). This analysis shows that the mutant cell line produces approximately 3-fold less  $\beta$ -globin mRNA than the wild type. Some  $\beta$ -globin mRNA was detected in the absence of Tat (lanes 3 & 5), indicating that the parental HEK293 cell line allows promiscuous HIV-LTR transcription initiation in the absence of Tat. The lack of a signal in the parental HEK293 cell line (lane 1) confirms that the endogenous  $\beta$ -globin gene is not expressed in these cells.

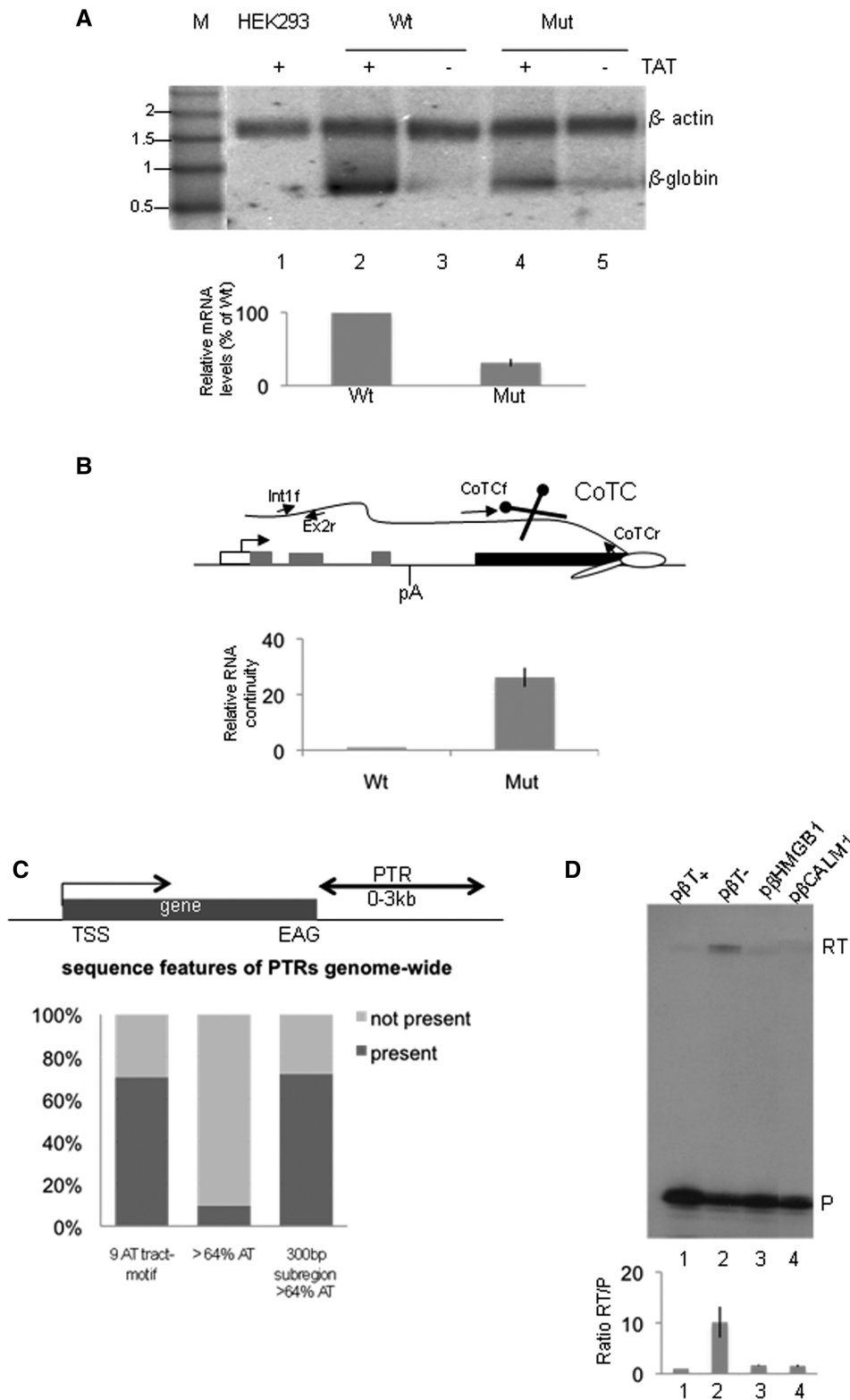
The main aim of producing these stable cell lines was to analyse the effects of CoTC oligoA tract mutations in a more physiological chromatin environment. Steady state RNA analysis shows a 3-fold decrease in  $\beta$ -globin gene expression with the mutated CoTC element. This is concordant with transient transfection experiments in West and Proudfoot (2009). We then used qRT-PCR as before (Figure 2) to analyse CoTC cleavage of the Terminator transcripts in the wild-type and mutant stable cell lines. We confirmed the results from the plasmid-based system by showing that the mutant Terminator is defective in CoTC activity (Figure 4B). Indeed, in this chromosomal environment, transcript cleavage is reduced by approximately 25-fold, a significantly higher effect than observed in the plasmid-based system. This shows that when the Terminator is placed in a chromosomal environment, the importance of the A-rich tracts for transcriptional termination is paramount.

#### Computational analysis of putative terminator regions genome-wide

Using the mutational analysis described in this study, we were able to define a set of functional human  $\beta$ -globin-derived terminators ( $\beta T^+$ ,  $\beta T^{con}$  and  $\beta Ta \rightarrow t$ ), as well as deficient ones ( $\beta Ta \rightarrow c$ ,  $\beta T^{rev}$  and  $\beta \alpha 30$ ), and used them to perform sequence analysis to identify motifs that could be important for CoTC element function. The positive set included additionally the two upstream  $\beta$ -globin terminator elements capable of directing independently CoTC-mediated transcription termination (called here T8 and T9, corresponding to regions 8 and 9 in Dye and Proudfoot, 2001).

Consistent with manual inspection, computational sequence analysis for pattern elicitation revealed that no simple conserved pattern was present within the CoTC terminator elements (Supplementary Figure S3 and data not shown). Instead, long, degenerate, AT-rich stretches were identified (motif 1–3, Supplementary Figure S3A–C). Simplification of the identified patterns leads to the regular expression  $[A/T_{5+}G/C_{1+N_{0-35}}]_8A/T_5$ , corresponding to nine dispersed AT tracts of minimum 5 nt length each separated by 1–36 nt containing at least one G or C residue (referred to as the 9 AT tracts-motif, Supplementary Figure S3C).

To estimate whether  $\beta$ -globin-CoTC-like elements have a more general occurrence, a genome-wide analysis searching for the identified patterns was performed. To this end, putative terminator regions (PTRs) were defined as regions 0–3 kb downstream from the end of annotated human RefSeq genes (Figure 4C). PTRs were further filtered so that they do not overlap any annotated gene, and unique regions were retained, resulting in 16 159 sequences. Pattern matching revealed that 70% of PTRs contain the 9 AT tracts-motif (Figure 4C). Another common feature of our set of studied terminators is an AT percentage of above 64% (on average, 67.72%). Ten per cent of PTRs are >64% AT rich, and more importantly 72% of PTRs contain a 300 bp subregion >64% AT rich (Figure 4C). Taken together these data support the view that CoTC-terminator elements may be widespread.



**Figure 4.** Dispersed oligoA tracts promote transcriptional termination in a chromosomal context, and are prevalent in PTRs. (A) Northern Blot analysis of nuclear RNA isolated from HEK293 cells containing stably integrated constructs with the Terminator elements from p $\beta$ T<sup>+</sup> and p $\beta$ Ta<sup>->c</sup> (wt and mut, respectively) and from the parental HEK293 line before the integration. Antisense RNA probe complimentary to exon 2 was used.  $\beta$ -actin was detected as a loading control. Levels of  $\beta$ -globin mRNA were normalized to levels of  $\beta$ -actin mRNA and corrected to the Wt, set at 100%. Error bars denote standard error of mean. (B) Top panel: Diagram showing the position of primers used in RT-PCR analysis. Lower panel (as in Figure 2A); The ratio of Terminator to intron1/exon2 RNA in mut is normalized to wt. Error bars denote standard deviation. (C) Top panel: Diagram showing the definition of a putative terminator region (PTR). TSS, transcription start site; EAG, end of annotated gene. Lower panel: Bar graph showing the genome-wide incidence of the identified CoTC-mediated terminator sequence motif and AT richness in PTRs. (D) RPA-RT analysis of p $\beta$ HMGB1 and p $\beta$ CALM1 as in Figure 1B. p $\beta$ T<sup>+</sup> and p $\beta$ T<sup>-</sup> were used as positive and negative controls.



To evaluate the significance of the 70% occurrence of the AT-tract motif in PTRs, we decided to investigate the prevalence of this motif in other genomic regions. Notably, 67% of equal size, extended promoter regions (defined as 3 kb upstream of TSS of annotated genes, non-overlapping any annotated gene) also contain this motif. Furthermore, when randomly selecting the same number and length of sequence from the human genome a 1000 times, 79% of such sequence contained this motif. As an additional control, we performed a similar analysis of the TATA box motif (TATAAAA). This motif is present in 37% of extended promoter regions, 40% of PTRs and 46% of random genomic sequences of the same number and length (average from 1000 examples). Overall, this analysis suggests that a well-studied sequence feature, such as the TATA box, is clearly common in the genome, yet, in the right context, can mediate transcription initiation. Likewise, we envisage that, the 9 AT tract motif is also common in the genome. However, when it is encountered in the correct context, such as by a termination-competent Pol II downstream of a functional PAS, then it can lead to transcription termination.

To confirm that PTRs containing the 9 AT tract motif can function as terminator elements, two such PTRs were cloned into the p $\beta$ T<sup>-</sup> construct, and their function assayed using the RPA-RT assay. The PTRs chosen (Supplementary Figure S4) are located 641 bp and 71 bp from the PAS of the HMGB1 and CALM1 genes respectively, and the constructs containing them are named p $\beta$ HMGB1 and p $\beta$ CALM1. The results of this analysis show that in both constructs, the RT/P ratio is similar to the RT/P ratio observed for p $\beta$ T<sup>+</sup> (Figure 4D). These results were repeated three times and the RT/P ratios are presented graphically (Figure 4D). The data presented here validate the predictions from the computational analysis and further support the idea that CoTC-mediated transcriptional termination may be a prevalent phenomenon.

## DISCUSSION

It has been shown that CoTC-mediated termination results in enhanced mRNA and protein levels (5), a finding which indicates that CoTC terminators are important regulators of gene expression. It is possible that this means of enhancing gene expression is used for highly expressed genes and/or regulated genes whose products need to be made available quickly. In this study, we describe the role of homopolymeric A/T tracts of the  $\beta$ -globin Terminator in Pol II termination and show that the integrity of these homopolymer tracts is required for co-transcriptional RNA cleavage and subsequent transcription termination. In addition, results indicating that transcriptional termination can be restored when the oligoA tract mutations are converted to Ts suggest that the strength of the RNA:DNA hybrid is important in the termination process. Even though A/T tracts are a required feature for CoTC terminators, it is apparent that other undetermined sequence features are also required. Thus, reversal of the Terminator sequence

(pTrev, Figure 1) causes loss of termination, even though the A/T tracts will persist. Possibly, overall secondary or tertiary structure is also a determinant factor in CoTC terminators. Furthermore, computational analysis reveals the widespread nature of the 9 AT tract motif, as defined bioinformatically, suggesting that these elements need to be found in the correct context to be functional in transcription termination.

Work in yeast and mammalian systems has shown the importance of weak A:T hybrids, in combination with *trans*-acting factors, in Pol I and Pol III termination (28). Moreover, the role of A:T hybrids in prokaryotic intrinsic terminators is well established. By demonstrating the importance of the oligoA/T tracts in Pol II termination, and their abundance in PTRs, we anticipate that definition of the mechanism of action and generality of CoTC terminators will become apparent.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table 1, Supplementary Figures 1–4 and Supplementary Materials and Methods.

## ACKNOWLEDGEMENTS

The authors are grateful to the NJP lab for advice and support throughout these studies, especially to Monika Gullerova. We also thank Pia Andersen for help and advice in the construction of the stable cell lines.

## FUNDING

Studies supported by a programme grant from Wellcome Trust (to N.J.P.). Funding for open access charge: Wellcome Trust.

*Conflict of interest statement.* None declared.

## REFERENCES

- Greger, I.H., Demarchi, F., Giacca, M. and Proudfoot, N.J. (1998) Transcriptional interference perturbs the binding of Sp1 to the HIV-1 promoter. *Nucleic Acids Res.*, **26**, 1294–1301.
- Mapendano, C.K., Lykke-Andersen, S., Kjems, J., Bertrand, E. and Jensen, T.H. Crosstalk between mRNA 3' end processing and transcription initiation. *Mol. Cell*, **40**, 410–422.
- O'Sullivan, J.M., Tan-Wong, S.M., Morillon, A., Lee, B., Coles, J., Mellor, J. and Proudfoot, N.J. (2004) Gene loops juxtapose promoters and terminators in yeast. *Nat. Genet.*, **36**, 1014–1018.
- Ansari, A. and Hampsey, M. (2005) A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping. *Genes Dev.*, **19**, 2969–2978.
- West, S. and Proudfoot, N.J. (2009) Transcriptional termination enhances protein expression in human cells. *Mol. Cell*, **33**, 354–364.
- Camblong, J., Iglesias, N., Fickentscher, C., Dieppois, G. and Stutz, F. (2007) Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in *S. cerevisiae*. *Cell*, **131**, 706–717.
- Gullerova, M. and Proudfoot, N.J. (2008) Cohesin complex promotes transcriptional termination between convergent genes in *S. pombe*. *Cell*, **132**, 983–995.
- Gromak, N., West, S. and Proudfoot, N.J. (2006) Pause sites promote transcriptional termination of mammalian RNA polymerase II. *Mol. Cell Biol.*, **26**, 3986–3996.

9. Skourti-Stathaki,K., Proudfoot,N.J. and Gromak,N. (2011) Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell*, **42**, 794–805.
10. West,S., Proudfoot,N.J. and Dye,M.J. (2008) Molecular dissection of mammalian RNA polymerase II transcriptional termination. *Mol. Cell*, **29**, 600–610.
11. Wollerton,M.C., Gooding,C., Wagner,E.J., Garcia-Blanco,M.A. and Smith,C.W. (2004) Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. *Mol. Cell*, **13**, 91–100.
12. Ashe,M.P., Griffin,P., James,W. and Proudfoot,N.J. (1995) Poly(A) site selection in the HIV-1 provirus: inhibition of promoter-proximal polyadenylation by the downstream major splice donor site. *Genes Dev.*, **9**, 3008–3025.
13. Plant,K.E., Dye,M.J., Lafaille,C. and Proudfoot,N.J. (2005) Strong polyadenylation and weak pausing combine to cause efficient termination of transcription in the human Ggamma-globin gene. *Mol. Cell. Biol.*, **25**, 3276–3285.
14. Ashe,H.L., Monks,J., Wijgerde,M., Fraser,P. and Proudfoot,N.J. (1997) Intergenic transcription and transinduction of the human beta-globin locus. *Genes Dev.*, **11**, 2494–2509.
15. Dye,M. (1999), Studies in transcriptional termination of Pol II genes, D.Phil Thesis. University of Oxford.
16. West,S., Gromak,N. and Proudfoot,N.J. (2004) Human 5' → 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature*, **432**, 522–525.
17. Dye,M.J., Gromak,N. and Proudfoot,N.J. (2006) Exon tethering in transcription by RNA polymerase II. *Mol. cell*, **21**, 849–859.
18. Bailey,T.L. and Elkan,C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.; ISMB*, **2**, 28–36.
19. Gentleman,R.C., Carey,V.J., Bates,D.M., Bolstad,B., Dettling,M., Dudoit,S., Ellis,B., Gautier,L., Ge,Y., Gentry,J. *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*, **5**, R80.
20. Kent,W.J., Sugnet,C.W., Furey,T.S., Roskin,K.M., Pringle,T.H., Zahler,A.M. and Haussler,D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.
21. Dye,M.J. and Proudfoot,N.J. (2001) Multiple transcript cleavage precedes polymerase release in termination by RNA polymerase II. *Cell*, **105**, 669–681.
22. Cullen,B.R. (1993) Does HIV-1 Tat induce a change in viral initiation rights? *Cell*, **73**, 417–420.
23. Dye,M.J. and Proudfoot,N.J. (1999) Terminal exon definition occurs cotranscriptionally and promotes termination of RNA polymerase II. *Mol. Cell*, **3**, 371–378.
24. Wuarin,J. and Schibler,U. (1994) Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. *Mol. Cell. Biol.*, **14**, 7219–7225.
25. Dye,M.J., Gromak,N., Haussecker,D., West,S. and Proudfoot,N.J. (2006) Turnover and function of noncoding RNA polymerase II transcripts. *Cold Spring Harb. Symp. Quant. Biol.*, **71**, 275–284.
26. West,S., Gromak,N., Norbury,C.J. and Proudfoot,N.J. (2006) Adenylation and exosome-mediated degradation of cotranscriptionally cleaved pre-messenger RNA in human cells. *Mol. Cell*, **21**, 437–443.
27. Tian,B., Hu,J., Zhang,H. and Lutz,C.S. (2005) A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res.*, **33**, 201–212.
28. Richard,P. and Manley,J.L. (2009) Transcription termination by nuclear RNA polymerases. *Genes Dev.*, **23**, 1247–1269.