

# SCIENTIFIC REPORTS



OPEN

## Genomewide Analysis of Clp1 Function in Transcription in Budding Yeast

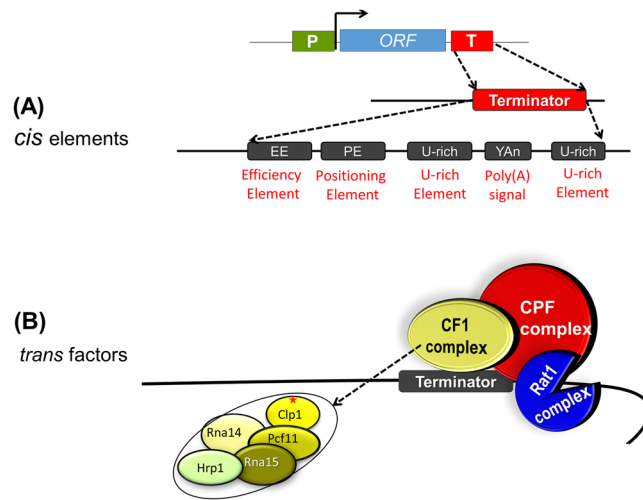
Nadra Al-Husini<sup>1</sup>, Ali Sharifi<sup>2,3,4</sup>, Seyed Ahmad Mousavi<sup>3</sup>, Hamidreza Chitsaz<sup>2</sup> & Athar Ansari<sup>1</sup>

In budding yeast, the 3' end processing of mRNA and the coupled termination of transcription by RNAPII requires the CF IA complex. We have earlier demonstrated a role for the Clp1 subunit of this complex in termination and promoter-associated transcription of *CHA1*. To assess the generality of the observed function of Clp1 in transcription, we tested the effect of Clp1 on transcription on a genomewide scale using the Global Run-On-Seq (GRO-Seq) approach. GRO-Seq analysis showed the polymerase reading through the termination signal in the downstream region of highly transcribed genes in a temperature-sensitive mutant of Clp1 at elevated temperature. No such terminator readthrough was observed in the mutant at the permissive temperature. The poly(A)-independent termination of transcription of snoRNAs, however, remained unaffected in the absence of Clp1 activity. These results strongly suggest a role for Clp1 in poly(A)-coupled termination of transcription. Furthermore, the density of antisense transcribing polymerase upstream of the promoter region exhibited an increase in the absence of Clp1 activity, thus implicating Clp1 in promoter directionality. The overall conclusion of these results is that Clp1 plays a general role in poly(A)-coupled termination of RNAPII transcription and in enhancing promoter directionality in budding yeast.

The CF IA 3' end processing complex of budding yeast is composed of four subunits: Rna14, Rna15, Pcf11 and Clp1 as shown in Fig. 1B<sup>1–3</sup>. The complex has been implicated in cleavage-polyadenylation and coupled termination of transcription<sup>4</sup>. The termination function of the CF IA complex was demonstrated using a nuclear run-on assay that revealed the readthrough of RNAPII beyond the poly(A) site in the mutants of all four CF IA subunits<sup>4,5</sup>. The transcription readthrough phenotype, however, was demonstrated for just a few selected yeast genes. ChIP analysis identified that the whole CF IA complex is localized at the 3' end of selected genes in accordance with its role in 3' end processing and termination of transcription<sup>5,6</sup>. Genomewide analysis found Pcf11 crosslinked to the 3' end of a majority of transcriptionally active yeast genes<sup>7</sup>. These observations suggest that the CF IA complex could be playing a general role in the termination of transcription in budding yeast. More direct evidence, however, was needed in support of a role for the CF IA complex in the termination of transcription on a genomewide scale and to firmly establish it as a general termination factor.

A vast majority of RNAPII-transcribed genes in yeast and higher eukaryotes exhibit antisense transcription initiating from both their 5' as well as the 3' ends<sup>8,9</sup>. In yeast, the 5' end initiated antisense transcripts are rapidly degraded by the RNA surveillance machinery of the cell and are therefore referred to as 'cryptic unstable transcripts' (CUTs)<sup>10</sup>. In contrast, the 3' end initiated antisense transcripts are stable and belong to the category of 'stable unannotated transcripts' (SUTs)<sup>8</sup>. At least 50% of SUTs in budding yeast are 3' end initiated antisense transcripts. A number of genes in yeast and mammalian cells exhibit 3' end initiated antisense transcription under repressed conditions. The *GAL10* gene of budding yeast, for example, exhibits robust sense transcription in the presence of galactose<sup>11,12</sup>. Upon shifting of cells to glucose containing medium, however, sense transcription is almost completely inhibited, and the 3' end initiated antisense transcription predominates. It has been proposed that when the 3' end initiated antisense transcript reaches the promoter end of the gene, it may adversely affect

<sup>1</sup>Department of Biological Science, Wayne State University, Detroit, MI, 48202, USA. <sup>2</sup>Department of Computer Science, Colorado State University, Fort Collins, CO, 80523, USA. <sup>3</sup>Department of Stem Cells and Developmental Biology, Cell Science Research Center, Royan Institute for Stem Cell Biology and Technology, ACECR, Tehran, Iran. <sup>4</sup>Department of Computer Engineering, Sharif University of Technology, Tehran, Iran. Correspondence and requests for materials should be addressed to H.C. (email: [chitsaz@chitsazlab.org](mailto:chitsaz@chitsazlab.org)) or A.A. (email: [ansari@biology.biosci.wayne.edu](mailto:ansari@biology.biosci.wayne.edu))



**Figure 1.** Termination of transcription in yeast requires DNA elements located near the 3' end of genes and accessory protein factors. **(A)** *Cis*-acting DNA elements required for termination of transcription in budding yeast. **(B)** CF 1, CPF and Rat1 are three multisubunit protein complexes required for termination of transcription in budding yeast. CF1 subunit Clp1 shown with an asterisk is the focus of this investigation.

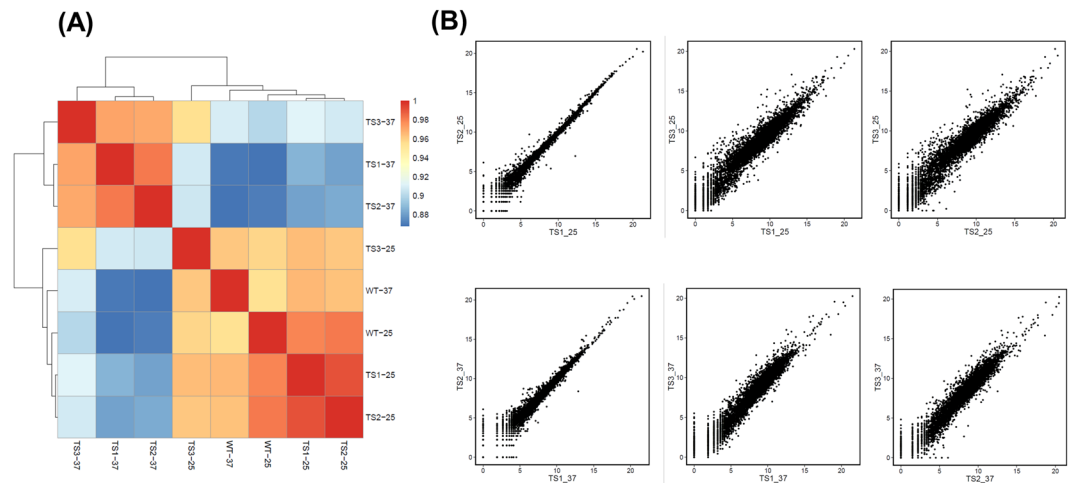
initiation/reinitiation either through transcriptional interference or through histone modifications<sup>13–15</sup>. Whether 3' end initiated antisense transcription is a general feature of termination defective mutants is not clear.

Since the CF IA complex is involved in the cleavage-polyadenylation of mRNA and termination of transcription, it is expected to be present near the 3' end of genes. It was, however, intriguing to find all four subunits of the CF IA complex occupying the 5' end of genes as well<sup>6</sup>. Furthermore, the Su72 subunit of the CPF complex, which is also linked to 3' end processing and termination of transcription in budding yeast, has been localized to the promoter end of genes<sup>16,17</sup>. The presence of termination factors towards the 5' end of genes is an evolutionarily conserved feature as a number of 3' end processing/termination factors of higher eukaryotes also have been found crosslinked to the distal ends of genes<sup>18</sup>. The promoter occupancy of termination factors remained an enigma until recently when it was demonstrated that these factors provide directionality to transcription by limiting divergent antisense transcription initiating from the promoter region<sup>5,10,18–20</sup>. The nucleosome free region located at the 5' end of a majority of protein coding genes in yeast and mammals contains two juxtaposed unidirectional promoters<sup>21–25</sup>. The transcription from these promoters initiates in both the downstream sense direction as well as the upstream antisense direction. The downstream sense transcription of the coding region produces mRNA, while the upstream antisense transcription produces upstream antisense RNA (uaRNA)<sup>10,19</sup>. Transcription in the upstream antisense direction is terminated within a few hundred bases from the transcription start site by the termination factors residing in the promoter region, while productive transcription in the sense direction is allowed to proceed until the polymerase reaches the 3' end of the gene<sup>10,18–21,26</sup>. Thus, promoter linked termination factors provide directionality to the inherently bidirectional eukaryotic promoter region by limiting synthesis of uaRNA. It has been demonstrated that gene looping plays a crucial role in the termination of uaRNA synthesis in budding yeast<sup>5,10</sup>. We recently showed that the CF IA complex is involved in providing transcription directionality during transcription of selected yeast genes<sup>5</sup>. Whether the CF IA complex has a general role in conferring promoter directionality during RNAPII-mediated transcription, however, needed further investigation.

Using a temperature-sensitive mutant of Clp1, we recently demonstrated the role of the CF IA complex in termination of transcription as well as in the promoter directionality of the *CHA1* gene in budding yeast<sup>5</sup>. Here we extend these studies to show that the CF IA complex has a global role in the poly(A)-dependent termination of transcription. Employing the GRO-Seq approach, we further showed that the CF IA complex also influence promoter directionality of subset of yeast genes. How the CF IA complex affects promoter-associated transcription, however, remains to be elucidated.

## Results

**Clp1 is required for the termination of transcription on a genome-wide scale.** A combination of biochemical and genetic approaches has identified two macromolecular complexes called CF I and CPF being essential for both 3' end processing and poly(A)-dependent termination of transcription in budding yeast (Fig. 1)<sup>2,27,28</sup>. In addition, the Rat1 complex is also required to accomplish the termination step. When termination is defective, RNAPII does not dissociate from the template beyond the poly(A) site, but reads through the poly(A) termination signal<sup>2,29</sup>. This transcription readthrough phenotype is characteristic of defective termination, and has been widely used to determine the involvement of a factor in the termination of transcription. A reliable approach used for detecting the terminator readthrough phenotype is the nuclear run-on (NRO) assay<sup>4</sup>. This assay measures the density of transcriptionally active polymerase over a specific genomic locus. The genome-wide version of this assay, called 'Global Run-On-Seq (GRO-Seq)', overcomes the limitations in the NRO assay in genome coverage and the strand specificity<sup>21</sup>. In GRO-Seq, the newly synthesized transcripts incorporate BrUTP,



**Figure 2.** GRO-Seq experiments performed in triplicates in *clp1* mutant exhibit high degree of correlation. **(A)** Heatmap of pairwise Pearson Correlation Coefficients among all samples. **(B)** Scatter plots of indicated pair of samples.

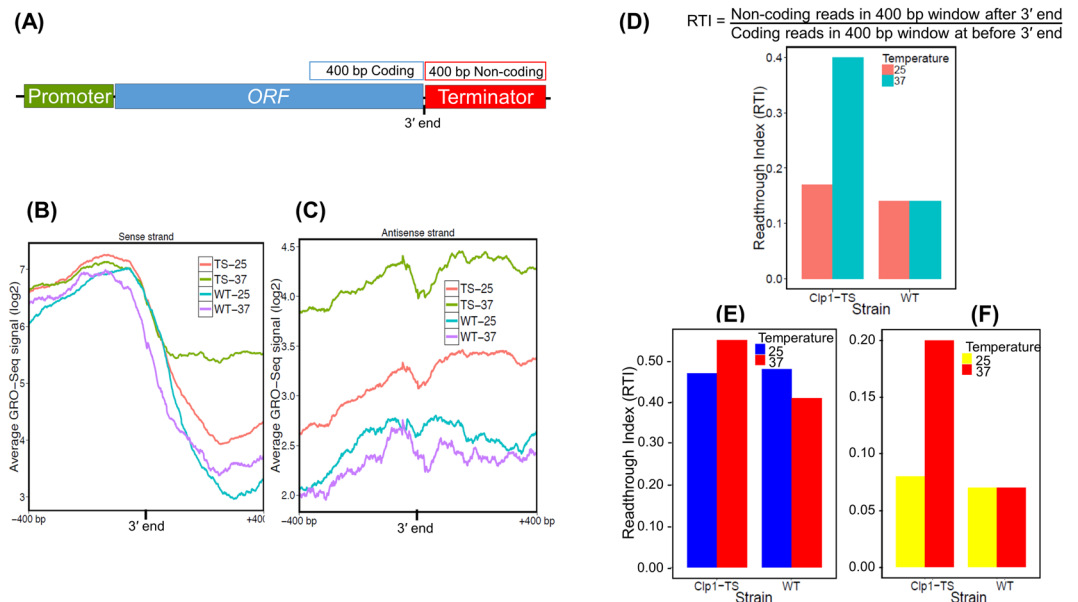
which allow for the affinity purification of these nascent transcripts on an anti-BrUTP column. The resultant nascent RNAs are subjected to high throughput sequencing. The outcome of GRO-Seq is a snapshot of the position as well as the density of the actively engaged polymerase in a strand specific manner.

To assess the role of the CF IA subunit Clp1 in transcription on a genomewide scale, we used the GRO-Seq approach. Comparing the GRO-Seq maps of the temperature-sensitive mutant of Clp1 and the isogenic wild type strain at the permissive (25 °C) and non-permissive (37 °C) temperatures of the mutant can reveal if Clp1 is a universal termination factor like the general transcription factors, or its role in termination is restricted to a subset of genes. Accordingly, we performed GRO-Seq analysis in the Clp1 mutant and the isogenic wild type strains in the cells grown at 25 °C and 37 °C. The experiments with the Clp1 mutant were performed in triplicates named TS-1, TS-2 and TS-3. The number of transcriptionally active genes was determined by using an experimentally determined threshold of 5 reads per kilo base. We found that ~5500 of the 6693 annotated ORFs (~82%) in the Saccharomyces Genome Database (SGD) were expressed as tags above the background. The heatmap of pairwise Pearson Correlation Coefficients among all samples shows that all TS-37 samples (TS1-37, TS2-37 and TS3-37) are clustered together, while TS-25 (TS1-25, TS2-25 and TS3-25), WT-25 and WT-37 samples are clustered together (Fig. 2A,B and Supplemental Fig. 1).

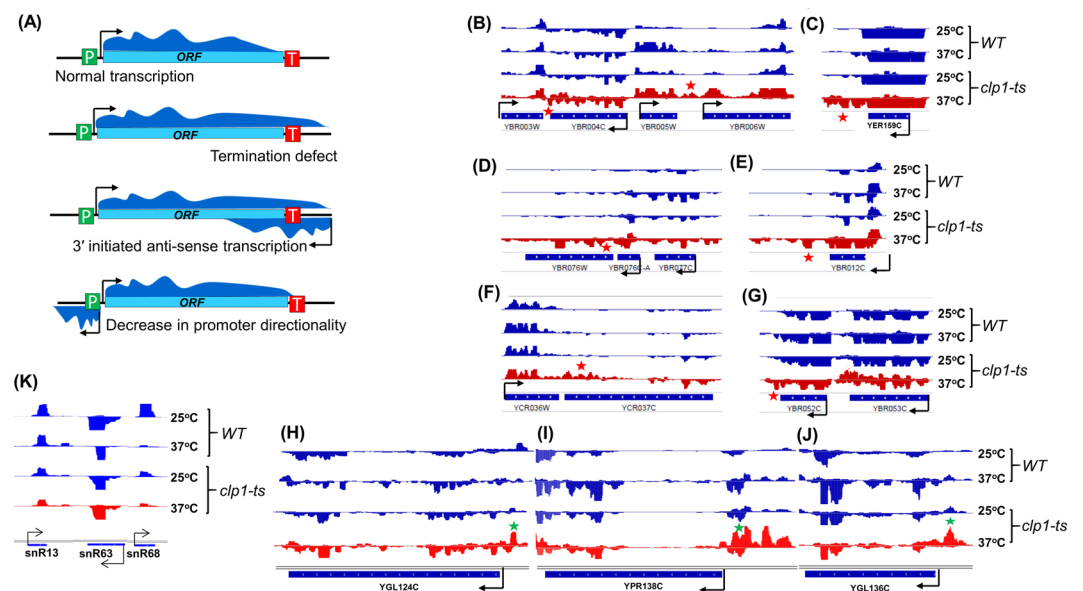
To investigate the role of Clp1 in the termination of transcription, we mapped GRO-Seq reads to the annotated 3' end of genes. Throughout this manuscript, the 5' end is the start of the open reading frame, while the 3' end is the end of the ORF of a gene. For this analysis, we removed all those genes whose next neighboring gene from the 5' and 3' end was less than 400 bp away. This was done because of the compact nature of the yeast genome, which often results in the terminator region of a gene overlapping with the promoter or terminator elements of the neighboring gene. We therefore restricted our analysis to 1038 genes whose 5' and 3' ends were at least 400 bp away from the neighboring ORF. All these 1038 genes were not transcriptionally active in the mutant and the wild type cells. The number of transcriptionally active genes varied from 664 to 844 as shown in Supplemental Table 1. We compared the transcription in the sense direction near the 3' end of genes in the mutant and isogenic wild type strains at 25 °C and 37 °C. The wild type strain exhibited a distinct peak of activity towards the 3' end of genes (Fig. 3B). The peak was followed by a sharp drop off beyond the presumed polyadenylation signal. A similar peak and drop in the GRO-Seq signal near the 3' end of genes was observed in the mutant strain at the permissive temperature (25 °C) (Fig. 3B). In contrast, there was no drop off, but an increase in the number of reads beyond the polyadenylation signal in the mutant at 37 °C compared to 25 °C (Fig. 3B). No such increase in the number of GRO-Seq signal was observed in the wild type cells at 37 °C (Fig. 3B). A logical conclusion of these results is that the polymerase was unable to read the poly(A) termination signal in the mutant at 37 °C leading to an enhanced readthrough signal in the 3' downstream region of the gene.

A number of gene specific examples are provided to illustrate the role of Clp1 in termination. These genes show an increase in GRO-Seq signal beyond their 3' ends in the mutant at 37 °C (Fig. 4B–G). Since GRO-Seq approach has never been used before to study termination of transcription in budding yeast, we wanted to confirm the effectiveness of the approach to study the global transcription termination in yeast. We therefore examined transcription of two genes, which exhibited the termination defective phenotype in our GRO-Seq analysis, by strand-specific TRO assay. We selected *RFS-1* (YBR052C) and *GPI18* (YBR004C) genes, which exhibit the termination defect in the GRO-Seq analysis in *clp1* mutant at 37 °C (Fig. 4G and B). The strand-specific TRO analysis of *RFS1* and *GPI18* revealed comparable results. Both genes exhibited terminator readthrough phenotype in the mutant at elevated temperature (Supplemental Fig. 2). These results corroborated the suitability of GRO-Seq approach in studying termination of transcription on a genomewide scale in yeast.

To further characterize the role of Clp1 in termination on a genomewide scale, we calculated the readthrough index (RTI) in the mutant and wild type cells grown at 25 °C and 37 °C. RTI was calculated by dividing the



**Figure 3.** Alignment of GRO-Seq reads to the 3' end in *clp1* mutant and the isogenic wild type strain revealed a role for Clp1 in general termination of transcription in budding yeast. **(A)** Schematic depiction of a gene showing the window of 400 bp coding and non-coding regions near the 3' end of genes where the GRO-Seq reads were aligned. **(B)** GRO-Seq reads aligned to the 400 bp window flanking the 3' end of genes in sense direction in *clp1-ts* mutant and the wild type strains at 25 °C and 37 °C. **(C)** GRO-Seq reads aligned to the 400 bp windows flanking the 3' end of genes in antisense direction in *clp1-ts* mutant and the wild type strains at the indicated temperatures. **(D)** Readthrough Index (RTI) was calculated as indicated in *clp1-ts* mutant and the wild type strains at 25 °C and 37 °C. **(E)** RTI plots of genes with low transcription activity (RPKM value of less than 24.22), and **(F)** of those with high transcription activity (RPKM value more than 24.22) in *clp1-ts* mutant and the wild type strains at 25 °C and 37 °C.



**Figure 4.** Snapshot of individual genes from GRO-Seq analysis. **(A)** Schematic depiction of a genes showing expected distribution of GRO-Seq reads on a gene when termination is normal or defective or when there is 3' end initiated or 5' end initiated antisense transcription. **(B–J)** GRO-Seq reads aligned to selected individual genes in sense and antisense direction in *clp1-ts* mutant and the wild type strains at 25 °C and 37 °C. Arrow indicates the transcription start site (TSS) and the direction of transcription. Name of each gene is indicated below. Red star shows transcription readthrough phenotype, while green star indicates promoter-associated upstream antisense transcription in the mutant at 37 °C. Scale of Y-axis in each gene snapshot is given in Supplemental Table 2. **(K)** GRO-Seq reads aligned to three snoRNA genes in sense and antisense direction in *clp1-ts* mutant and the wild type strains at 25 °C and 37 °C. Scale of Y-axis is given in Supplemental Table 2.

number of reads beyond the 3' end of a genes with the number of reads before the 3' end within the coding region. A termination defect will result in more GRO-Seq reads downstream of the 3' end of a genes and therefore a higher RTI value. Accordingly, RTI value increased by about 2-fold in the mutant at 37 °C compared to at 25 °C (Fig. 3D). No such increase was observed in the wild type strain at elevated temperature (Fig. 3D). These results strongly suggested that Clp1 is required for termination of transcription by RNA polymerase II in budding yeast. It was, however, not clear from this analysis if Clp1 is required for termination of transcription of all RNA polymerase II-transcribed genes or its function is restricted to a subset of class II genes. To address the issue, we divided 664 transcriptionally active genes in WT-25 sample (see Supplemental Table 1) into two categories, one with high expression level (higher than median in WT-25) and the other with low expression level (lower than median in WT-25). The low expression group consisted of 355 genes, while the high expression group comprised of 309 genes. We calculated RTI values for both categories in the mutant and the wild type cells at 25 °C and 37 °C as described above. The genes with low transcription activity exhibited inefficient termination even in the wild type cells and exhibited only a marginal increase in RTI in the mutant at elevated temperature (Fig. 3E). In contrast, genes with higher expression level showed strong dependence on Clp1 for termination of transcription as their RTI value increased by nearly 2.5-fold in the absence of Clp1 function (Fig. 3F).

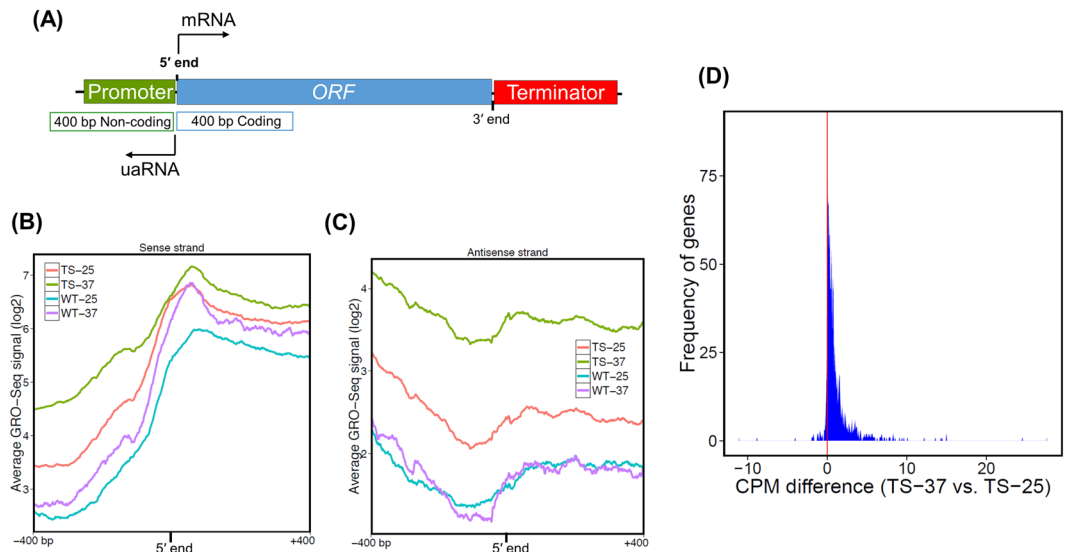
Next, we examined if Clp1 is required for poly(A)-dependent termination of transcription only, or for the poly(A)-independent termination as well. The transcription termination of the snoRNA genes occurs in a poly(A)-independent manner. Most of the genes coding for snoRNA are located in the crowded region of yeast genome. We, therefore, looked at the termination of three snoRNA genes, snR13, snR63 and snR68, which are separated from the neighboring genes by several hundred base pairs. No termination defect was observed for these three genes in the *clp1* mutant at elevated temperature (Fig. 4K). These results suggest that the role of Clp1 may be restricted to the poly(A)-coupled termination of transcription by RNAII in budding yeast.

**Clp1 and 3' initiated antisense transcription.** A number of genes both in yeast and mammalian systems exhibit 3' initiated antisense transcription upon defective termination. It has been shown that the 3' initiated antisense transcription inhibits the recruitment of transcription machinery at the promoter region for initiation of transcription. We, therefore, examined the role of Clp1 in antisense transcription originating from the 3' end of genes. We choose the same set of 1038 genes for this analysis whose 5' and 3' ends were at least 400 bp away from the neighboring gene. Alignment of GRO-Seq reads to the 3' end of these genes revealed enhanced antisense transcription in the Clp1 mutant at 37 °C compared to at 25 °C (Fig. 3C). No 3' end antisense transcript enhancement was observed in the wild type strain at 37 °C (Fig. 3C). It is, however, difficult to distinguish the 3' initiated antisense transcription from the 5' initiated antisense transcription of the neighboring downstream gene, and the readthrough transcription from the convergent downstream gene. The relationship between 3' initiated antisense transcription and the sense transcription in the Clp1 mutant therefore requires a more thorough scrutiny.

**Role of Clp1 in promoter-associated transcription.** We have earlier showed that Clp1 plays a crucial role in promoter-associated transcription of *CHA1* by facilitating reinitiation and conferring promoter directionality<sup>5</sup>. We, therefore, examined if Clp1 is facilitating initiation/reinitiation of transcription on a genomewide scale. If Clp1 was facilitating initiation/reinitiation of transcription, we expected a decrease in the density of mRNA transcribing polymerase in the mutant at 37 °C. When we mapped GRO-Seq reads of the mutant and the wild type cells grown at 37 °C, to the annotated 5' end of genes, there was no decrease in the GRO-Seq signal for sense transcribing polymerase in the mutant at elevated temperature (Fig. 5B). In fact, there was a slight increase in the density of mRNA-transcribing polymerase in the promoter-proximal 400 bp coding region in the mutant at 37 °C (Fig. 5B). A similar polymerase peak was observed in the promoter-proximal downstream region in the mutants of termination factors in mammalian cells<sup>19</sup>. These results argue against a role for Clp1 in recruitment of polymerase for initiation/reinitiation of transcription.

Next, we determined the role of Clp1 in transcription directionality. To determine the generality of Clp1 function in the termination of uaRNA synthesis, and thus the maintenance of promoter directionality, we mapped GRO-Seq reads in the mutant and the wild type cells grown at 25 °C and 37 °C, to the annotated 5' end of genes. For this analysis, we chose the same set of 1038 genes whose 5' and 3' ends were at least 400 bp away from the neighboring gene. The GRO-Seq reads in antisense direction upstream of the 5' end of genes represent uaRNA, while GRO-Seq reads in the promoter downstream sense direction represent mRNA (Fig. 5A). As described above, there was only a marginal increase in mRNA reads in 400 bp promoter downstream coding window both in the mutant and wild type cells at elevated temperature (Fig. 5B). In contrast, a two fold increase in the intensity of uaRNA signal was observed in the mutant at 37 °C compared to 25 °C in 400 bp promoter upstream non-coding window (Fig. 5C). No such increase was observed in the wild type cells at elevated temperature. Thus, there was a decrease in promoter directionality in the mutant at 37 °C. A number of gene specific examples are provided to illustrate the role of Clp1 in promoter directionality (Fig. 4H–J). These genes show an increase in GRO-Seq signal in the promoter-upstream antisense direction in the mutant at 37 °C.

To compute the number of genes exhibiting a decrease in promoter directionality in terms of enhanced uaRNA transcription in *clp1* mutant at elevated temperature, we calculated ratio of GRO-Seq reads in 100 bp promoter upstream region in the antisense direction for individual genes in the mutant at 37 °C and 25 °C. We computed the difference in CPM (Count Per Million) reads in the mutant at the permissive and non-permissive temperatures as described in 'Methods' section. Our results indicate that of around 700 transcriptionally active genes analyzed in this study, 330 genes exhibited the CPM difference greater than +1 (in log<sub>2</sub> scale, which means at least 2-fold higher uaRNA expression in the mutant at 37 °C), while only 13 genes had CPM difference less than -1 (Fig. 5D). These results clearly demonstrate that the uaRNA transcription significantly increased in the *clp1* mutant at 37 °C for 330 genes (*p*-value < 1E-15, Wilcoxon signed rank test), while remaining approximately 60% of the transcriptionally active genes analyzed in this study did not exhibit any change in promoter directionality at elevated temperature. The overall conclusion of these results is that Clp1 influences the promoter directionality



**Figure 5.** Alignment of GRO-Seq reads to the 5' end suggests a role for Clp1 in promoter directionality of genes. (A) Schematic depiction of a gene showing the window of 400 bp coding and non-coding regions near the 5' end of genes where the GRO-Seq reads were aligned. (B) GRO-Seq reads aligned to the 400 bp window flanking the 5' end of genes in sense direction in *clp1-ts* mutant and the wild type strains at 25 °C and 37 °C. (C) GRO-Seq reads aligned to the 400 bp windows flanking the 5' end of genes in antisense direction in *clp1-ts* mutant and the wild type strains at the indicated temperatures. (D) A histogram showing the number of genes exhibiting a decrease in promoter directionality upon shifting the mutant cells to non-permissive temperature. A plot of CPM difference in the 100 bp non-coding window upstream of the promoter was calculated in the mutant at 37 °C and 25 °C and plotted as a function of fraction of genes.

of nearly 40% of transcriptionally active genes analyzed in our study. These include both highly expressed and lowly expressed genes.

## Discussion

Of the three steps of the RNAPII transcription cycle (initiation, elongation and termination), initiation is the most well understood step. The general initiation factors are well characterized, and it is possible to perform initiation of transcription from a defined promoter using highly purified factors under *in vitro* conditions. In contrast, termination is the least understood step of the transcription cycle. The accessory factors required for termination are not thoroughly characterized. Consequently, termination of transcription under *in vitro* conditions using purified yeast factors has not been achieved so far in budding yeast. A thorough understanding of the process of termination by RNAPII necessitates identification of the general termination factors in line with the general transcription factors that are required for initiation of transcription. A number of approaches like PAR-CLIP, GRO-Seq, PRO-Seq, and Net-Seq can be used to determine the genomewide role of an identified factor in termination of transcription. All these approaches may give similar results, but there are subtle differences in the precise information they reveal. We used GRO-Seq approach because of its high resolution and sensitivity in detecting defective termination.

An important conclusion of the GRO-Seq analysis performed here is that Clp1 is required for termination of transcription of RNAPII-transcribed genes in budding yeast. Our analysis further show that Clp1 is required for poly(A)-dependent termination of transcription of relatively highly expressed genes. Since there is no report of Clp1 acting alone or being a part of a complex other than CF 1A, we extrapolate these results to propose that the whole CF 1A complex, and not its individual components separately, is required for termination of transcription. Our results are corroborated by a recent finding that another CF 1A subunit Pcf11 is required for termination of transcription on a genomewide scale<sup>30</sup>. A similar genomewide analysis was performed for CstF64, which is the mammalian homologue of another CF 1A subunit Rna15<sup>19</sup>. This analysis identified CstF64 as a general termination factor in mammalian RNAPII transcription cycle. Thus, the role of CF 1A complex and its mammalian counterpart in termination of transcription has been conserved during evolution. Besides CF 1A complex, termination of transcription in yeast requires CPF complex. It was recently demonstrated using PAR-CLIP approach that the CPF subunit Ysh1 is required for global termination of transcription<sup>31</sup>. The termination defective phenotype of Ysh1, however, was quite different from that of Clp1. Unlike Clp1, depletion of Ysh1 did not lead to readthrough transcription, but resulted in the accumulation of polymerase beyond the poly(A) site. These results suggested a role for Ysh1 in dissociating polymerase from the template beyond the poly(A) site. In contrast, CF 1A complex could be facilitating slowing down of polymerase near the poly(A) site, which is an important step in termination of transcription by RNAPII.

During transcription, the distal ends of a gene physically interact with each other forming a looped gene architecture<sup>16,32</sup>. The proximity of the promoter and terminator regions in a gene loop allows promoter-bound factors to influence termination of transcription<sup>6,33-36</sup>. The gene looping also invokes the possibility of termination

factors affecting the promoter-associated transcription. Accordingly, it was demonstrated that Ssu72, which is a component of CPF termination complex, affects directionality of promoter-initiated transcription<sup>10</sup>. We also showed that Clp1 affects transcription directionality of *CHAT*<sup>5</sup>. Clp1 is also required for gene looping of *CHAT*<sup>5</sup>. The results of this genomewide analysis demonstrate that, apart from its role in termination, Clp1 also function in regulating promoter directionality of RNAPII-transcribed yeast genes. Clp1 suppresses the promoter-initiated upstream antisense transcription of 330 out of about 700 transcriptionally active genes used in this analysis. The promoter-initiated downstream antisense transcription either remained unaffected or exhibited a marginal increase for a majority of genes in the absence of Clp1 activity. Thus, the role of Clp1 in promoter directionality is similar to that of CPF subunit Ssu72<sup>10</sup>. CAF-1 complex, which is associated with chromatin assembly pathway, also affected promoter-initiated divergent non-coding transcription of yeast genes in a manner similar to Clp1<sup>37</sup>. Both Ssu72 and CAF-1 complex suppress promoter-initiated uRNA transcription. The transcription of promoter-associated mRNA, however, either remains unaffected or exhibited a small increase in the absence of Ssu72 or CAF-1 activities. The NET-Seq analysis performed in mammalian cells likewise revealed the role of CstF64 and Xrn2 termination factors in transcription directionality<sup>19</sup>.

Overall conclusion of these results is that the CF 1A complex plays a role in poly(A)-coupled termination of transcription, and in limiting promoter-associated divergent antisense transcription on a genomewide scale in budding yeast. Not all RNAPII-transcribed yeast genes, however, are dependent on CF 1A complex for either termination of their transcription or promoter directionality. This is expected as there is an alternate pathway for termination of transcription in budding yeast that requires Nrd1-Nab3-Sen1 complex. Nrd1-dependent termination has been shown to affect termination of non-polyadenylated mRNA as well as uRNA belonging to the category of CUTs. Is it gene looping that governs if termination of transcription and promoter directionality will be accomplished by the CF 1A complex or the Nrd1 complex remains to be elucidated?

## Methods

**Yeast Strains.** The yeast strains used in this study are BY4733 (*MATa his3Δ200 trp1Δ63 leu2Δ0 met15Δ0 ura3Δ0*) and *clp1-769-5* (*MATa ura3Δ0 leu2Δ0 his3Δ1lys2Δ0can1Δ::LEU2-MFA1pr::His3 clp1ts::URA3*).

**GRO-Seq approach.** GRO-Seq analysis was performed by the modification of method described in Core *et al.*<sup>21</sup>.

**Step I: Preparation of permeabilized cells.** The *clp1-769-5* mutant or the wild type cells were grown at 25 °C in 100 ml of YPD (Yeast extract-Peptone-Dextrose) medium to an OD<sub>600</sub> of 0.4. The 100 ml culture was transferred to two 50 ml sterile tubes, and centrifuged at 1860 × g for 5 minutes. Cell pellets were resuspended in 50 ml of 25 °C or 37 °C prewarmed YPD medium, and grown at the permissive (25 °C) or non-permissive temperature (37 °C) for 2 hours. This normally takes cells to an OD<sub>600</sub>~0.8. Equal number of cells were pelleted by centrifugation at 1860 × g for 5 minute at 4 °C. Cell pellets were resuspended in 10 ml of ice-cold TMN buffer (10 mM Tris-HCl pH 7.5, 5 mM MgCl<sub>2</sub>, 100 mM NaCl), incubated in ice for 10 minutes, centrifuged at 1860 × g for 5 minutes, and then resuspended in 940 μl of DEPC (Diethylpyrocarbonate)-treated ice-cold water. Cells were permeabilized by incubating with 60 μl 10% sarkosyl at 4 °C for 25 minutes with gentle shaking. Permeabilized cells were spun down at 1200 × g for 6 minutes at 4 °C. The pellets were stored at 4 °C till all samples were ready for the next step.

**Step II: Transcription run-on reaction and isolation of nascent RNA.** Transcription run-on reaction was performed by the modification of methods described in Birse *et al.*, (1997). Cell pellets from step I above were resuspended in 150 μl of transcription run-on reaction buffer (50 mM Tris-HCl pH 7.5, 100 mM KCl, 10 mM MgCl<sub>2</sub>, 2 mM DTT, 0.5 mM each of ATP, CTP, GTP and Brd-UTP) containing 5 μl of freshly added RNase Inhibitor (40 units/μl) and incubated at 30 °C for 5 minutes. The reaction was stopped by adding 500 μl of ice-cold Trizol reagent. About 250 μl of acid-washed glass beads were added to each tube. Cells were immediately lysed by vigorous shaking at 4 °C for 20 minutes. The lysate was transferred to a new 1.5 ml microfuge tube containing 500 μl of Trizol and 200 μl of chloroform. The tubes were shaken vigorously and centrifuged at 14000 rpm at 25 °C for 20 minutes. The supernatant was extracted with phenol-chloroform three times. Total RNA was precipitated by adding 3 volumes of ice-cold 100% ethanol, 2 μl of glycogen (20 mg/ml) and NaCl to a final concentration of 0.3 M followed by overnight incubation at -20 °C. The precipitated RNA was recovered by centrifugation at 16000 × g for 30 minutes at 4 °C. The RNA pellet was washed once with ice-cold 75% ethanol, air dried and resuspended in 55 μl of DEPC-treated water.

The extracted RNA was subjected to partial hydrolysis by incubating with 500 mM NaOH on ice for 20 minutes. The NaOH was neutralized by adding 30 μl of 1 M Tris-HCl (pH 6.6). Partially hydrolyzed RNA was purified using RNeasy kit (Qiagen) to remove any unincorporated nucleotides and to exchange buffer. RNA was eluted from the RNeasy column twice with 50 μl of DEPC treated water.

**Step III: Immunopurification of Brd-UTP-labeled RNA.** About 25 μl of anti-BrdU-antibody-conjugated agarose beads were washed three times with 500 μl of binding buffer (0.25x SSPE buffer, 1 mM EDTA, 0.05% Tween20, 37.5 mM NaCl). Beads were blocked by gently shaking with 500 μl of blocking buffer (485 μl binding buffer containing 5 μl of 10% polyvinylpyrrolidone and 10 μl of Ultrapure BSA) for two hours at 4 °C. The blocked beads were washed twice with 500 μl of binding buffer and then resuspended in 400 μl of the same buffer. The RNA from step II was incubated at 65 °C for 5 minutes to get rid of any secondary structure and immediately placed on ice for 2 minutes. The RNA was then bound to the processed anti-BrdU beads with gentle shaking for one hour at 4 °C. Beads were then washed sequentially with 500 μl of binding buffer, 500 μl of low salt buffer (0.2x SSPE, 1 mM EDTA, 0.05% Tween20) and 500 μl of high salt buffer (0.25x SSPE, 1 mM EDTA, 0.05% Tween-20, 100 mM NaCl) once each, followed by two washes of 500 μl of TET buffer (1xTE buffer, 0.05% Tween20). All washes were performed on nutator for three minutes each. All centrifugation steps between the washes were at 194 × g for 1 minute each, with incubation of tubes on ice for 30 seconds before

removing wash buffer. The bound nascent RNA was eluted two times with 125  $\mu$ l and one time with 250  $\mu$ l of elution buffer (20 mM DTT, 150 mM NaCl, 50 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.1% SDS) for a total elution time of 10 minutes in 42 °C water bath. The eluted RNA was precipitated with NaCl as described previously.

RNA pellet was washed with ice-cold 75% ethanol, air dried and resuspended in 12  $\mu$ l of DEPC-treated water and quantified using a Nanodrop spectrophotometer.

**Step IV: rRNA depletion, library preparation and sequencing.** The nascent, purified RNA samples were first depleted of rRNA. The depletion of rRNA was performed by the Ribo-Zero™ Magnetic Kit (Epicenter) following manufacturer's recommendations. GRO-Seq libraries were constructed using the ScriptSeq™ complete Kit (Epicenter) following manufacturer's instructions. DNA Sequencing was performed at the DNA Core Facility of Cornell University.

**Gene expression analysis and RTI calculation.** We used Cutadapt<sup>38</sup> for quality control of raw fastq files to trim adapters and remove low quality reads. All high quality fastq files were mapped to *Saccharomyces* genome (Sacc3, downloaded from UCSC database) using Tophat2<sup>39</sup> (version 2.1.1, parameters -i 30 -I 800) and HTSeq<sup>40</sup> (version 0.6) for calculating read counts. Subsequently DESeq. 2 was performed for read count normalization. Then reads per kilobase of transcript per million (RPKM) mapped reads  $\geq 5$  RPKM was done for determining number of expressed genes<sup>41</sup>.

Pheatmap package is applied to generate pairwise Pearson correlation coefficients analysis on normalized count files, and ggplot2 package applied for creating pairwise scatter plot among all replicates.

To calculate 'Read Through Index' (RTI) score and expression difference of samples regards to their coding strand, again all high quality fastq files were mapped to three modified Gene Transfer format (GTF) files TSS, TTS and gene body regions of *Sacchar3* genome by Tophat2. In addition HTSeq and DESeq. 2 applied to obtain normalized read count. The output is used to calculate RTI score for each gene as shown in Fig. 3.

**RTI = 3' end-400 bp noncoding-sense/3' end-400 bp coding-sense.** All scripts for calculating RTI were performed using R and ggplot2 package employed to generate density plots and scatter plots.

**Calculating fraction of genes showing loss of promoter directionality in the mutant at elevated temperature.** To determine the number of genes showing loss of promoter directionality due to increased uaRNA transcription in the mutant at 37 °C, we computed the count of reads in each sample in the 100 bp non-coding window upstream to the TSS. We performed cross-sample normalization by dividing the RAW count values to the total number of aligned reads (in millions) per sample to compute Count Per Million (CPM) reads. Then we computed the Log2 fold-change of mean CPM values between TS-37 vs. TS-25 samples and the sum of CPM values for TS-37 samples minus TS-25 samples. The CPM difference greater than +1 indicates more than 2-fold higher expression in TS-37 samples, while a CPM difference less than -1 shows more than 2-fold higher expression in TS-25 sample. The mean CPM difference value for all genes is +1.136, and the median value is +0.545.

**Accession number.** The data present in this work was deposited in NCBI's Gene Expression Omnibus (GEO) database under the accession number GSE71624.

## References

- Gross, S. & Moore, C. Five subunits are required for reconstitution of the cleavage and polyadenylation activities of *Saccharomyces cerevisiae* cleavage factor I. *Proc Natl Acad Sci USA* **98**, 6080–6085, doi:10.1073/pnas.101046598 (2001).
- Richard, P. & Manley, J. L. Transcription termination by nuclear RNA polymerases. *Genes Dev* **23**, 1247–1269, doi:10.1101/gad.1792809 (2009).
- Minvielle-Sebastia, L., Preker, P. J., Wiederkehr, T., Strahm, Y. & Keller, W. The major yeast poly(A)-binding protein is associated with cleavage factor IA and functions in premessenger RNA 3'-end formation. *Proc Natl Acad Sci USA* **94**, 7897–7902 (1997).
- Birse, C. E., Minvielle-Sebastia, L., Lee, B. A., Keller, W. & Proudfoot, N. J. Coupling termination of transcription to messenger RNA maturation in yeast. *Science* **280**, 298–301 (1998).
- Al Husini, N., Kudla, P. & Ansari, A. A role for CF1A 3' end processing complex in promoter-associated transcription. *PLoS Genet* **9**, e1003722, doi:10.1371/journal.pgen.1003722 (2013).
- Medler, S. *et al.* Evidence for a complex of transcription factor IIB with poly(A) polymerase and cleavage factor I subunits required for gene looping. *J Biol Chem* **286**, 33709–33718, doi:10.1074/jbc.M110.193870 (2011).
- Mayer, A. *et al.* Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol* **17**, 1272–1278, doi:10.1038/nsmb.1903 (2010).
- Xu, Z. *et al.* Antisense expression increases gene expression variability and locus interdependency. *Mol Syst Biol* **7**, 468, doi:10.1038/msb.2011.1 (2011).
- Hah, N. *et al.* A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell* **145**, 622–634, doi:10.1016/j.cell.2011.03.042 (2011).
- Tan-Wong, S. M. *et al.* Gene loops enhance transcriptional directionality. *Science* **338**, 671–675, doi:10.1126/science.1224350 (2012).
- Murray, S. C. *et al.* A pre-initiation complex at the 3'-end of genes drives antisense transcription independent of divergent sense transcription. *Nucleic Acids Res* **40**, 2432–2444, doi:10.1093/nar/gkr1121 (2012).
- Malik, S., Durairaj, G. & Bhaumik, S. R. Mechanisms of antisense transcription initiation from the 3' end of the GAL10 coding sequence *in vivo*. *Mol Cell Biol* **33**, 3549–3567, doi:10.1128/MCB.01715-12 (2013).
- Hongay, C. F., Grisafi, P. L., Galitski, T. & Fink, G. R. Antisense transcription controls cell fate in *Saccharomyces cerevisiae*. *Cell* **127**, 735–745, doi:10.1016/j.cell.2006.09.038 (2006).
- Camblong, J., Iglesias, N., Fickentscher, C., Dieppois, G. & Stutz, F. Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in *S. cerevisiae*. *Cell* **131**, 706–717, doi:10.1016/j.cell.2007.09.014 (2007).
- Berretta, J., Pinskaya, M. & Morillon, A. A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in *S. cerevisiae*. *Genes Dev* **22**, 615–626, doi:10.1101/gad.458008 (2008).
- Ansari, A. & Hampsey, M. A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping. *Genes Dev* **19**, 2969–2978, doi:10.1101/gad.1362305 (2005).



17. Zhang, D. W. *et al.* Ssu72 phosphatase-dependent erasure of phospho-Ser7 marks on the RNA polymerase II C-terminal domain is essential for viability and transcription termination. *J Biol Chem* **287**, 8541–8551, doi:10.1074/jbc.M111.335687 (2012).
18. Almada, A. E., Wu, X., Kriz, A. J., Burge, C. B. & Sharp, P. A. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* **499**, 360–363, doi:10.1038/nature12349 (2013).
19. Nojima, T. *et al.* Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* **161**, 526–540, doi:10.1016/j.cell.2015.03.027 (2015).
20. Agarwal, N. & Ansari, A. Enhancement of Transcription by a Splicing-Competent Intron Is Dependent on Promoter Directionality. *PLoS Genet* **12**, e1006047, doi:10.1371/journal.pgen.1006047 (2016).
21. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848, doi:10.1126/science.1162228 (2008).
22. Seila, A. C., Core, L. J., Lis, J. T. & Sharp, P. A. Divergent transcription: a new feature of active promoters. *Cell Cycle* **8**, 2557–2564, doi:10.4161/cc.8.16.9305 (2009).
23. Duttke, S. H. *et al.* Human promoters are intrinsically directional. *Mol Cell* **57**, 674–684, doi:10.1016/j.molcel.2014.12.029 (2015).
24. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169, doi:10.1093/bioinformatics/btu638 (2015).
25. Scruggs, B. S. *et al.* Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin. *Mol Cell* **58**, 1101–1112, doi:10.1016/j.molcel.2015.04.006 (2015).
26. Neil, H. *et al.* Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* **457**, 1038–1042, doi:10.1038/nature07747 (2009).
27. Mandel, C. R. & Tong, L. How to get all “A”s in polyadenylation. *Structure* **15**, 1024–1026, doi:10.1016/j.str.2007.08.002 (2007).
28. Buratowski, S. Connections between mRNA 3' end processing and transcription termination. *Current opinion in cell biology* **17**, 257–261, doi:10.1016/j.cub.2005.04.003 (2005).
29. Mischo, H. E. & Proudfoot, N. J. Disengaging polymerase: terminating RNA polymerase II transcription in budding yeast. *Biochim Biophys Acta* **1829**, 174–185, doi:10.1016/j.bbagr.2012.10.003 (2013).
30. Baejen, C. *et al.* Genome-wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes. *Mol Cell* **66**, 38–49 e36, doi:10.1016/j.molcel.2017.02.009 (2017).
31. Schaughency, P., Merran, J. & Corden, J. L. Genome-wide mapping of yeast RNA polymerase II termination. *PLoS Genet* **10**, e1004632, doi:10.1371/journal.pgen.1004632 (2014).
32. El Kaderi, B., Medler, S., Raghunayakula, S. & Ansari, A. Gene looping is conferred by activator-dependent interaction of transcription initiation and termination machineries. *J Biol Chem* **284**, 25015–25025, doi:10.1074/jbc.M109.007948 (2009).
33. Wang, Y., Fairley, J. A. & Roberts, S. G. Phosphorylation of TFIIB links transcription initiation and termination. *Curr Biol* **20**, 548–553, doi:10.1016/j.cub.2010.01.052 (2010).
34. Henriques, T. *et al.* Transcription termination between polo and snap, two closely spaced tandem genes of *D. melanogaster*. *Transcription* **3**, 198–212, doi:10.4161/trns.21967 (2012).
35. Mukundan, B. & Ansari, A. Srb5/Med18-mediated termination of transcription is dependent on gene looping. *J Biol Chem* **288**, 11384–11394, doi:10.1074/jbc.M112.446773 (2013).
36. Medler, S. & Ansari, A. Gene looping facilitates TFIIB kinase-mediated termination of transcription. *Sci Rep* **5**, 12586, doi:10.1038/srep12586 (2015).
37. Marquardt, S. *et al.* A chromatin-based mechanism for limiting divergent noncoding transcription. *Cell* **157**, 1712–1723, doi:10.1016/j.cell.2014.04.036 (2014).
38. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* **17**, 10–12 (2011).
39. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology* **14**, 1 (2013).
40. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, btu638 (2014).
41. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods* **5**, 621–628 (2008).

## Acknowledgements

We are grateful to Dr. Leighton Core for helpful suggestions in standardizing the GRO-Seq protocol in budding yeast. We are thankful to Dr. Lori Pile (Wayne State University) for critical reading of the manuscript. We thank Dr. Philip Hieter for kindly providing the *clp1* mutant.

## Author Contributions

A.A. conceived the idea. A.A. and N.A.-H. designed experiments, N.A.-H. performed GRO-Seq experiments. Bioinformatics analysis of GRO-Seq data was performed by H.C., A.S. and S.A.M. A.A. wrote the paper.

## Additional Information

**Supplementary information** accompanies this paper at doi:10.1038/s41598-017-07062-6

**Competing Interests:** The authors declare that they have no competing interests.

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



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

© The Author(s) 2017