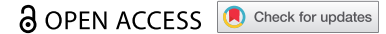



RESEARCH PAPER



CFIm25 regulates human stem cell function independently of its role in mRNA alternative polyadenylation

Yi Ran^{a, #}, Shanshan Huang^{a, #}, Junjie Shi^{a, #}, Qiumin Feng^a, Yanhui Deng^a, Andy Peng Xiang^a, and Chengguo Yao ^{a, b}

^aCenter for Stem Cell Biology and Tissue Engineering, Key Laboratory for Stem Cells and Tissue Engineering, Ministry of Education, Sun Yat-Sen University, Guangzhou, China; ^bDepartment of Genetics and Cell Biology, Zhongshan School of Medicine, Sun Yat-Sen University, Guangzhou, China

ABSTRACT

It has recently been shown that CFIm25, a canonical mRNA 3' processing factor, could play a variety of physiological roles through its molecular function in the regulation of mRNA alternative polyadenylation (APA). Here, we used CRISPR/Cas9-mediated gene editing approach in human embryonic stem cells (hESCs) for CFIm25, and obtained three gene knockdown/mutant cell lines. CFIm25 gene editing resulted in higher proliferation rate and impaired differentiation potential for hESCs, with these effects likely to be directly regulated by the target genes, including the pluripotency factor *rex1*. Mechanistically, we unexpectedly found that perturbation in CFIm25 gene expression did not significantly affect cellular mRNA 3' processing efficiency and APA profile. Rather, we provided evidences that CFIm25 may impact RNA polymerase II (RNAPII) occupancy at the body of transcribed genes, and promote the expression level of a group of transcripts associated with cellular proliferation and/or differentiation. Taken together, these results reveal novel mechanisms underlying CFIm25's modulation in determination of cell fate, and provide evidence that the process of mammalian gene transcription may be regulated by an mRNA 3' processing factor.

ARTICLE HISTORY

Received 17 January 2022
Revised 15 April 2022
Accepted 20 April 2022

KEYWORDS



CFIm25; hESC pluripotency; mRNA alternative polyadenylation; gene transcription; *rex1*

Introduction


Processing of pre-mRNA 3' end is a key step in eukaryotic gene expression [1,2]. Based on current models, processing of human canonical mRNA 3' end involves two coupled steps, namely cleavage and polyadenylation [3–6]. Specifically, cleavage requires two core multi-subunit complexes, namely cleavage and polyadenylation specificity factors (CPSF) and cleavage stimulation factor (CstF). On the other hand, polyadenylation involves addition of a poly(A) tail at the 3' end of pre-mRNA upon cleavage by Poly(A) polymerase (PAP). At the molecular level, mRNA 3' processing often occurs co-transcriptionally [7–9], and is tightly connected with all the three steps of mRNA transcription, namely initiation, elongation, and termination. For example, TFIID, one of the general transcription factors required for transcription initiation, has been implicated in regulation of mRNA 3' processing by associating with CPSF [10], the core subunit of 3' processing complex. More recent studies have shown that transcription activity at 5' end of genes could significantly affect mRNA 3' processing, though the detailed mechanisms remain elusive [11–13]. Another example is the phosphorylation of serine 2 residues (Ser2P) at the C-terminal domain (CTD) of heptad repeats of RPB1, the largest subunit of RNAPII. As transcription approaches termination, Ser2P facilitates recruitment of 3' processing factors to nascent transcripts [14,15]. Aside from the impact of transcription on mRNA 3' processing, emerging

evidences have shown that mRNA 3' processing, in turn, might impact transcription. For example, several yeast 3' processing factors reportedly interact with 5' end of genes, thereby impacting transcription through gene looping [16–19]. In human cells, 3' end formation has been shown to play a stimulatory role in transcription, possibly by recycling factors required for initiation/elongation [20]. Another example is U1 snRNP telescoping, a phenomenon linking premature transcription termination with mRNA 3' processing at numerous intronic polyadenylation sites (PASs) [21–23]. However, despite the crucial role played by mRNA 3' processing in mRNA maturation and function, its benefits to transcription is often underestimated and less studied [20,24].

The human CFIm complex, which comprises CFIm68/CFIm59 and CFIm25, was initially identified as a basic subunit of canonical 3' processing complex [3,25,26]. Recent studies suggest that it serves as an activator of canonical mRNA 3' processing and is a master regulator of alternative polyadenylation (APA) [27–31]. Accumulating evidences have indicated that this complex might play a role in gene transcription. Firstly, CFIm, together with CPSF and CstF, can be cross-linked with transcription initiation region for transcribed genes [9,32–34]. Secondly, researchers used RNAPII ChIP-seq to reveal transcription changes in a subset of genes following depletion of CFIm [35,36]. To date, nothing is known on whether the CFIm complex can directly regulate

CONTACT Chengguo Yao  yaochguo@mail.sysu.edu.cn  Center for Stem Cell Biology and Tissue Engineering, Key Laboratory for Stem Cells and Tissue Engineering, Ministry of Education, Sun Yat-Sen University, Guangzhou 510080, China

#These authors contributed equally to this work

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/15476286.2022.2071025>

© 2022 Sun Yat-Sen University. Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

transcription of genes. This is, at least partially, due to a global APA shift upon CFIm depletion in previously reported cell systems [27,30,31,37–40], and the effect of CFIm depletion on gene transcription may be neglected.

Recent studies have implicated CFIm25, the key component of the CFIm complex, in development of multiple cancer types and determination of cell fate [39,41–43]. Given the primary molecular function of CFIm25 in regulation of mRNA 3' processing and APA, most of the reported CFIm25-associated cellular phenotypes have been attributed to its role in PAS choice of target genes thus far [30,37–47]. In the present study, we used CFIm25 knockdown/mutant H9 cell lines, to elucidate an alternative underlying mechanism through which CFIm25 participates in gene regulatory network. Our results revealed that CFIm25 depletion/mutation has little effect on efficiency of cellular mRNA 3' processing and global APA profile in H9 cell lines. Strikingly, disruption of CFIm25 gene expression significantly impacted RNAPII binding, at transcribed genes, and down-regulated transcription output of several key genes associated with the phenotype, including *rex1* gene. Overall, these results reveal a potential role played by CFIm25 in regulation of gene transcription.

Results

Generation of three CFIm25 knockdown/mutation cell lines in hESCs (human embryonic stem cells) using CRISPR/Cas9 technology

CFIm25 was recently shown to be a determinant factor of cell fate in mouse cells [41]. To examine its role in human stem cells, we initially wished to perform CFIm25 gene knockout (KO) in H9 cells, a commonly used human embryonic stem cell line, by using the episomal vector-based CRISPR/Cas9 technology [48]. T7 endonuclease I assays revealed that transfection of target gRNAs resulted in a relatively high mutation rate at the CFIm25 gene locus (Figure S1A; Supplemental source file 1). After antibiotics selection, we picked more than 300 clones for western blot analysis (Supplemental source file 2), and results showed that at least three of them showed gene KO using a CFIm25 primary antibody (sc-81,109, santa cruz) (Fig. 1A; Supplemental source file 3), which recognizes the N-terminus of CFIm25 protein. We noted that our gRNAs were designed near the start codon (Figure S1B), it is possible that these clones may harbour mutation at CFIm25 N-terminus, which may not be targeted by this antibody. Indeed, using another antibody that can recognize the full length CFIm25 (10,322-1-AP, Proteintech), we observed a faint band near 25 kDa for the selected clones (the knockdown efficiency reached approximately 90% for all the three clones) (Fig. 1A; Supplemental source file 3). We further applied DNA sequencing and found that each clone at least one allele has been deleted a multiple of 3 nts (Figure S1B), which results in 12 to 17 amino acids N-terminus deletion proteins. These results are in line with the observation that the molecular weight of the band in mutant cells is slightly smaller than that in control cells using the CFIm25 antibody from Proteintech (Fig. 1A; Supplemental source

file 3). Consistent with previous reports [30,49], we observed that CFIm59, but not CFIm68, showed a mild decrease in expression level upon CFIm25 depletion (Fig. 1A; Supplemental source file 3). Taken together, we presumed that CFIm25 is essential for human cells, and we generated three CFIm25 gene knockdown and small N-terminus deletion mutant clones in H9 cells. For simplicity, they were designated as CFIm25-mutants (CFIm25 m). As mock controls in subsequent experiments, we randomly picked two clones that were transfected with CRISPR/Cas9 empty vector.

CFIm25 regulates growth rate and pluripotency of hESCs

During cell culture, we observed that all of these mutant cell lines grew faster than mock control cells. Therefore, we decided to decipher the potential physiological roles of CFIm25 in hESCs using these cells. Cell proliferation assessment using Cell Counting Kit-8 (CCK-8) showed that CFIm25 mutation caused a significant increase in cell proliferation rate in all three clones (Fig. 1C). It is important to note that the growth rate of human stem cell is largely dependent on the starting cell density, we repeated the same experiment with another starting cell density, and the results showed similar trend (Figure S1C). This observation is consistent with previous reports that CFIm25 knock-down increased the rate of cell proliferation in multiple cancer cell lines [30,43], and suggests that ESCs could be sharing a common feature with cancer cells. Notably, since all three mutant clones were obtained following transfection of three independent gRNAs (Figure S1B), the observed phenotype may not be due to potential indirect effects caused by gRNA off-targeting. Consequently, we combined all three into one dataset (CFIm25-m) for simplicity, it not indicated otherwise, owing to the high similarity among phenotypes and deep sequencing results (Fig. 1C; Supplemental Table 2 and 3).

To elucidate the role of CFIm25 in proliferation of hESCs, we performed a rescue experiment by re-expressing CFIm25 in the mutant cells (CFIm25-m3) using a lentivirus-mediated gene overexpression system. Western blotting revealed that CFIm25 expression was restored to a level comparable to that of endogenous protein (Fig. 1B; Supplemental source file 4), thereby re-establishing the cell proliferation phenotype (Fig. 1C). Additionally, results from cell cycle analysis showed that mutation in CFIm25 significantly shortened the G1 phase and lengthened the G2 phase during cell cycle progression (Fig. 1D; Figure S1D). Overall, these results demonstrated that CFIm25 plays an active role in proliferation of hESCs.

Next, we investigated whether CFIm25 might affect hESCs self-renewal capacity and differentiation potential, two main features characteristic of ESCs. To test this, we first performed qRT-PCR analysis targeting a panel of canonical pluripotency and differentiation markers in both mock control and CFIm25-mutant hESCs. Results revealed no significant changes in expression of the tested pluripotency markers (Figure S1E and 1F). This is consistent with the observation from cell morphology analysis (Figure S1G), and suggests that mutations in CFIm25 might not affect self-renewal capacity of hESCs. In keeping with this, the expression of marker genes

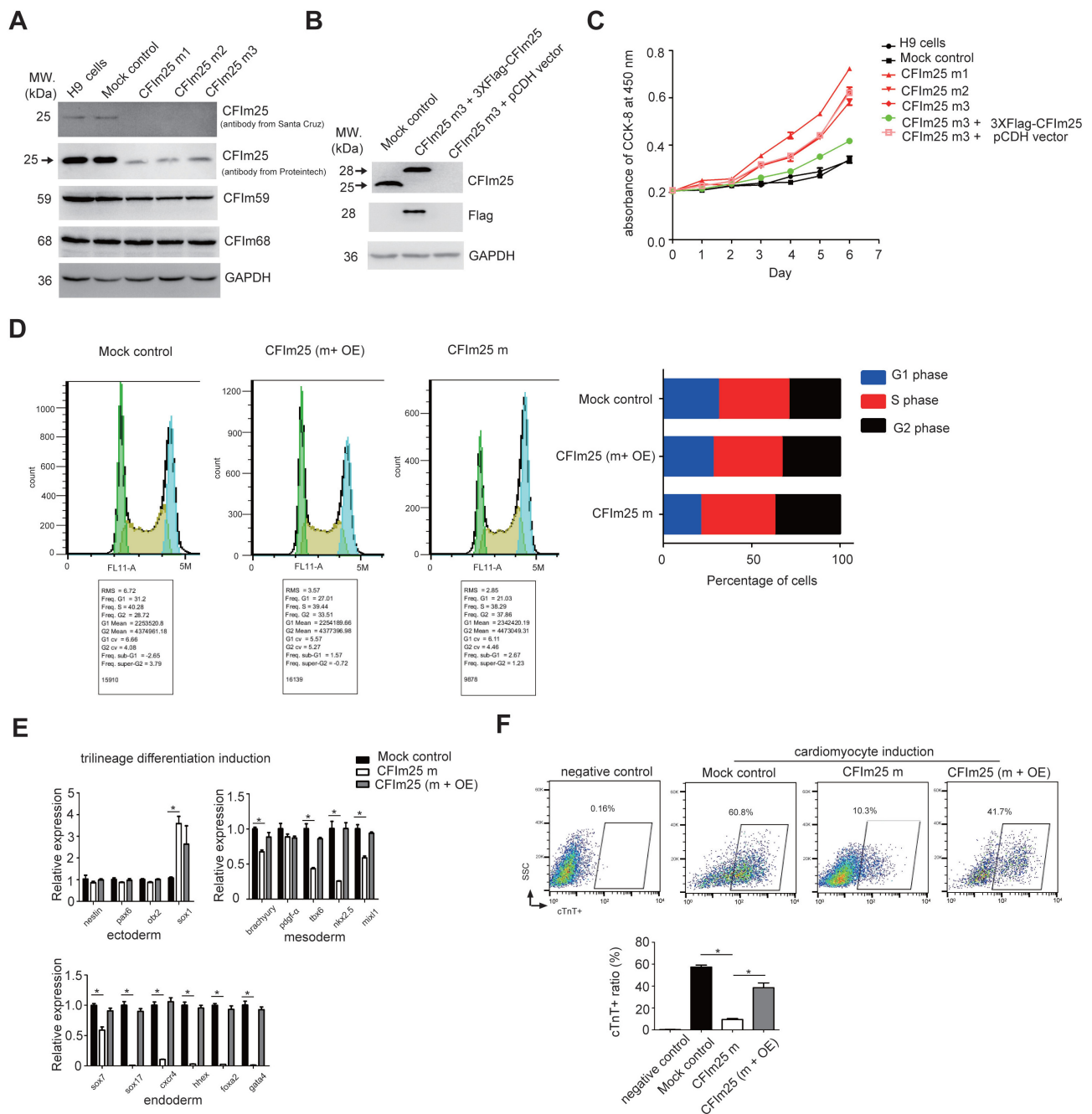


Figure 1. CFim25 knockdown/mutation impacts the cell proliferation rate and differentiation potential of H9 cell line. (A) Western blot analysis of CFim25 and CFim59/68 proteins in cell lysates prepared from two controls and three CFim25 gene-edited H9 cell lines. GAPDH serves as sample loading control (Control: H9 cells; Mock control: eCRISPR empty vector-transfected H9 cells; CFim25 m1-3: eCRISPR-CFim25 gRNAs transfected H9 cells). The primary antibody 1 against CFim25 is from Santa Cruz Company, and antibody 2 is from Proteintech. (B) Western blot analysis of CFim25 and indicated protein (peptide) in cell lysates prepared from Mock, CFim25-m, and CFim25-m plus 3XFlag-CFim25 over-expression cells. At least three independent experiments have been carried out and representative images are shown. The primary antibody for CFim25 is from Santa Cruz (sc-81,109). (C) Cell proliferation rate measurement by CCK-8 kit for indicated cell lines. The growth rate of hESCs is largely dependent on starting cell density. The starting cell density in this experiment is 5000 per well of 96 well plates. Three independent experiments have been carried out and representative results are shown. (D) Flow Cytometry analysis of cell cycle using a Propidium iodide Flow Cytometry Kit in the indicated cell lines (m: mutant; OE: overexpression). The right panel shows the representative result of the percent-ages of cells during different stages of cell cycle. The quantification of three independent experiments is shown in Figure S1D. (E) RT-qPCR analysis of the expression level of corresponding lineage differentiation markers in indicated cell lines during trilineage differentiation. Three independent experiments have been carried out and quantified. Student's t-test was used to estimate the significance of the change. * $p < 0.05$. ns: non-significant. (F) Quantification of the yield of cardiomyocytes by performing fluorescence-activated cell sorting (FACS) analysis in the indicated cell lines during cardiomyocytes differentiation from three independent experiments. cTnT antibodies were used in FACS experiment. Bottom panel is the quantification from three representative experiments (m: mutant; OE: overexpression). Student's t-test was used to estimate the significance: * $p < 0.05$.

across the three germ layers remained undetectable (Supplemental Table 1). Additionally, we used a well-defined Trilineage Differentiation Kit followed by qRT-PCR analysis of differentiation markers to compare the differentiation potential of mock control and CFIm25-mutant hESCs. Strikingly, mutation in CFIm25 appeared to interfere with endoderm, and to a lesser extent, mesoderm differentiation, as evidenced by downregulation of all tested endoderm markers as well as some in the mesoderm markers (Fig. 1E).

To validate the positive role played by CFIm25 in hESC mesoderm/endoderm differentiation, we used a Cardiomyocyte Differentiation kit to generate cardiomyocytes from mock control and CFIm25-mutant hESCs, owing to the fact that cardiomyocyte specification requires both primitive endoderm and nascent mesoderm [50,51]. Cardiomyocyte induction resulted in a ~ 5-fold decrease in efficiency of CFIm25-mutants, as evidenced by the percentages of cardiac troponin T-positive (cTnT+) cells (Fig. 1F). Additionally, cardiomyocytes derived from mock control hESCs exhibited spontaneous beating on day 15, whereas less activity was observed in those from CFIm25-mutant hESCs (Supplemental Video 1 and 2). CFIm25 re-expression in CFIm25-mutant hESCs increased the induction efficiency by about 4 fold (Fig. 1F), further affirming CFIm25's role in cardiomyocytes. Taken together, these results indicated that CFIm25 regulates cell proliferation and differentiation potential in H9 cell line.

CFIm25 regulates mRNA expression level in a subset of genes in hESCs

Next, we explored the molecular mechanisms through which CFIm25 regulates hESCs proliferation and pluripotency. Given its function in choice of polyadenylation site (PAS) and APA regulation [28,41,43], we hypothesized that the observed phenotype is, at least in part, caused by aberrant APA profile in CFIm25-mutant cells. To test this, we characterized global polyadenylation profiles in CFIm25-mutant hESCs alongside controls via high-throughput mRNA 3' end sequencing. Unexpectedly, results showed that mutations in CFIm25 induced insignificant APA changes in the hESC transcriptome (Fig. 2A; Supplemental Table 2), in contrast with what has previously been reported [30,37,41,42,45,46]. As can be seen in Figure S2A, two well documented CFIm25 APA target, *cyclinD1* and *dicer1* genes, did not show apparent APA shift in CFIm25-mutant cells. Previous studies have shown that CFIm25-mediated APA regulation is associated with enhanced canonical PAS processing [31]. Therefore, we compared the overall canonical PAS processing efficiency between CFIm25-mutant and mock control hESCs using a luciferase reporter assay that is applied elsewhere [31,52–54]. Results indicated that mutations in CFIm25 did not lower the processing activity of SVL PAS (Figure S2B), a widely used canonical PAS in the mRNA processing field. These results suggest that CFIm25-mutant is sufficient to support canonical mRNA 3' processing in cells. To verify these findings, we performed a SVL PAS RNA-biotin based pull-down assay followed by western blot analysis using nuclear extracts (NEs) in CFIm25-mutant and mock control hESCs, and

observed that core 3' processing factors, such as CFIm68, Fip1 and CPSF30, were pulled down with similar efficiency (Figure S2C; Supplemental source file 5). As negative controls, much less proteins were detected in the pull-down sample using SVL PAS RNA mutant, which harbours a point mutation at the core AAUAAA hexamer (Figure S2C). Consistently, the SVL PAS RNA mutant showed little PAS processing activity in vivo (Figure S2B).

To further unravel the mechanisms underlying the cellular phenotype, we performed RNA-seq analysis to determine differential expression of genes between CFIm25-mutant and control H9 cells. At a cut-off value of $P < 0.05$ and fold change >1 , we found a total of 587 differentially expressed genes between the groups, of which 277 and 310 were down-regulated and up-regulated, respectively. On the other hand, 99 and 129 genes were down-regulated and up-regulated, respectively, at a cut-off value of $P < 0.05$ and fold change >2 (Fig. 2B; Supplemental Table 3). Additionally, both RNA-seq and PAS-seq approaches were efficient in quantification of gene/isoform expression, as evidenced by good agreement between respective results (Fig. 2C; Figure S2D-E). Gene ontology analysis of the 99/277 down-regulated genes revealed significantly enrichment of genes involved in cellular differentiation, as well as development and negative regulation of growth (Fig. 2D), which is consistent with the earlier results on phenotypes in CFIm25-mutant cells (Fig. 1C-1F). In contrast, enrichment analysis of the 128/309 up-regulated genes revealed no terms associated with cellular proliferation or differentiation (Supplemental Table 3).

To further validate the RNA-seq results, we performed qRT-PCR analysis targeting 11 down-regulated genes, and found consistent expression patterns (Fig. 2C, 2E; Figure S2E). Analysis of RNA-seq data from the aforementioned CFIm25 over-expression and mock control hESCs revealed that CFIm25 re-expression restored expression of most down-regulated genes (Fig. 2E; Figure S2F, S2G; Supplemental Table 4), suggesting that CFIm25 may be playing a direct role in regulating expression of these transcripts. Based on these findings, we hypothesized that CFIm25 might be regulating expression of a subset of cellular proliferation/differentiation-associated transcripts independent of its canonical role in promoting the processing of canonical PASs in hESCs.

CFIm25 promotes *rex1* gene expression at the transcription level

Given that *rex1* is a well-established pluripotency marker and its expression showed the most significant change upon CFIm25 gene editing (Fig. 2E) [55,56], we further wished to understand how CFIm25 promotes *rex1* gene expression in hESCs. We considered several hypotheses. Firstly, *rex1* PAS 3' processing might not be efficient in CFIm25-mutant cells, and may cause transcription read-through at the PAS region as well as subsequent mRNA decay; secondly, CFIm25 might protect *rex1* mRNA from degradation and promote its stability in the nucleus; thirdly, CFIm25 may promote *rex1* gene transcription. To test the first hypothesis, we compared the levels of extended transcript beyond *rex1* PAS via qRT-PCR (Fig. 3A), and observed that CFIm25 mutations did not

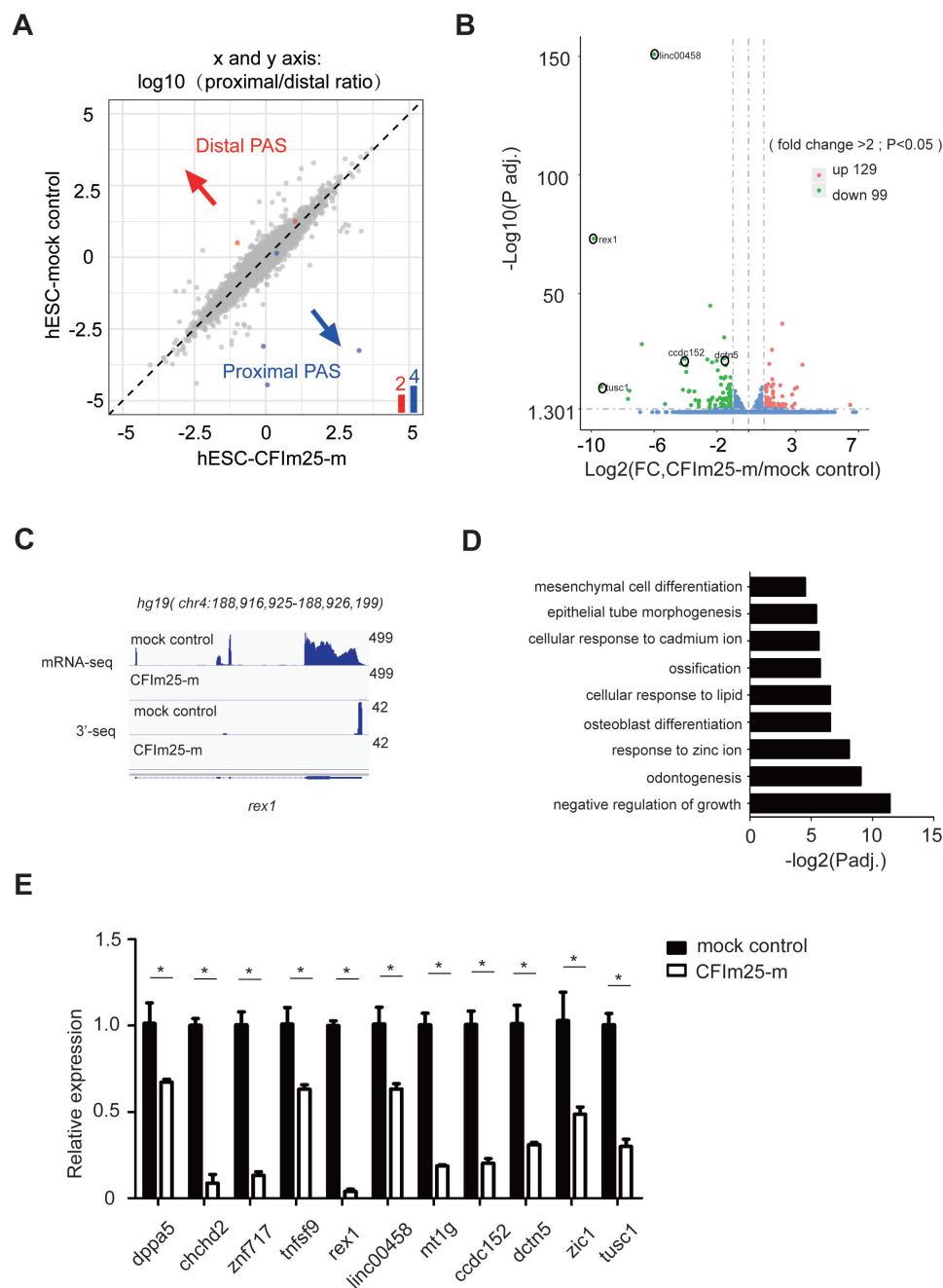


Figure 2. Effect of CFim25 knockdown/mutation on the global mRNA alternative polyadenylation (APA) profile and expression level of poly(A+) transcripts in H9 cell line. (A) mRNA alternative polyadenylation (APA) change in mock and CFim25-m(1–3) H9 cell lines. 3'-seq analysis of APA in mock and CFim25-m H9 cells, \log_2 (proximal/distal ratio) are plotted for mock (y-axis) and CFim25-m H9 cells (x-axis). Statistically significant changes are highlighted in blue (distal to proximal shift) and red (proximal to distal shift). The numbers of APA changes are shown in the column graph. (B) Volcano plot showing the expression level change of poly(A+) mRNA in mock and CFim25-m (1–3) H9 cells. Significant changes ($p < 0.05$, fold change >2) were coloured red (up-regulated in CFim25-m cells in comparison to mock cells) or green (down-regulated in CFim25-m cells in comparison to mock cells), blue dots shows the changes either not statistically significant ($p > 0.05$) or less reliable (fold change <2). Genes for subsequent studies are circled. (C) IGV track screen shots showing mRNA-seq and 3'-seq results for *rex1* gene in mock and CFim25-m H9 cells. (D) Gene ontology analysis of the group of down-regulated genes (277 genes) upon CFim25-m using the Gene Ontology Consortium platform (<http://geneontology.org/>). Gene ontology terms (y axis) and corresponding p-values (x axis) are shown. (E) RT-qPCR analysis of the expression level of indicated genes in mock and CFim25-m (1–3) H9 cells. The results of three independent experiments have been quantified. Student's t-test was used to estimate the significance of the change. * $P < 0.05$.

increase the yield of read-through transcript at PAS region (Fig. 3B). To test the second scenario, we measured the half-life of *rex1* mRNA by first treating cells with Actinomycin D (Act D), followed by qRT-PCR analysis. Similarly, we found no marked difference between mock control and CFim25-mutant cells within 2-hour periods (Figure S3A). In fact,

rex1 mRNA was hardly detectable in CFim25-mutant cells following longer time Act D treatment. At least three lines of evidences suggest that CFim25 regulates *rex1* gene expression at the transcriptional level. Firstly, qRT-PCR-based comparison of *rex1* pre-mRNA expression, using primers targeting its intronic region, revealed a fold change that was comparable to

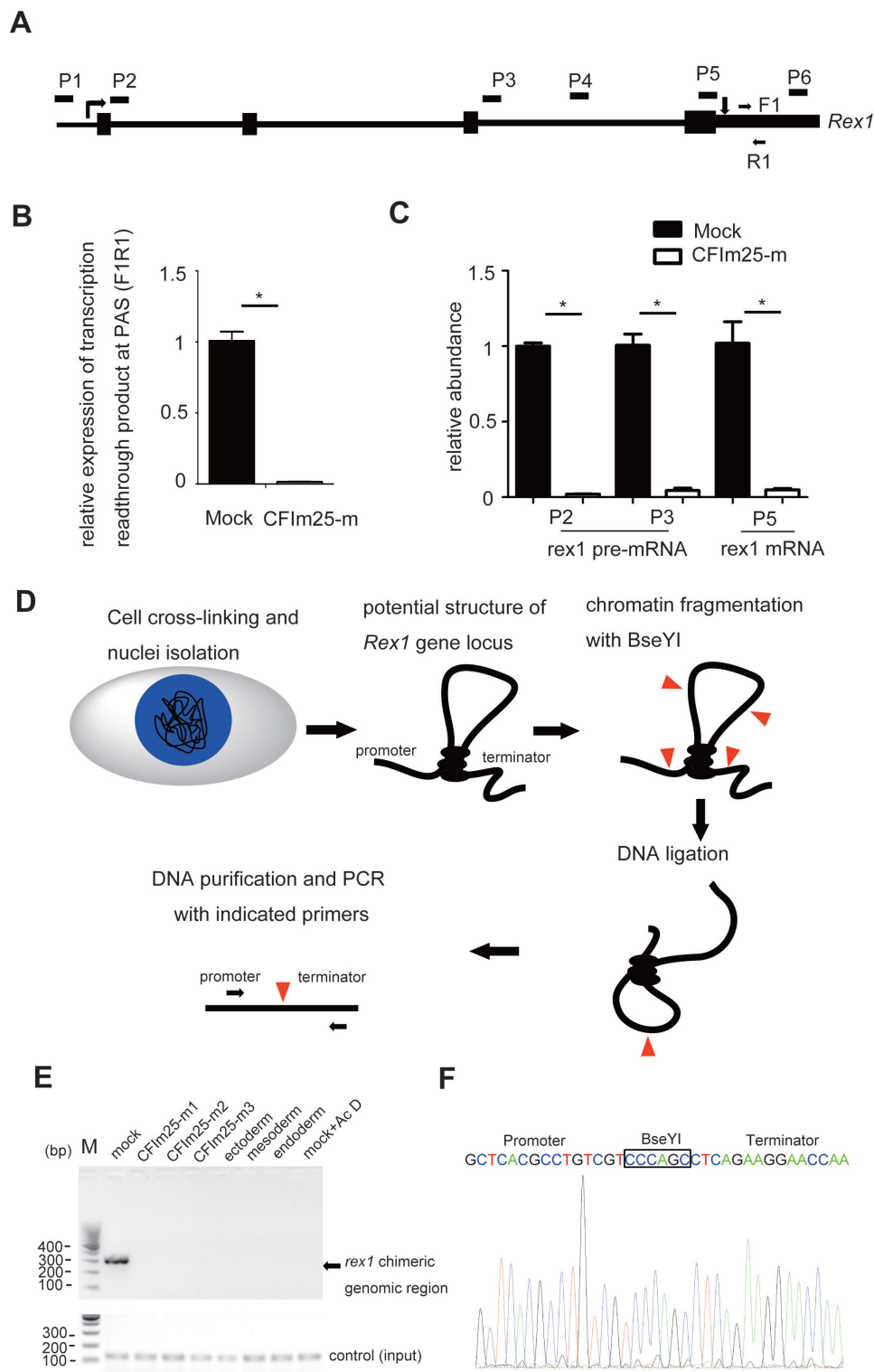


Figure 3. CFIm25 knockdown/mutation impacts *rex1* gene transcription in H9 cells. (A-B) A pair of primers (F1/R1) was designed to detect the gene transcription readthrough beyond *rex1* PAS. The relative expression of extended transcript in mock and CFIm25-m H9 cells was estimated by RT-qPCR analysis shown in Figure 38. *Gapdh* gene expression serves as internal control. Student's t-test was used to estimate the significance of the change. * $p < 0.05$. (C) RT-qPCR analysis of the expression level of *rex1* pre-mRNA and mRNA using indicated primers in mock and CFIm25-m cells. Student's t-test was used to estimate the significance of the change. * $P < 0.05$. (D) Outline of the 3C procedure used to detect chromatin interactions between promoter and terminator region for *rex1* gene. (E) PCR product resulting from 3C library amplification using the primers located in the promoter and terminator regions. PCR product targeting specific internal coding region of *rex1* gene serves as input. Cells used for 3C library preparation are indicated above the gel image. (F) Sanger sequencing shows that the PCR products correspond to the ligated DNA fragments of the two regions located at the promoter and terminator of *rex1*.

that of *rex1* mRNA expression (Fig. 3C). Secondly, we enriched nascent RNAs by purifying chromatin-associated RNAs and subsequently quantified their expression via qRT-PCR using the aforementioned primers. We found consistent results, evidenced by lower levels of *rex1* pre-mRNAs in CFIm25-mutant, relative to control cells (Figure S3B). Thirdly, we directly enriched nascent RNAs via metabolic pulse-chase labelling of RNA using bromouridine (BrU), then purified them with anti-BrU antibody. As expected, we obtained similar results after qRT-PCR (Figure S3C).

Given the essential role played by a promoter in gene transcription, we hypothesized that CFIm25 might be regulating activity of *rex1*'s promoter. To test this, we cloned *rex1* gene promoter into pGL3-basic plasmid, then measured its luciferase activity. Results showed no significant differences in luciferase activities between control and CFIm25-mutant cells (Figure S3D), suggesting that other elements might be involved in CFIm25's role in regulating *rex1* transcription initiation/elongation.

Previous studies have shown that in yeast, some mRNA 3' end processing factors may regulate transcription by bridging the interaction between the promoter and terminator regions of specific genes, a phenomenon termed gene looping [16–19]. In the present study, we used chromatin conformation capture (3C) analysis to test this model in *rex1* gene, based on following observations. Firstly, mouse *rex1* gene locus is characterized by long-range DNA–DNA interactions [57]. Secondly, ChIP-qPCR results suggest that CFIm25 is moderately enriched at both ends of the *rex1* gene in H9 cells (Figure S3E). Therefore, we constructed 3C libraries by digesting nuclei prepared from control and CFIm25-mutant cells with BseYI restriction enzyme. Then, we applied DNA ligation with T4 DNA ligase and PCR amplification targeting the indicated genomic sites (Fig. 3D). BseYI enzyme was chosen as both *rex1* gene promoter and terminator region harbour this restriction enzyme site. Results revealed clear band, indicative of a genomic interaction between the *rex1* promoter and terminator regions, in control H9 cell lines, but not in CFIm25-mutant cells (Fig. 3E; Supplemental source file 6). However, the amount of DNA input, which was amplified from a specific internal coding region of *rex1* gene between two BseYI sites, was comparable in the parallel experiments. Result of sanger sequencing analysis confirmed that the amplified PCR product were similar to those obtained near *rex1* promoter and terminator regions (Fig. 3F). Additionally, we showed that this PCR product is dependent on formaldehyde-mediated crosslinking and DNA ligation (Figure S3F; Supplemental source file 6), indicating the PCR band may reflect the chromatin interaction.

Next, we evaluated whether the detected *rex1* promoter/terminator interaction was correlated with its expression. Strikingly, results from 3C-PCR analysis revealed no significant interaction in promoter/terminator interaction across differentiated cells expressing low levels of *rex1* transcript (Fig. 3E; Figure S3G; Supplemental source file 6), suggesting that this gene looping may be associated with gene expression. Similarly, treatment of the cells with transcription inhibitor Act D significantly abolished the observed interaction (Fig. 3E; Supplemental source file 6). Overall, these results

indicated that CFIm25 might promote *rex1* expression in hESCs, in part, by facilitating formation of gene looping, a chromatin conformation status associated with transcription activation or enhancement.

Gene looping was reported to affect promoter directionality [16], we further examined if this could also be applied for *rex1* gene. In contrast to previous report, we failed to detect apparent expression of anti-sense transcript near *rex1* gene promoter in both control and CFIm25 mutant cells based on mRNA-seq data and RT-qPCR analysis (Fig. 2C; Supplemental Table 1), indicating that the *rex1* gene looping regulated by CFIm25 might not be related with promoter directionality.

CFIm25 significantly affects gene transcription dynamics in hESCs

We hypothesized that CFIm25 could be promoting expression of other mRNA targets via transcriptional mechanisms. To this end, we performed RNA polymerase II (RNAPII) ChIP-seq in two mock control and CFIm25-mutant hESCs (the mutant for all the ChIP-seqs refers to m3, two biological replicates, unless otherwise noted), and found that CFIm25 globally promoted RNAPII's association with actively transcribed genomic region, as evidenced by the metagene plots for all expressed genes (FPKM>1, 14,274 genes based on mRNA-seq) and the group of down-regulated genes (Fig. 4A; Figure S4A). This trend was pronounced for both high- and low-abundance genes (Figure S4A), suggesting that CFIm25-regulated gene transcription might be a general phenomenon. A common change, as shown in representative genome browser view of *gapdh* gene, involved a mild decrease in the signal at transcription start site (TSS) and gene body following gene editing of CFIm25 (Figure S4B). We noted that RNAPII ChIP-seq signal near TES decreased, rather than increased, in CFIm25-mutant cells (Fig. 4A), further supporting the aforementioned conclusion that overall PAS processing efficiency was not affected, as inefficient PAS processing often leads to retarded transcription termination and RNAPII accumulation downstream of transcription end site (TES) [58]. Next, we performed peak calling by MACS2 and identified differential binding events using DiffBind package. As expected, we identified thousands of differential binding sites, with a majority of them located in intron (Fig. 4B; Supplemental Table 5). Consistently, 22% of the genes showing expression level changes upon CFIm25 gene editing harbour differential RNAPII binding sites (Fig. 4B). To validate the ChIP-seq results, we randomly selected nine regions, subjected them to ChIP-qPCR (three biological replicates), and observed consistent trend for most of the sites (Fig. 4C). Thus, we presumed that CFIm25 might significantly regulate the expression of mRNAs at the transcription level. Additionally, the RNAPII binding on the group of up-regulated genes (310 genes) appeared to be less affected by CFIm25 gene editing (Figure S4A). However, this might not be significant, because up-regulated genes intrinsically require more RNAPII binding to produce more transcripts.

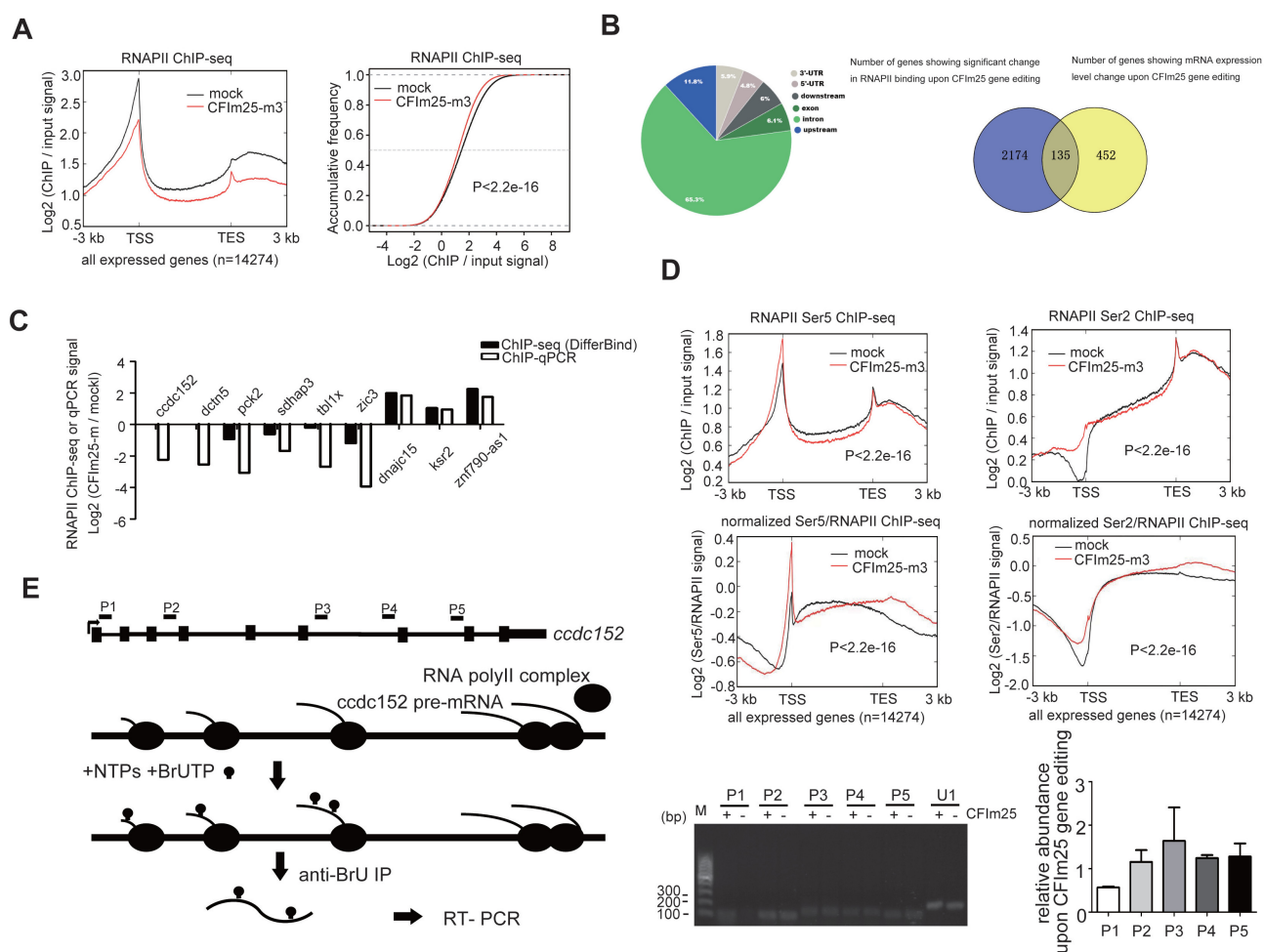


Figure 4. CFIm25 knockdown/mutation globally impacts transcription dynamics in H9 cells. (A) (left panel) Metagenes plots of RNAPII ChIP-seq reads in mock and CFIm25-m H9 cells for actively transcribed genes (FPKM > 1 based on mRNA-seq, 14,274 genes in total), and its corresponding cumulative frequency plot (right panel). K-S test was used to examine the significance of the difference between the two plots. (B) (left panel) Pie plot showing the genomic annotations of 4024 sites that displayed differential RNAPII binding. Peak calling was performed using MACS2 software and DiffBind package was used to identify the differential binding events. (right panel) Venn diagram showing the numbers of overlapping and non-overlapping genes that displayed differential RNAPII binding and mRNA expression level change. (C) Comparison of RNAPII ChIP-seq and ChIP-qPCR results in mock and CFIm25-m H9 cells for the tested genomic sites. Y axis represents the average fold changes from replicates (ChIP-seq: two replicates; ChIP-qPCR: three replicates). (D) Metagenes plots of RNAPII Ser5 ChIP-seq and RNAPII Ser2 ChIP-seq reads for actively transcribed genes in mock and CFIm25-m H9 cells. K-S test was used to examine the significance of the difference between the two plots. Bottom panels show the signals normalized to RNAPII. (E) Nuclear run-on assay on the nascent *ccdc152* transcript. The gene structure and the primer positions are indicated on the top. The diagram for the nuclear run-on assay is shown in the middle. A representative set of RT-PCR data are shown in the bottom panel. Left gel image: nuclear run on assays followed RT-PCR using primers targeting P1-P5 region. 'CFIm25⁻' represents CFIm25-m cell nuclei, whereas 'CFIm25⁺' represents mock cell nuclei. Right Bar graph represents RT-qPCR data from three independent experiments. U1 snRNA was assayed as normalization control. Student's t-test was used to estimate the significance of the change. *P < 0.05.

Given that *rex1* is a transcription-related pluripotency factor, whose expression is down-regulated in CFIm25-mutant cells, we further investigated whether the aberrant RNAPII occupancy might be caused by its depletion. Here, we generated a stable cell line expressing *rex1* shRNA, and subsequently analysed it using RNAPII ChIP-seq assay. Results from qRT-PCR and western blot analysis revealed that *rex1* was moderately depleted (Figure S4C; Supplemental source file 7). Significantly, rather than detecting a decrease, we observed an increase in RNAPII ChIP-seq signal in *rex1* RNAi cells (Figure S4C). Furthermore, the same bioinformatics pipeline and statistical analysis revealed no presence of differential binding sites, which is in contrast with results from RNAPII ChIP-seq in CFIm25-mutant cells (Supplemental Table 5). Additionally, we observed that other CFIm25 mRNA targets (84 down-regulated genes and 129 up-

regulated genes upon CFIm25 gene editing) exhibited no clear general transcription-associated molecular functions (Supplemental Table 3). Thus, we combined these observations with the aforementioned data in which CFIm25 overexpression rescued the gene expression phenotype, and concluded that CFIm25 may be directly responsible for the observed transcription effect.

To further validate this finding, we performed ChIP-seq analysis using antibodies against RNAPII Ser5/Ser2, two modification status associated with transcription initiation/elongation/termination [59–61]. Consistent with the finding from previous studies that RNAPII Ser5 has major binding peak at TSS and Ser2 signal gradually increases towards TES [61], our results revealed similar patterns in hESCs. Results from both metagenes plot and differential peak identification revealed potential transcription initiation/elongation disturbance at

the transcriptome level (Fig. 4D; Figure S4D; Supplemental Table 5). Further normalization of Ser5/Ser2 signal to total RNAPII signal also confirmed this point (Fig. 4D). The genome browser views for two representative genes (down-regulated upon CFIm25 gene editing) were shown in Figure S4E. Taken together, our results suggest that CFIm25 might regulate gene transcription.

To identify additional evidence supporting CFIm25's role in regulating gene transcription, we performed Cleavage Under Targets and Tagmentation (CUT&Tag) analysis, an improved ChIP-seq method, using antibodies against CFIm25. Strikingly, in addition to the peak observed at TES, we detected a sharp peak at TSS in the CFIm25 binding profiles (Figure S4F). Consistent with the knowledge that core 3' processing factors associate with RNAPII during transcription [7], we observed positive CFIm25 binding signal throughout the gene body for all expressed genes, including the group of down-regulated/up-regulated genes (Figure S4F). Peak calling analysis showed that CFIm25 has at least one significant binding peak for 3735 genes, with 1242 genes harbouring at least one peak near TSS region (-2 kb to 2 kb) (Supplemental Table 5). For these 1242 genes, 202 of them have at least one RNAPII differential binding peak. These results indicate that the presence of CFIm25 may directly affect RNAPII binding for specific group of genes. Notably, although results from both CUT&Tag analysis are often limited by non-specificity of target antibodies, our findings in CFIm25 are reliable owing to prior normalization of the signals by backgrounds from CFIm25-mutant cells.

Furthermore, we performed a nuclear run-on assay, which provides a measure of transcription and minimizes the effect of RNA stability, and analysed expression levels of two target pre-mRNAs by qRT-PCR, these two genes were selected owing to their differential mRNA expression as well as effect on their transcription processes as shown by RNAPII/Ser5 ChIP-seq and RNAPII ChIP-qPCR results (Fig. 4C; Figure S4E). Notably, we detected low levels of transcription product near the promoter region in CFIm25-mutant cells, but not in the middle or at the end point of these genes (Fig. 4E; Figure S4H; Supplemental source file 8). Taken together, these results suggest that CFIm25 may be playing a cellular role in the early stages of gene transcription in specific genes, including *ccdc152* and *dctn5*.

To validate our findings, we analysed a randomly selected dataset in which CFIm25 was depleted in human cancer cell line [62]. In addition to APA change, we observed differential expression of thousands of transcripts upon CFIm25 depletion (Figure S4I; Supplemental Table 6). Interestingly, these two groups of genes did not show a striking overlap, further indicating that CFIm25 might have functions other than PAS usage.

CFIm25 might affect gene transcription through its association with LEO1

Next, we explored the mechanism through which CFIm25 regulates transcription. Since some splicing factors could regulate transcription process through their interaction with general transcription factors [63,64], we hypothesized that

CFIm25 might also utilize such a mechanism to regulate transcription. To this end, we took advantage of the aforementioned 3XFLAG-CFIm25 H9 cell line and performed anti-FLAG immunoprecipitation (IP) assays followed by mass spectrometry (MS) analysis. As expected, CFIm68 and CFIm59, two known CFIm25 interaction partners, were highly enriched in the FLAG IP sample based on cell lysates prepared from hESCs overexpressing FLAG-CFIm25 (Supplemental Table 7). LEO1, an RNAPII associated factor, was selected among the candidates owing to its direct association with transcription [65].

MS results were further confirmed by western blot analysis (Fig. 5A; Supplemental source file 9). As our FLAG IP/MS was carried out in the absence of ribonuclease, it is possible that the detected interaction was mediated by RNAs. To confirm a direct protein-protein interaction, we performed aforementioned FLAG IP in the presence of RNase A to avoid RNA-mediated effects. Western blot analysis revealed similar result to that without RNase A treatment, indicating a direct association of CFIm25 with LEO1 (Figure S5A; Supplemental source file 9). To further confirm this, we carried out GST-pull down assays using recombinant GST-tagged CFIm25 protein and His-tagged LEO1 protein (Figure S5B; Supplemental source file 10). Western blot analysis suggested that LEO1 indeed is able to associate with GST-CFIm25, but not control GST proteins (Fig. 5B). Furthermore, we observed that LEO1 C-terminus fragment, but not other fragments (truncation fragment 4 and 5 were apparently detectable in western blotting analysis using the anti-His Tag antibody, whereas their expression were not readily detectable using Coomassie blue staining), showed detectable association with CFIm25 under physiological conditions, as shown by the western blot analysis (Fig. 5C; Supplemental source file 10).

Inspired by the above result, we further tested the associations of LEO1 C-terminus fragment with several CFIm25 N-terminus mutants. Three mutants were designed to mimic the three small N-terminus deletion/mutant proteins produced by CFIm25 gene-edited cells (Figure S1B; Figure S5B; Fig. 5D). Strikingly, we observed that none of these three mutants were able to associate with LEO1 C-terminus truncation fragment, in comparison with wild type GST-CFIm25 (Fig. 5D; Supplemental source file 11), providing evidence that CFIm25 may associate with LEO1 through its N-terminus. It must be noted, nevertheless, that this interaction is relatively weak in vitro based on the pull-down efficiency (Fig. 5B-D).

To explore the functional impact of CFIm25-LEO1 association, we carried out ChIP-seq analysis on LEO1 in CFIm25-mutant alongside control hESCs, and found that LEO1 exhibited a significant decrease in binding frequency on transcribed genes, including the group of down-regulated genes, upon CFIm25 gene editing (Fig. 5E), although the overall ChIP efficiency is relatively lower than that of RNAPII (Figure 4A and 5E). Interestingly, the decrease trend seemed more obvious for high-abundance genes (Figure S5C), as shown in representative genome browser view of *gapdh* gene (Figure S5D). This observation is in agreement with aforementioned finding that RNAPII occupancy is globally down-regulated in CFIm25-mutant hESCs

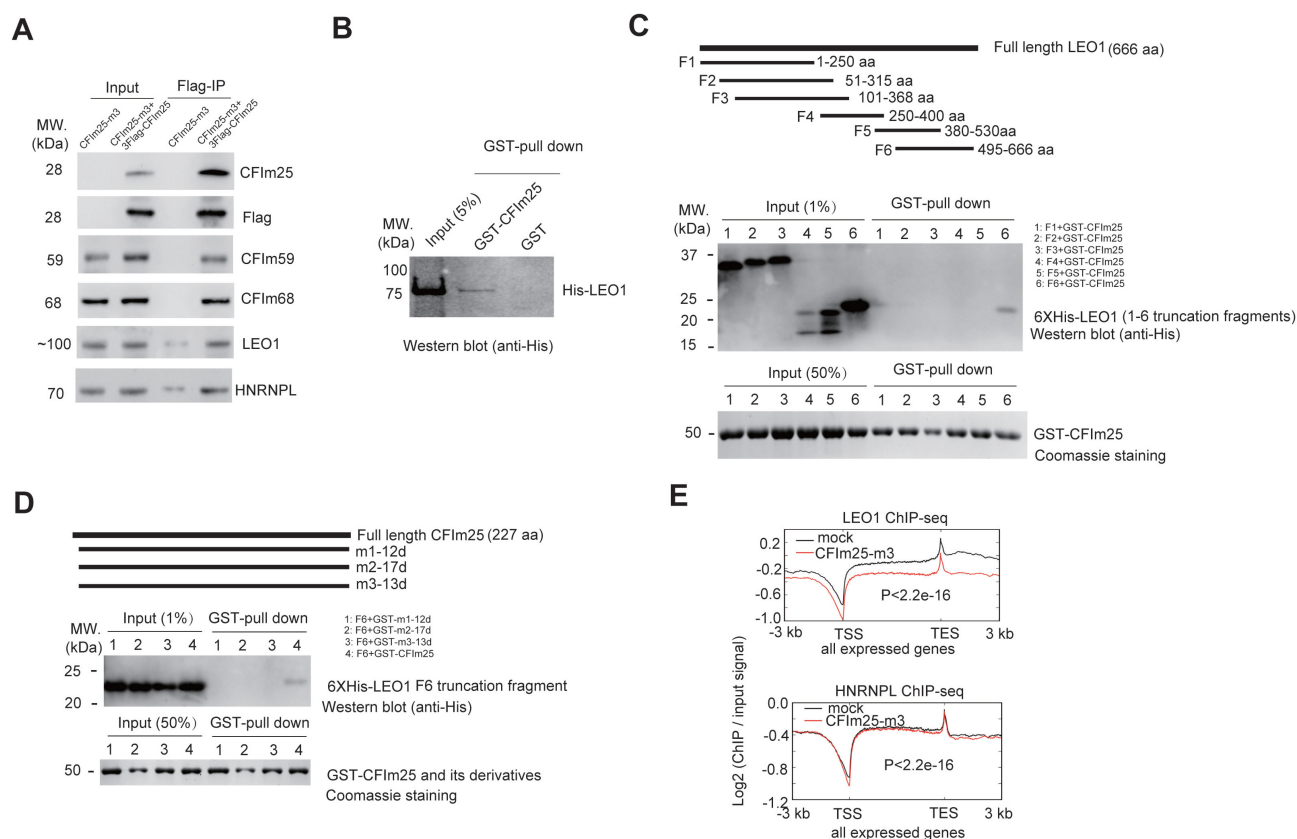


Figure 5. CFIm25 interacts with LEO1 and impacts the DNA genomic binding profiles of LEO1. (A) Western blot analysis of the abundance of indicated proteins in Flag-IPed sample. Flag-IP was performed using extracts from CFIm25-m and CFIm25 (m + 3XFlag-CFIm25 overexpression) H9 cells. Input: 1% of the lysates for IP. The primary antibody for CFIm25 is from Santa Cruz (sc-81,109). (B) GST pull-down assay using recombinant GST/GST-CFIm25 and His-LEO1 protein. The pull-down efficiency was estimated with western blot analysis using the anti-His antibody. (C) Schematic representation of full length human LEO1 protein and truncation fragments (F1-F6) (upper). Truncation fragments were fused to pET-28a vector (BamHI and XhoI) for recombinant His-tag protein expression. At least three independent experiments have been performed, and a representative western blotting result of GST-pull-down assay using anti-His antibody is shown in the middle picture. As the bait protein, recombinant GST-CFIm25 was stained with Colloidal Coomassie G-250 (Bottom). The percentage of input is indicated in the bracket. (D) Schematic representation of human CFIm25 protein in CFIm25-mutants (m1-12d/m2-17d/m3-13d) represent the 12/17/13 amino acids deletion/mutation proteins produced in CFIm25-mutant (m1-m3) cells respectively (upper). Full length and mutant CFIm25 proteins were fused to pGEX-4T3 vector (BamHI and XhoI) for recombinant GST-tag protein expression. In the GST pull-down assay, recombinant His-tag LEO1-F6 protein was used as the prey protein. At least three independent experiments have been performed, and a representative western blotting result of GST-pull-down assay using anti-His antibody is shown in the middle picture. GST-fused bait proteins are stained with Colloidal Coomassie G-250. The percentage of input is indicated in the bracket. (E) Meta-gene plots of LEO1 ChIP-seq and HNRNPL ChIP-seq reads for actively expressed genes in mock and CFIm25-m cells. K-S test was used to examine the significance of the difference between the two plots.

(Fig. 4A). It is important to point out that the input samples gave approximately the same signal in our RNAPII and LEO1 ChIP-seqs (Figure S4B; Figure S5D), and thus the detected discrepancy between ChIP samples did not appear to be caused by DNA heterogeneity in input samples. In contrast, the overall DNA binding pattern of HNRNPL, another protein that has potential interaction with CFIm25 (Fig. 5A; Supplemental Table 7; Supplemental source file 9), showed no apparent change in CFIm25-mutant cells (Fig. 5E). The LEO1/HNRNPL ChIP-seq results were further validated by ChIP-qPCR, by amplifying two specific genomic regions of *gapdh* gene. As shown in Figure S5D, the ChIP efficiency of LEO1 was increased by more than 5 fold at both tested sites in control cells, whereas HNRNPL did not show significant change. Importantly, antibodies against negative IgG control showed extremely low affinity with the same targets, indicating both LEO1 and HNRNPL ChIP-seq signals were above the background level. Taken together, these results suggest that CFIm25 potentially affects the genomic binding pattern of its

associated transcription factor LEO1, thereby providing a potential mechanism underlying CFIm25-mediated transcription regulation.

CFIm25 targets associate with the phenotypes of CFIm25 gene editing in hESCs

The above results suggest that CFIm25 may affect gene transcription process and enhance expression of a subset of mRNA targets. To understand the effect of CFIm25 gene editing on cellular phenotypes, we used overexpression and knockdown experiments on several high-confidence CFIm25 targets, then analysed the resulting cellular phenotypes. Strikingly, depletion of *rex1* significantly impaired differentiation of the endoderm in hESCs, as evidenced by down-regulation of endoderm-lineage markers following differentiation induction (Fig. 6A). Notably, we found no significant changes in expression of most of the tested self-

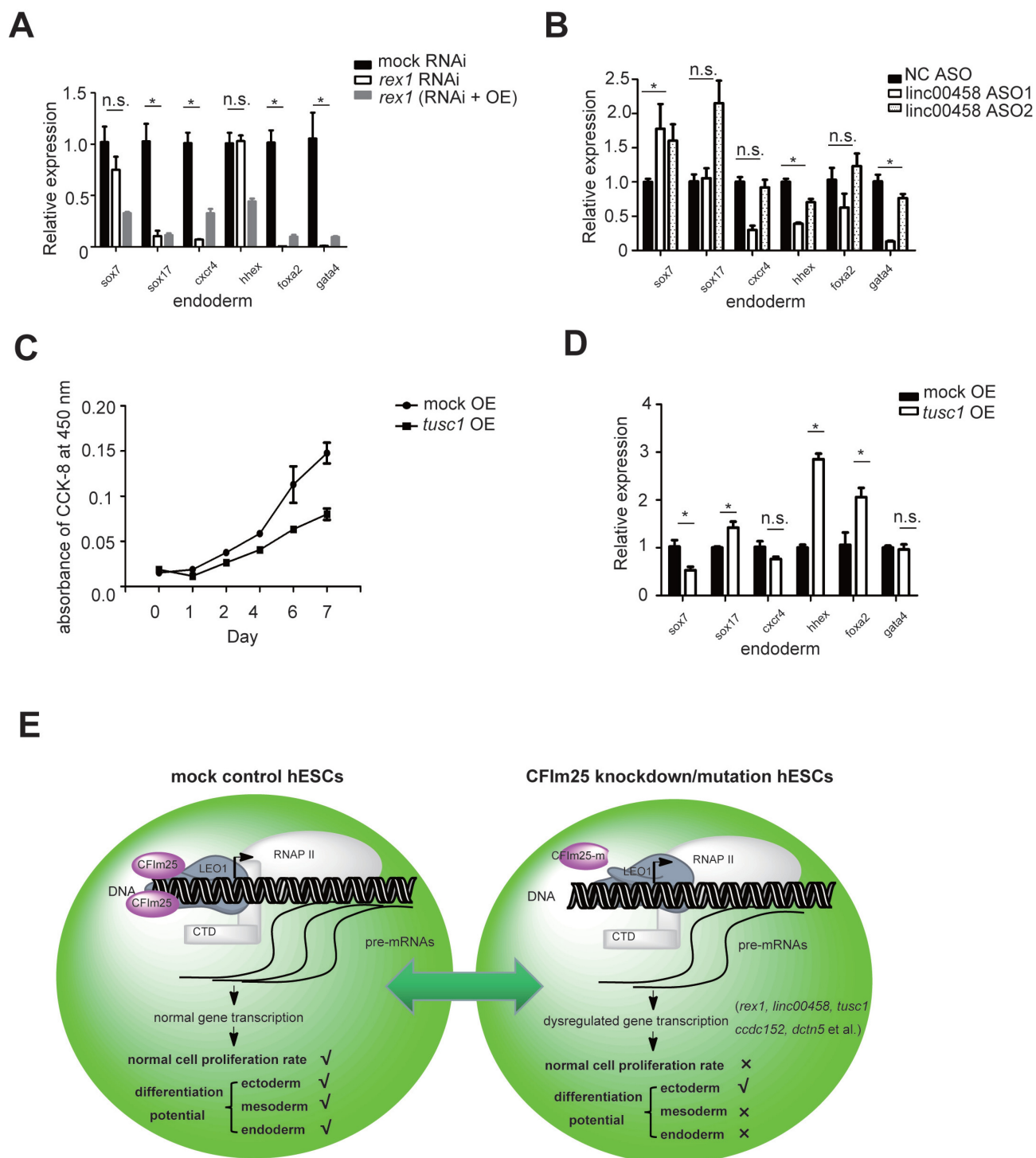


Figure 6. CFIm25 targets play roles in hESCs cell proliferation and pluripotency. (A-B; D) RT-qPCR analysis of the expression level of 6 endoderm lineage differentiation markers in indicated cell lines during endoderm lineage differentiation (OE: overexpression; ASO: antisense oligo). Student's t-test was used to estimate the significance of the change. * $P < 0.05$; n.s.: non-significant. (C) Cell proliferation rate measurement by CCK-8 kit for mock and *tusc1* gene overexpression H9 cell lines. (E) A schematic model summarizing the key finding in this study. Human CFIm25 protein might be generally impacting gene transcription in H9 cells, thereby enhancing the expression level of a group of transcripts associated with pluripotency (such as *rex1* gene) and cell proliferation (such as *tusc1* gene). CFIm25 depletion/mutation predominantly caused defects in the endoderm/mesoderm differentiation and accelerated the rate of cell growth in H9 cells.

renewal markers and the cell morphology in both CFIm25-mutant and *rex1* RNAi cells (Figure S6A and B), suggesting that *rex1* depletion did not affect self-renewal of H9 cells. These results are consistent with the findings of previous reports in which mouse *rex1* was reportedly dispensable for self-renewal of ES cells [55]. In fact, knocking it out in mouse

ES cells was implicated in impaired differentiation of the visceral endoderm [55]. Furthermore, we carried out *rex1* gene overexpression in *rex1* RNAi hESCs and performed parallel experiments. Indeed, *rex1* overexpression could partially rescue the differentiation potential phenotype induced by *rex1* depletion (Fig. 6A).

Since *rex1* depletion could not fully recapitulate the endoderm lineage differentiation phenotype caused by CFIm25 gene editing (Fig. 1E; Fig. 6A), we tested the function of another CFIm25 high-confidence target, *linc00458*, a long noncoding RNA that has been associated with endodermal lineage specification [66]. Results showed that *linc00458* knockdown using Antisense Oligonucleotides (ASO) technology significantly down-regulated endoderm-specific genes *gata4* and *hhex* during induction of endoderm differentiation (Fig. 6B). Overall, these results suggest that the observed phenotype in hESCs lacking CFIm25 might be caused by synergistic effects of CFIm25 mutation in target genes.

We further tested the function of several other targets that might be associated with cell proliferation phenotype upon CFIm25 gene editing. As expected, overexpression of *tuscl1*, a tumour-associated suppressor gene [67], caused an apparent suppression in the rate of cell proliferation (Fig. 6C). When cells overexpressing *tuscl1* gene were subjected to endoderm differentiation, they appeared to be dysregulated during this process, as evidenced by the expression of molecular markers (Fig. 6D). This result was consistent with the findings from previous reports in which some tumour suppressor genes were found to play a crucial role in ESCs pluripotency [68,69]. Taken together, our results indicate that phenotypes of CFIm25-mutant hESCs result from down-regulation of a subset of CFIm25-regulated RNA transcripts.

Discussion

In the field of co-transcriptional mRNA processing, most previous reports studying co-transcriptional mRNA 3' processing have focused on how transcription facilitates mRNA 3' processing, and the effect of 3' processing on mRNA alternative polyadenylation (APA). In this study, we present evidence that CFIm25, a canonical mRNA 3' processing factor, may promote gene transcription in H9 cell line, and the mechanism might be involved in its interaction with LEO1, an RNAPII associated factor. Importantly, CFIm25 as well as its targets plays a direct role in H9 cell function. A schematic model is presented in Fig. 6E. Our findings not only provide novel insights into the critical role played by CFIm25 (and possibly other 3' processing factors) in gene regulation, aside from its traditionally studied function in mRNA 3' processing and APA regulation, but also expand our understanding of its role in determination of cell fate.

Researchers have long hypothesized that mRNA 3' processing factors may be playing a role in transcription. For example, the co-purification of CPSF with TFIID was discovered more than twenty years ago [10]. Recent studies have shown that CstF64 and CPSF73 regulate RNAPII activity at transcription end sites (TES) [58], and CFIm25/CFIm68 depletion in HeLa cells affects RNAPII occupancy in a subset of genes [35,36]. However, our results are significant in at least two major respects. Firstly, we excluded the possibility that the observed transcription phenotypes might be caused by impaired transcription termination, upon CFIm25 gene editing. Therefore, our findings provide more direct evidence that mRNA 3' processing factor may be playing an active role in early transcription rather than passively interacting with

transcription termination. Secondly, results from our global analyses and nuclear run-on assays for specific genes affirm reliability of our results, while the findings of our CFIm25 overexpression rescue experiment validate the conclusion.

Multiple lines of evidences indicate that CFIm25 might affect transcription elongation. Firstly, ChIP-seq analysis using antibodies against a transcription initiation-associated factor TBP (TATA box binding protein) revealed no marked difference near TSS region (Figure S6C). Secondly, the majority of differential peaks identified from RNAPII ChIP-seq are located in introns (Fig. 4B). Thirdly, RNAPII Ser5/Ser2 signals showed significant changes at the gene body when normalized to total RNAPII (Fig. 4D). Fourthly, CFIm25 could bind throughout gene body based on CUT&Tag analysis (Figure S4F). Lastly, CFIm25 appears to associate with LEO1 (Fig. 5B-D), which is best known for its role in transcription elongation. Given the global effect of RNAPII occupancy on transcribed genes in CFIm25-mutant cells, it remains unclear why CFIm25 gene editing only affected the steady level of a specific subset of genes, as we observed no significant difference in total poly(A+) RNA yield between control and CFIm25-mutant hESCs (Figure S6D). We attribute this to two scenarios. Firstly, the steady levels of mRNAs are controlled by multiple factors, such as transcription, mRNA processing and stability [70], while we cannot rule out existence of unknown mechanisms that regulate this balance in mRNA expression upon CFIm25 gene editing. For example, we noted a slight increase, albeit statistically insignificant, in canonical SVL PAS processing efficiency (Figure S2B). Therefore, it is plausible that the steady levels in a majority of genes with no apparent change in expression might be balanced by decreased transcription and increase in 3' processing efficiency. Secondly, transcription itself is controlled by autoregulatory mechanisms. For example, paused RNAPII reportedly inhibits new transcriptional initiation [71]. In the present study, we used RNAPII ChIP-seq analysis to reveal defects in the observed global transcription. However, the extent to which the occupancy of RNAPII contribute to the transcription output in our system remain unknown.

Further studies are required to fully understand the role of CFIm25 in transcriptional regulation in the context of co-transcriptional mRNA processing. Firstly, although it is unlikely that this phenomenon is unique to hESCs, we cannot fully exclude this possibility. Similar assays in other cell types are imperative to validate these findings and unravel the precise underlying molecular mechanisms. Secondly, previous studies have shown that CFIm25 can regulate global mRNA alternative polyadenylation (APA) in many cell types [30,31,37,39,41–43,45,46], while recent reports demonstrated that it could also regulate mRNA splicing in specific genes [72,73]. Future explorations are expected to reveal whether they are associated with CFIm25's potential role in transcriptional regulation, and to elucidate mechanisms underlying coordination of these multiple regulatory roles. Finally, we envisage that further explorations will generate a deeper understanding of the functional significance of CFIm25-mediated regulation of transcription. Previous studies have shown that CFIm25 plays important cellular roles under normal physiological conditions, while its dysregulation has been

associated with a variety of diseases, such as cancer, learning deficits and dermal fibrosis [30,37,39,41–43,45,46]. Results of the present study corroborated the aforementioned findings, as evidenced by enhanced cell proliferation and impaired differentiation potential in hESCs upon CFIm25 mutation. It is plausible that CFIm25-mediated transcription regulation may also be involved in other reported cellular systems. A key challenge for future investigations is emergence of multiple molecular functions of CFIm25. For example, although CFIm25 might regulate mRNA abundance, splicing and APA for the same group of genes, approaches for delineating their respective contributions to cellular phenotype remain limited. With the growing trend in generating related data, we believe a clearer picture will be painted with regards to the functional significance of CFIm25-mediated regulation in transcription.

Materials and Methods

Cell culture and plasmids transfections

H9 hESCs were purchased from the National Collection of Authenticated Cell Cultures (Shanghai, Catalogue SCSP-302) and were maintained in mTeSR (Stem Cell Technology) on Matrigel-coated plates at 37°C. CRISPR/Cas9-mediated CFIm25 gene editing was carried out using a previously reported eCRISPR plasmid [48], based on the following gRNA target sequences: g1: CAGCCGGTCTGCGAGCGATT, g2: CCGAACTGAGTGACCCCCCG, and g3: CCAATCGCTCGCAGACCGGC. Cultures were selected on puromycin, single-cell clones picked for further expansion. pLKO.puro shRNA vectors were used for *rex1* RNAi, with target sequence GCATGCAAATACGAACAAGAA, while lentiviral plasmids were used for gene overexpression. Briefly, 3XFlag-CFIm25 cDNA was cloned into CD533A-2 pCDH-EF1-MCS-IRES-Neo (SBS), whereas *rex1* and *tuscl* overexpression plasmids were purchased from Fulengen, Catalogues EX-T4815-Lv242; EX-I2275-Lv233. Transfections for plasmids and Antisense oligos (ASOs) were performed using Lipofectamine 3000 and Lipofectamine RNAiMAX (Life Technology), respectively. ASOs targeting *linc00458* were ordered from RiboBio. Cells were harvested at the suggested time points for further analysis upon transfection.

Cell growth measurement and differentiation induction

Cell growth monitoring and analysis of hESCs trilineage differentiation were performed using the Cell Counting Kit-8 (Dojindo) and STEMdiff™ trilineage differentiation kits (Stem Cell Technology), according to the manufacturer's protocols. For cell growth measurement, cells were seeded at 5000 cells per well on 96-well plate at day 0. After the addition of CCK-8 solution, the absorbance at 450 nm was measured using a microplate reader at the indicated time points (day 0, 1, 2, et al.). It is important to note that the growth rate of human embryonic stem cells is sensitive to the quality and

density of starting cells. Therefore, it is essential to keep the starting cell numbers at the same level and make sure tested cells were treated in parallel in this experiment. For stem cell trilineage differentiation induction experiments, cells were seeded in 12-well plate at the suggested cell density. The induction time is approximately within one week. After induction, total RNAs were harvested, and subsequent RT-qPCR analysis of lineage expression markers was carried out to estimate the induction efficiency. Moreover, differentiation of hESCs cardiomyocytes was performed using the STEMdiff™ Cardiomyocyte Differentiation Kit (Stem Cell Technology), whereas analysis of cardiomyocyte induction efficiency was conducted via FACS using the cTnT+ primary antibody (Thermo Scientific, MA5-12,960).

Luciferase reporter assays

hESCs were transfected for 24 h with pPASPORT-SVL PAS or pGL3-basic (promoter sequence inserts)+pRL-TK plasmids, harvested, then subjected to analysis of Luciferase activity using the Promega Dual-Luciferase Reporter kit and Berthold Sirius detection system.

RNA-biotin based pull-down assay

SVL PAS RNA and the corresponding point mutant RNA (CPSF recognition motif 'AAUAAA' hexamer was mutated to 'AACAAA') were made by in vitro transcription using SP6 polymerase, and biotinylated at 3' end using a biotinylation Kit (ThermoFisher). H9 cell nuclear extracts (NEs) were made following the described protocol [3,74]. Approximately 15 µg biotinylated RNAs were first bound to the streptavidin beads, and then incubated with 100 µl pre-cleared NE in the polyadenylation condition [40% NE, 8.8 mM HEPES (pH 7.9), 44 mM KCl, 0.4 mM DTT, 0.7 mM MgCl₂, 1 mM ATP, and 20 mM creatine phosphate] for 20 minutes, after biotin-streptavidine binding, washing, pull-down sample were heated (75°C for 5 minutes) in 1XSSC buffer (150 mM NaCl, 15 mM sodium citrate) for elution. The eluted sample was further subjected to western blot analysis.

Metabolic pulse-chase RNA labelling with bromouridine (BrU)

Cellular pre-mRNA labelling was performed with bromouridine, according to a published protocol [75]. Briefly, hESCs were grown to approximately 50% confluency in 3 10-cm plates, then incubated with bromouridine (final 2 mM), at a pulse time of 30 min. BrU containing pre-mRNA was purified with 2 µg anti-BrdU antibodies (BD Pharmingen) prior to use in downstream RT-qPCR analysis.

Nuclear run-on assay

Nuclear run-on assays were performed using previously described protocol with minor modifications [64,76]. Briefly, 1×10^7 hESCs were permeabilized with digitonin, and nuclei was isolated via low-speed centrifugation. A nuclear run-on reaction was initiated by mixing the nuclei with 60 µl reaction

buffer (50 mM Tris-HCl, pH7.4, 10 mM MgCl₂, 150 mM NaCl, 25% (v/v) glycerol, 0.5 mM PMSF and 25 U ml⁻¹ RNasin) and 40 µl BrU-containing NTPs mixture (1.8 mM ATP, 0.5 mM CTP and GTP, 0.375 mM UTP, 0.125 mM BrU), with a 15-min incubation at 25°C. After the reaction, RNA was extracted from BrU-containing cells using the Trizol reagent (Thermo Scientific), and further isolated by 2 µg anti-BrU antibodies (BD Pharmingen). Purified RNAs were used for downstream RT-qPCR analysis.

Chromosome conformation capture (3C) analysis

A 3C analysis was carried out according to a published protocol [77], with minor modifications. Briefly, 1×10^7 cells were cross-linked in 1% formaldehyde solution and quenched with 125 mM glycine. Cells were permeabilized, their nuclei were isolated via centrifugation then resuspended in a 0.5 ml solution comprising 1.2× restriction enzyme NEBuffer™ r3.1 (NEB) containing 0.3% SDS. After incubation, shaking, and digestion with BseYI enzyme (NEB), the cross-linked chromatin was ligated using T4 DNA ligase (NEB). The DNA was de-crosslinked, purified via phenol/chloroform extraction and ethanol precipitation, then subjected to 3C-PCR analysis using primers listed in supplementary table 8. PCR reactions were set up by mixing: 500 ng of DNA template; 25 pmol of each primer; 5 µl of 10× PCR buffer; 1 µl of 10 mM dNTP mix; 1 µl of 5 U/µl taq DNA polymerase; H₂O to bring the final volume to 50 µl. Run PCR with the following thermal cycling parameters. 1 cycle: 2 min 95°C (initial denaturation); 30 cycles: 30 sec 95°C (denaturation), 30 sec 55°C (annealing), 1 min 72°C (extension); 1 cycle: 4 min 72°C (final extension).

3'-seq, mRNA-seq, ChIP-seq

We performed 3'-seq analysis using QuantSeq Rev 3' mRNA sequencing library prep kit (Lexogen), on the NovaSeq platform. Raw reads were reverse complemented and mapped to the human genome (hg19), allowing up to two mismatches using Bowtie2 with the settings 'bowtie2 -p 28 -N 1 -k 1'. The 3' end of the read maps was considered a poly(A) junction. The bioinformatics analysis for reads filtering and clustering, internal priming removal, poly(A) site identification and subsequent APA analysis shown in Fig. 2A and Figure S4H, were performed essentially as previously described [52,54].

Preparation of mRNA-seq library, sequencing and analysis of sequence data were performed in accordance with the standard protocol described by Illumina and Novogene. Identification of differentially expressed genes was done using the DESeq2 tool.

ChIP-seq libraries were prepared using the ChIP-IT® Express Enzymatic Shearing (Active Motif) and ChIP-seq library preparation (Vazyme) kits. Primary antibodies used for ChIP included RNAPII (39,497, Active Motif); hnRNPL (18,354-1-AP, Proteintech); RNAPII Ser5 (61,986, Active Motif); RNAPII Ser2 (61,984, Active Motif); LEO1 (PAB14102, Abnova); CFIm68 (A301-358A,

Bethyl); CFIm59 (A301-359A, Bethyl); IgG (AC005, Abclonal). CUT & Tag libraries for CFIm25 were prepared using the CUT & Tag Hyperactive In-Situ ChIP Library Prep Kit (Vazyme) and primary antibody (10,322-1-AP, Proteintech). Compared to traditional ChIP-seq, CUT&Tag has the advantage that the entire process from antibody binding to adaptor ligation occurs *in vivo*. These libraries were sequenced on the NovaSeq platform, and all ChIP-seq data processing and analysis performed according to the ENCODE ChIP-seq pipeline (<https://www.encodeproject.org/data-standards/chip-seq/>). Metagene plots were generated with DeepTools2 computeMatrix tool with a bin size of 50 bp and plotProfile - outFileData tool. Graphs representing the (IP/Input) signal (ChIP-seq) were then created with R packages. Metagene profiles are shown as the average of two biological replicates. P-values were computed with a Kolmogorov-Smirnov (K-S) test.

Affinity purification of CFIm25-associated proteins

A total of 10×10^7 hESCs cells that stably overexpress Flag-tag CFIm25 or negative control were harvested by centrifugation at 2000 g for 5 min. Cells were lysed with 3 ml IP lysis buffer (87,787, Thermo Scientific) in the presence of protease inhibitor cocktail (Roche). After incubation at 4°C for 20 min and centrifugation at 15,000 g for 10 min, cell extracts (3 ml of the supernatant) were incubated with Anti-Flag Affinity Gel (Bimake) at 4°C for 3 h. After three washes, each with 1 ml Wash Buffer (30 mM Tris-Cl, pH 7.4, 150 mM NaCl, and 0.5% Triton X-100), proteins were eluted from the beads using elution buffer (30 mM Tris-Cl, pH 7.4, 150 mM NaCl, 0.5% Triton X-100, and 400 µg/ml Poly FLAG peptide). Eluted samples were resolved in an SDS-polyacrylamide gel, followed by mass spectrometry (Mass Spectrometry Facility at Novogene, Beijing). Aliquots of the eluted proteins were used for western blotting.

GST pull-down assay

Human CFIm25 and corresponding N-terminus mutants were inserted into vector pGEX-4T3 and expressed as GST-CFIm25 fusion protein in BL21 (DE3) strain. The fusion protein was purified with ProteinIso® GST Resin (TRANS). LEO1 truncation fragments were inserted into vector pET-28a vector and expressed as His-LEO1 fusion proteins in BL21 (DE3) strain. Full length His-LEO1 DNA construct was purchased from Fulgen (Catalogue: EX-T3600-B31). The fusion proteins were purified with HisPur Cobalt Resin (Thermo Scientific). For GST pull-down assay, two proteins (approximately 10 µM for each) were mixed in binding buffer (20 mM HEPES, pH 7.9, 150 mM NaCl, 1 mM MgCl₂, 0.2 mM EDTA, 0.1% NP-40, protease inhibitor). After binding, washing, proteins were eluted from the beads using elution buffer (20 mM HEPES, pH 7.9, 150 mM NaCl, 1 mM MgCl₂, 0.2 mM EDTA, 0.1% NP-40, protease inhibitor, 20 mM Glutathione). Eluted proteins were used for western blotting or Coomassie blue staining.

Quantitative real time polymerase chain reaction (qRT-PCR) and western blot analysis

Quantitative real-time PCR was performed in 96-well plates, on the LightCycler® 480 qPCR system (Roche). Briefly, RNAs were quantified on a NanoDrop™ 1000 Spectrophotometer (Anti-BrU antibodies purified pre-mRNAs were not quantified due to low yield). The cDNA was synthesized from extracted RNA using the superscript III reverse transcriptase kit (Life Technology). The cDNA was used for qRT-PCR amplification targeting genes outlined in Supplementary Table 8. Expression data were analysed using the $\Delta\Delta C_t$ method, and normalized based on appropriate controls. All the qPCR parameters and results including reaction conditions, input volumes and Ct values, have been listed in MIQE form as Supplemental Table 9. Western blot assay was conducted using standard techniques, with the following primary antibodies; CFIm25 (10,322-1-AP, Proteintech or sc-81,109, Santa Cruz), REX1 (MA5-38,664, Thermo Scientific), CFIm68 (A301-358A, Bethyl), CFIm59 (A301-359A, Bethyl), GAPDH (sc-32,233, Santa Cruz), Flag (HT201-01, TRANS), His (HT501-01, TRANS).

Accession numbers

All the deep sequencing data have been deposited to GEO database with the accession no.GSE178194.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Guangzhou Municipal Science and Technology Project [201803040017]; National Natural Science Foundation of China [31970613].

ORCID

Chengguo Yao  <http://orcid.org/0000-0001-9224-0156>

References

- [1] Colgan DF, Manley JL. Mechanism and regulation of mRNA polyadenylation. *Genes Dev.* 1997;11(21):2755–2766.
- [2] Cramer P. Organization and regulation of gene transcription. *Nature.* 2019;573(7772):45–54.
- [3] Shi Y, Di GDC, Taylor D, et al. Molecular architecture of the human pre-mRNA 3' processing complex. *Mol Cell.* 2009;33(3):365–376.
- [4] Sun Y, Hamilton K, Tong L. Recent molecular insights into canonical pre-mRNA 3'-end processing. *Transcription.* 2020;11(2):83–96.
- [5] Sun Y, Zhang Y, Hamilton K, et al. Molecular basis for the recognition of the human AAUAAA polyadenylation signal. *Proc Natl Acad Sci U S A.* 2018;115(7):E1419–E1428.
- [6] Zhang Y, Sun Y, Shi Y, et al. Structural Insights into the Human Pre-mRNA 3'-End Processing Machinery. *Mol Cell.* 2020;77(4):800–809 e806.
- [7] Bentley DL. Coupling mRNA processing with transcription in time and space. *Nat Rev Genet.* 2014;15(3):163–175.
- [8] Fusby B, Kim S, Erickson B, et al. Coordination of RNA Polymerase II Pausing and 3' End Processing Factor Recruitment with Alternative Polyadenylation. *Mol Cell Biol.* 2016;36(2):295–303.
- [9] Glover-Cutter K, Kim S, Espinosa J, et al. RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nat Struct Mol Biol.* 2008;15(1):71–78.
- [10] Dantonel JC, Murthy KG, Manley JL, et al. Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA. *Nature.* 1997;389(6649):399–402.
- [11] Ji Z, Luo W, Li W, et al. Transcriptional activity regulates alternative cleavage and polyadenylation. *Mol Syst Biol.* 2011;7(1):534.
- [12] Nagaike T, Logan C, Hotta I, et al. Transcriptional activators enhance polyadenylation of mRNA precursors. *Mol Cell.* 2011;41(4):409–418.
- [13] Rosonina E, Bakowski MA, McCracken S, et al. Transcriptional activators control splicing and 3'-end cleavage levels. *J Biol Chem.* 2003;278(44):43034–43040.
- [14] Davidson L, Muniz L, West S. 3' end formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II CTD are reciprocally coupled in human cells. *Genes Dev.* 2014;28(4):342–356.
- [15] Licatalosi DD, Geiger G, Minet M, et al. Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. *Mol Cell.* 2002;9(5):1101–1111.
- [16] Al Husini N, Kudla P, Ansari A. A role for CF1A 3' end processing complex in promoter-associated transcription. *PLoS Genet.* 2013;9(8):e1003722.
- [17] Allepuz-Fuster P, O'Brien MJ, Gonzalez-Polo N, et al. RNA polymerase II plays an active role in the formation of gene loops through the Rpb4 subunit. *Nucleic Acids Res.* 2019;47(17):8975–8987.
- [18] Ansari A, Hampsey M. A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping. *Genes Dev.* 2005;19(24):2969–2978.
- [19] El Kaderi B, Medler S, Raghunayakula S, et al. Gene looping is conferred by activator-dependent interaction of transcription initiation and termination machineries. *J Biol Chem.* 2009;284(37):25015–25025.
- [20] Mapendano CK, Lykke-Andersen S, Kjems J, et al. Crosstalk between mRNA 3' End Processing and Transcription Initiation. *Mol Cell.* 2010;40(3):410–422.
- [21] Kaida D, Berg MG, Younis I, et al. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature.* 2010;468(7324):664–668.
- [22] Ran Y, Deng Y, Yao C. U1 snRNP telescripting: molecular mechanisms and beyond. *RNA Biol.* 2021;18(11):1512–1523.
- [23] Venters CC, Oh JM, Di C, et al. U1 snRNP Telescripting: suppression of Premature Transcription Termination in Introns as a New Layer of Gene Regulation. *Cold Spring Harb Perspect Biol.* 2019;11(2):a032235.
- [24] Cavallaro M, Walsh MD, Jones M, et al. 3' (')-5' (') crosstalk contributes to transcriptional bursting. *Genome Biol.* 2021;22(1):56.
- [25] Ruegsegger U, Beyer K, Keller W. Purification and characterization of human cleavage factor Im involved in the 3' end processing of messenger RNA precursors. *J Biol Chem.* 1996;271(11):6107–6113.
- [26] Ruegsegger U, Blank D, Keller W. Human pre-mRNA cleavage factor Im is related to spliceosomal SR proteins and can be reconstituted in vitro from recombinant subunits. *Mol Cell.* 1998;1(2):243–253.
- [27] Kim S, Yamamoto J, Chen Y, et al. Evidence that cleavage factor Im is a heterotetrameric protein complex controlling alternative polyadenylation. *Genes Cells.* 2010;15(9):1003–1013.
- [28] Kubo T, Wada T, Yamaguchi Y, et al. Knock-down of 25 kDa subunit of cleavage factor Im in HeLa cells alters alternative polyadenylation within 3'-UTRs. *Nucleic Acids Res.* 2006;34(21):6264–6271.
- [29] Martin G, Gruber AR, Keller W, et al. Genome-wide analysis of pre-mRNA 3' end processing reveals a decisive role of human

- cleavage factor I in the regulation of 3' UTR length. *Cell Rep.* **2012**;1(6):753–763.
- [30] Masamha CP, Xia Z, Yang J, et al. CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. *Nature.* **2014**;510(7505):412–416.
- [31] Zhu Y, Wang X, Forouzmeh E, et al. Molecular Mechanisms for CFIm-Mediated Regulation of mRNA Alternative Polyadenylation. *Mol Cell.* **2018**;69(1):62–74 e64.
- [32] Calvo O, Manley JL. Strange bedfellows: polyadenylation factors at the promoter. *Genes Dev.* **2003**;17(11):1321–1327.
- [33] Garrido-Lecca A, Saldi T, Blumenthal T. Localization of RNAPII and 3' end formation factor CstF subunits on *C. elegans* genes and operons. *Transcription.* **2016**;7(3):96–110.
- [34] Katahira J, Okuzaki D, Inoue H, et al. Human TREX component Thoc5 affects alternative polyadenylation site choice by recruiting mammalian cleavage factor I. *Nucleic Acids Res.* **2013**;41(14):7060–7072.
- [35] Tellier M, Hardy JG, Norbury CJ, et al. Effect of CFIm25 knock-out on RNA polymerase II transcription. *BMC Res Notes.* **2018**;11(1):894.
- [36] Tellier M, Hardy JG, Norbury CJ, et al. Effect of CFIm68 knock-down on RNA polymerase II transcription. *BMC Res Notes.* **2019**;12(1):554.
- [37] Alcott CE, Yalamanchili HK, Ji P, et al. Partial loss of CFIm25 causes learning deficits and aberrant neuronal alternative polyadenylation. *Elife.* **2020**;9. DOI:10.7554/eLife.50895.
- [38] Sommerkamp P, Altamura S, Renders S, et al. Differential Alternative Polyadenylation Landscapes Mediate Hematopoietic Stem Cell Activation and Regulate Glutamine Metabolism. *Cell Stem Cell.* **2020**;26(5):722–738 e727.
- [39] Tan S, Li H, Zhang W, et al. NUDT21 negatively regulates PSMB2 and CXXC5 by alternative polyadenylation and contributes to hepatocellular carcinoma suppression. *Oncogene.* **2018**;37(35):4887–4900.
- [40] Weng T, Huang J, Wagner EJ, et al. Downregulation of CFIm25 amplifies dermal fibrosis through alternative polyadenylation. *J Exp Med.* **2020**;217(2). DOI:10.1084/jem.20181384.
- [41] Brumbaugh J, Di Stefano B, Wang X, et al. Nudt21 Controls Cell Fate by Connecting Alternative Polyadenylation to Chromatin Signaling. *Cell.* **2018**;172(1–2):106–120 e121.
- [42] Chu Y, Elrod N, Wang C, et al. Nudt21 regulates the alternative polyadenylation of Pak1 and is predictive in the prognosis of glioblastoma patients. *Oncogene.* **2019**;38(21):4154–4168.
- [43] Najaf Abadi MH J, Shafabakhsh R, Asemi Z, et al. CFIm25 and alternative polyadenylation: conflicting roles in cancer. *Cancer Lett.* **2019**;459:112–121.
- [44] Gennarino VA, Alcott CE, Chen CA, et al. NUDT21-spanning CNVs lead to neuropsychiatric disease and altered MeCP2 abundance via alternative polyadenylation. *Elife.* **2015**;4. DOI:10.7554/eLife.10782.
- [45] Huang J, Weng T, Ko J, et al. Suppression of cleavage factor Im 25 promotes the proliferation of lung cancer cells through alternative polyadenylation. *Biochem Biophys Res Commun.* **2018**;503(2):856–862.
- [46] Weng T, Ko J, Masamha CP, et al. Cleavage factor 25 deregulation contributes to pulmonary fibrosis through alternative polyadenylation. *J Clin Invest.* **2019**;129(5):1984–1999.
- [47] Zhou Z, Qu J, He L, et al. Stiff matrix instigates type I collagen biogenesis by mammalian cleavage factor I complex-mediated alternative polyadenylation. *JCI insight.* **2020**;5(3): e133972. DOI:10.1172/jci.insight.133972.
- [48] Xie Y, Wang D, Lan F, et al. An episomal vector-based CRISPR/Cas9 system for highly efficient gene knockout in human pluripotent stem cells. *Sci Rep.* **2017**;7(1):2320.
- [49] So BR, Di C, Cai Z, et al. A Complex of U1 snRNP with Cleavage and Polyadenylation Factors Controls Telescripting, Regulating mRNA Transcription in Human Cells. *Mol Cell.* **2019**;76(4):590–599 e594.
- [50] Rowton M, Guzzetta A, Rydeen AB, et al. Control of cardiomyocyte differentiation timing by intercellular signaling pathways. *Semin Cell Dev Biol.* **2021**;118:94–106.
- [51] Ruan H, Liao Y, Ren Z, et al. Single-cell reconstruction of differentiation trajectory reveals a critical role of ETS1 in human cardiac lineage commitment. *BMC Biol.* **2019**;17(1):89.
- [52] Lackford B, Yao C, Charles GM, et al. Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. *EMBO J.* **2014**;33(8):878–889.
- [53] Shi J, Deng Y, Huang S, et al. Suboptimal RNA-RNA interaction limits U1 snRNP inhibition of canonical mRNA 3' processing. *RNA Biol.* **2019**;16(10):1448–1460.
- [54] Yao C, Biesinger J, Wan J, et al. Transcriptome-wide analyses of CstF64-RNA interactions in global regulation of mRNA alternative polyadenylation. *Proc Natl Acad Sci U S A.* **2012**;109(46):18773–18778.
- [55] Masui S, Ohtsuka S, Yagi R, et al. Rex1/Zfp42 is dispensable for pluripotency in mouse ES cells. *BMC Dev Biol.* **2008**;8(1). DOI:10.1186/1471-213X-8-45
- [56] Son MY, Choi H, Han YM, et al. Unveiling the critical role of REX1 in the regulation of human stem cell pluripotency. *Stem Cells.* **2013**;31(11):2374–2387.
- [57] Zhang S, Deng T, Tang W, et al. Epigenetic regulation of REX1 expression and chromatin binding specificity by HMGNs. *Nucleic Acids Res.* **2019**;47(9):4449–4461.
- [58] Nojima T, Gomes T, Grosso ARF, et al. Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell.* **2015**;161(3):526–540.
- [59] Hsin JP, Manley JL. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev.* **2012**;26(19):2119–2137.
- [60] Lyons DE, McMahon S, Ott M. A combinatorial view of old and new RNA polymerase II modifications. *Transcription.* **2020**;11(2):66–82.
- [61] Tellier M, Zaborowska J, Caizzi L, et al. CDK12 globally stimulates RNA polymerase II transcription elongation and carboxyl-terminal domain phosphorylation. *Nucleic Acids Res.* **2020**;48(14):7712–7727.
- [62] Routh A, Ji P, Jaworski E, et al. Poly(A)-ClickSeq: click-chemistry for next-generation 3-end sequencing without RNA enrichment or fragmentation. *Nucleic Acids Res.* **2017**;45(12):e112.
- [63] Caizzi L, Monteiro-Martins S, Schwab B, et al. Efficient RNA polymerase II pause release requires U2 snRNP function. *Mol Cell.* **2021**;81(9):1920–1934 e1929.
- [64] Lin S, Coutinho-Mansfield G, Wang D, et al. The splicing factor SC35 has an active role in transcriptional elongation. *Nat Struct Mol Biol.* **2008**;15(8):819–826.
- [65] Xie Y, Zheng M, Chu X, et al. Paf1 and Ctr9 subcomplex formation is essential for Paf1 complex assembly and functional regulation. *Nat Commun.* **2018**;9(1):3795.
- [66] Chen YF, Li YJ, Chou CH, et al. Control of matrix stiffness promotes endodermal lineage specification by regulating SMAD2/3 via lncRNA LINC00458. *Sci Adv.* **2020**;6(6): eaay0264.
- [67] Shan Z, Shakoobi A, Bodaghi S, et al. TUSC1, a putative tumor suppressor gene, reduces tumor cell growth in vitro and tumor growth in vivo. *PLoS One.* **2013**;8(6):e66114.
- [68] Fu X, Wu S, Li B, et al. Functions of p53 in pluripotent stem cells. *Protein Cell.* **2020**;11(1):71–78.
- [69] Langer LF, Ward JM, Archer TK. Tumor suppressor SMARCB1 suppresses super-enhancers to govern hESC lineage determination. *Elife.* **2019**;8. DOI:10.7554/eLife.45672
- [70] Slobodin B, Bahat A, Sehwat U, et al. Transcription Dynamics Regulate Poly(A) Tails and Expression of the RNA Degradation Machinery to Balance mRNA Levels. *Mol Cell.* **2020**;78(3):434–444 e435.
- [71] Shao W, Zeitlinger J. Paused RNA polymerase II inhibits new transcriptional initiation. *Nat Genet.* **2017**;49(7):1045–1051.

- [72] Gao CC, Xu QQ, Xiao FJ, et al. NUDT21 suppresses the growth of small cell lung cancer by modulating GLS1 splicing. *Biochem Biophys Res Commun.* 2020;526(2):431–438.
- [73] Scarborough AM, Flaherty JN, Hunter OV, et al. SAM homeostasis is regulated by CFIm-mediated splicing of MAT2A. *Elife.* 2021;10. DOI:10.7554/eLife.64930
- [74] Huang C, Shi J, Guo Y, et al. A snoRNA modulates mRNA 3' end processing and regulates the expression of a subset of mRNAs. *Nucleic Acids Res.* 2017;45(15):8647–8660.
- [75] Paulsen MT, Veloso A, Prasad J, et al. Use of Bru-Seq and BruChase-Seq for genome-wide assessment of the synthesis and stability of RNA. *Methods.* 2014;67(1):45–54.
- [76] Roberts TC, Hart JR, Kaikkonen MU, et al. Quantification of nascent transcription by bromouridine immunocapture nuclear run-on RT-qPCR. *Nat Protoc.* 2015;10(8):1198–1211.
- [77] El Kaderi B, Medler S, Ansari A. *Curr Protoc Cell Biol*; 2012. Chapter 22: Unit22.15, Analysis of interactions between genomic loci through Chromosome Conformation Capture (3C); DOI:10.1002/0471143030.cb2215s56.