

DATA NOTE

Open Access



Effect of CFIm25 knockout on RNA polymerase II transcription

Michael Tellier^{*} , Jessica G. Hardy, Chris J. Norbury and Shona Murphy

Abstract

Objectives: Transcription of eukaryotic protein-coding genes by RNA polymerase II (pol II) is a highly regulated process. Most human genes have multiple poly(A) sites, which define different possible mRNA ends, suggesting the existence of mechanisms that regulate which poly(A) site is used. Poly(A) site selection may be mediated by cleavage factor I (CFIm), which is part of the cleavage and polyadenylation (CPA) complex. CFIm comprises CFIm25, CFIm59 and CFIm68 subunits. It has been documented that the CPA complex also regulates pol II transcription at the start of genes. We therefore investigated whether CFIm, in addition to its role in poly(A) site selection, is involved in the regulation of pol II transcription.

Data description: We provide genome-wide data of the effect of reducing by 90% expression of the CFIm25 constituent of CFIm, which is involved in pre-mRNA cleavage and polyadenylation, on pol II transcription in human cells. We performed pol II ChIP-seq in the presence or absence of CFIm25 and with or without an inhibitor of the cyclin-dependent kinase (CDK)9, which regulates the entry of pol II into productive elongation.

Keywords: RNA polymerase II, CFIm25, Transcription, Transcription termination, Cleavage and polyadenylation complex

Objective

The production of a eukaryotic protein-coding mRNA requires the recognition of a specific poly(A) site sequence at the end of the gene. More than half of all human genes contain more than one poly(A) site with evidence of widespread regulation of gene expression through alternative polyadenylation [1]. Poly(A) site recognition is essential for pre-mRNA cleavage and polyadenylation and requires around 85 proteins [2]. Four multi-subunits complexes are essential for pre-mRNA cleavage: cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), and cleavage factors I (CFIm) and II (CFIIm) [3]. The role of CFIm in cleavage is still unclear but this complex binds 40–50 nt upstream of the poly(A) site [4]. CFIm comprises two CFIm25 subunits, which binds RNA, and two larger subunits, CFIm59 and CFIm68 [5, 6].

Previous studies have shown that depletion of CFIm25 or CFIm68 promotes proximal poly(A) site usage and thus a shortening of the 3'untranslated region (3'UTR) of many mRNAs [7–9]. This suggests that CFIm normally promotes recognition of the distal poly(A) site. Misregulation of CFIm has been linked to both tumorigenicity of glioblastoma and some neuropsychiatric diseases through changes to mRNAs 3'UTR length [10, 11]. Proteins involved in pre-mRNA cleavage, such as the CPSF complex, regulate pol II activity at the beginning and end of the transcription cycle [12]. To determine if depletion of CFIm25 also affects pol II transcription, we used a CRISPR/Cas9 approach to reduce the expression of CFIm25 and performed pol II ChIP-seq in the absence or presence of a CDK9 inhibitor, which is the kinase regulating pol II entry into productive elongation [13]. Understanding the function of CFIm in pol II transcription could provide insights into transcriptional changes when CFIm is misregulated. Our data should be of interest to the scientific community working on pol II transcription and co-transcriptional processes.

*Correspondence: michael.tellier@path.ox.ac.uk
Sir William Dunn School of Pathology, University of Oxford, South Park
Roads, Oxford OX1 3RE, UK



Table 1 Overview of data files

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	293 DMSO Input	Fastq.gz (raw files), bigwig (processed files)	ENA accession number (fastq.gz): PRJNA490093 GEO accession number (bigwig): GSE119712
Data file 2	293 DMSO Pol II	Fastq.gz (raw files), bigwig (processed files)	ENA accession number (fastq.gz): PRJNA490093 GEO accession number (bigwig): GSE119712
Data file 3	293 DRB Input	Fastq.gz (raw files), bigwig (processed files)	ENA accession number (fastq.gz): PRJNA490093 GEO accession number (bigwig): GSE119712
Data file 4	293 DRB Pol II	Fastq.gz (raw files), bigwig (processed files)	ENA accession number (fastq.gz): PRJNA490093 GEO accession number (bigwig): GSE119712
Data file 5	CFIm25KO DMSO Input	Fastq.gz (raw files), bigwig (processed files)	ENA accession number (fastq.gz): PRJNA490093 GEO accession number (bigwig): GSE119712
Data file 6	CFIm25KO DMSO Pol II	Fastq.gz (raw files), bigwig (processed files)	ENA accession number (fastq.gz): PRJNA490093 GEO accession number (bigwig): GSE119712
Data file 7	CFIm25KO DRB Input	Fastq.gz (raw files), bigwig (processed files)	ENA accession number (fastq.gz): PRJNA490093 GEO accession number (bigwig): GSE119712
Data file 8	CFIm25KO DRB Pol II	Fastq.gz (raw files), bigwig (processed files)	ENA accession number (fastq.gz): PRJNA490093 GEO accession number (bigwig): GSE119712

Data description

HEK293 cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM, Sigma) supplemented with 10% fetal bovine serum (FBS, Gibco) and 100 units/ml penicillin + 100 µg/ml streptomycin (Gibco). Two of the three copies of the *CPSF5* gene that encodes CFIm25 were knocked out using CRISPR/Cas9 gene editing and confirmed by sequencing of the edited *CPSF5* locus and by western blotting with an antibody against CFIm25 (NUDT21 10322-1-AP, rabbit polyclonal, ProteinTech), which indicated an approximately 90% reduction in CFIm25 expression in the CFIm25KO cells. HEK293 and CFIm25KO cells were treated prior to ChIP-seq with DMSO or 100 µM DRB (Sigma) for 30 min (Table 1).

ChIP was performed as previously described [14]. Briefly, 293 and CFIm25KO cells were crosslinked at room temperature with 1% formaldehyde and quenched with 125 mM glycine for 5 min. Nuclear extracts were sonicated twice for 15 min at high amplitude, 30 s ON/30 s OFF using a Bioruptor (Diagenode). 80 µg of chromatin was incubated overnight at 4 °C with 2 µg of an antibody against IgG (sc-2027, Santa Cruz) as an IP negative control or against pol II (sc-899X, Santa Cruz). After recovery of immune complexes with BSA-saturated protein G Dynabeads and extensive washes, crosslinks were reversed by incubation at 65 °C for 5 h. After ethanol precipitation and proteinase K treatment, DNA was purified using a PCR Purification Kit (Qiagen). ChIP samples were analysed by deep sequencing using Illumina HiSeq 4000 75 bp paired-end reads (Wellcome Trust Centre for Human Genetics, University of Oxford).

To analyse data, adapters were trimmed with Cutadapt v. 1.9.1 [15] with the following constant parameters: --minimum-length 10 -q 15, 10--max-n 1. Obtained

sequences were mapped to the human hg19 reference sequence with Bowtie2 v. 2.2.5 [16]. Unmapped reads were removed with SAMtools v. 1.3.1 [17]. Mapped reads were then de-duplicated using Picard to remove PCR duplicates. Bam files were sorted and indexed with SAMtools. The total number of mapped reads was comprised between 33 and 59 million paired end reads. Bigwig files were created after data normalization to Reads Per Genomic Content (RPGC) by employing deepTools2 v. 2.2.4 [18] bamCoverage tool with the following parameters: -bs 10-normalizeTo1 × 2451960000-e-p max.

Limitations

The effect of CFIm25 KD on pol II transcription is not as strong as the effect observed with knock-down of CFIm68, another member of the CFIm complex [8]. The knockdown efficiency of CFIm25 was about 90%, which may not be sufficient to completely abrogate the role of CFIm25 in regulation of pol II transcription. The ChIP-seq was also performed only once and in only one cell line; HEK293.

Abbreviations

Pol II: RNA polymerase II; DRB: 5,6-dichlorobenzimidazole-1-β-D-ribofuranoside; ChIP: chromatin immunoprecipitation; RPGC: reads per genomic content; DMEM: Dulbecco's Modified Eagle's Medium; FBS: fetal bovine serum; 3'UTR: 3' untranslated region; CDK9: cyclin-dependent kinase 9; CPA: cleavage and polyadenylation complex; CFIm: cleavage factor I.

Authors' contributions

MT, JGH, CJN and SM designed different aspects of the research. JGH produced the CFIm25KO cell line, MT and JGH carried out the ChIP-seq, MT performed the bioinformatics analysis. MT, CJN and SM drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank the Oxford Genomics Centre at the Wellcome Centre for Human Genetics (funded by Wellcome Trust Grant Reference 203141/Z/16/Z) for the generation of sequencing data.

Competing interests

The authors declare that they have no competing interests.

Availability of data materials

The data described in this Data note can be freely and openly accessed on the GEO website under the Accession Number: GSE119712 [19]. Please see Table 1 and reference list for details and links to the data.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Funding

This work was supported by a Wellcome Trust Senior Investigator Grant WT106134AIA to SM and a Cancer Research UK (CR-UK) Grant Number C38302/A13012, through an Oxford Cancer Research Centre Prize D.Phil. Studentship to JGH. The funders had no role in the design of the study and collection, analysis and interpretation of data and in writing the manuscript.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 12 September 2018 Accepted: 10 December 2018

Published online: 14 December 2018

References

- Derti A, Garrett-Engle P, Macisaac KD, Stevens RC, Sriram S, Chen R, et al. A quantitative atlas of polyadenylation in five mammals. *Genome Res.* 2012;22(6):1173–83.
- Shi Y, Manley JL. The end of the message: multiple protein–RNA interactions define the mRNA polyadenylation site. *Genes Dev.* 2015;29(9):889–97.
- Takagaki Y, Ryner LC, Manley JL. Four factors are required for 3'-end cleavage of pre-mRNAs. *Genes Dev.* 1989;3(11):1711–24.
- Rueggsegger 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):6107–13.
- Kim S, Yamamoto J, Chen Y, Aida M, Wada T, Handa H, et al. Evidence that cleavage factor Im is a heterotetrameric protein complex controlling alternative polyadenylation. *Genes Cells.* 2010;15(9):1003–13.
- Yang Q, Gilmartin GM, Doublet S. Structural basis of UGUA recognition by the Nudix protein CFI(m)25 and implications for a regulatory role in mRNA 3' processing. *Proc Natl Acad Sci USA.* 2010;107(22):10062–7.
- Kubo T, Wada T, Yamaguchi Y, Shimizu A, Handa H. 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):6264–71.
- Hardy JG, Tellier M, Murphy S, Norbury CJ. The RS domain of human CFI_m68 plays a key role in selection between alternative sites of pre-mRNA cleavage and polyadenylation. *bioRxiv.* 2017. <https://doi.org/10.1101/177980>.
- Zhu Y, Wang X, Forouzmand E, Jeong J, Qiao F, Sowd GA, et al. Molecular mechanisms for CFI_m-mediated regulation of mRNA alternative polyadenylation. *Mol Cell.* 2018;69(1):62–74.
- Masamha CP, Xia Z, Yang J, Albrecht TR, Li M, Shyu AB, et al. CFI_m25 links alternative polyadenylation to glioblastoma tumour suppression. *Nature.* 2014;510(7505):412–6.
- Gennarino VA, Alcott CE, Chen CA, Chaudhury A, Gillentine MA, Rosenfeld JA, et al. NUDT21-spanning CNVs lead to neuropsychiatric disease and altered MeCP2 abundance via alternative polyadenylation. *Elife.* 2015;4:e10782.
- Nojima T, Gomes T, Grosso ARF, Kimura H, Dye MJ, Dhir S, et al. Mammalian NET-Seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell.* 2015;161(3):526–40.
- Laitem C, Zaborowska J, Isa NF, Kufs J, Dienstbier M, Murphy S. CDK9 inhibitors define elongation checkpoints at both ends of RNA polymerase II-transcribed genes. *Nat Struct Mol Biol.* 2015;22(5):396–403.
- Laitem C, Zaborowska J, Tellier M, Yamaguchi Y, Cao Q, Egloff S, et al. CTCF regulates NELF, DSIF and P-TEFb recruitment during transcription. *Transcription.* 2015;6(5):79–90.
- Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 2011;17(1):3.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9(4):357–9.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–9.
- Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 2016;44(W1):W160–5.
- Tellier M, Hardy JG, Norbury CJ, Murphy S. Effect of CFI_m25 knock-out on RNA polymerase II transcription. *GEO GSE119712.* 2018. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE119712>.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

