THE EMBO JOURNAL

Transcription activation depends on the length of the RNA polymerase II C-terminal domain

Patrick Cramer, Anna Sawicka, Gabriel Villamil, Michael Lidschreiber, Xavier Darzacq, Claire Dugast-Darzacq, and Björn Schwalb
DOI: 10.15252/embj.2020107015

Corresponding author: Patrick Cramer (patrick.cramer@mpibpc.mpg.de)

Review Timeline:

Submission Date:	12th Oct 20
Editorial Decision:	13th Nov 20
Revision Received:	24th Nov 20
Editorial Decision:	17th Dec 20
Revision Received:	4th Jan 21
Accepted:	13th Jan 21

Editor: Stefanie Boehm

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. Depending on transfer agreements, referee reports obtained elsewhere may or may not be included in this compilation. Referee reports are anonymous unless the Referee chooses to sign their reports.)

13th Nov 2020

Re: EMBOJ-2020-107015
The length of the RNA polymerase II CTD controls transcription activation in human cells

Dear Patrick,

Thank you for submitting your manuscript EMBOJ-2020-107015 for consideration by The EMBO Journal. We have now received three referee reports on your study, which are included below for your information. Given the referees' comments, we would like to invite you to prepare and submit a revised manuscript.

As you will see, the reviewers are overall positive and appreciate the additional insights into the role of the RNA polymerase II CTD in transcription, but nonetheless raise some concerns that should be addressed in the revised version. In particular, the discussion of the effect of CTD shortening on enhancers versus promoters should be revised and the specific questions of referee #2 (point 1, 2) and referee #3 (point 2) clarified. In addition, the comments of referee #1 (point 1) and #3 (point 1) regarding the CTD truncations used should be discussed. Please also add references to the indicated studies where appropriate and carefully respond to all other referee comments.

Please note that it is our policy to allow only a single round of major revision. We realize that lab work worldwide is currently affected by the COVID-19/SARS-CoV-2 pandemic and that an experimental revision may be delayed. However, given the recently published related work, and the fact that the majority of the concerns can likely be addressed by textual edits, we would like to strongly encourage you to resubmit your manuscript as soon as possible. If you foresee any potential issues that may significantly delay a revision, please contact us to discuss this. Please also feel free to contact me should you have any other questions.

Thank you again for the opportunity to consider your work for publication. I look forward to receiving your revised manuscript.

Kind regards,

Stefanie

Stefanie Boehm
Editor
The EMBO Journal

Referee #1:

The study by Sawicka et al. analyses the impact of the length of the C-terminal domain (CTD) of RNA polymerase II on transcription. The RNA pol II CTD is a unique structure of 52 hepta-repeats of the consensus sequence YSPTSPS that is required for transcriptional regulation and the recruitment of pre-mRNA processing factors, particularly for splicing. While factor binding to the CTD relies on the interaction with short regions comprising only 2 to 3 hepta-repeats, the length of the CTD is though to correlated with the complexity of the organism, ranging from 26 repeat in S.cerevisiae to 52 repeat in human. The authors perform a systematic multi-omics approach of two different CTD forms transfected in different cell types to investigate the effect of the CTD length on the transcriptome, pre-mRNA splicing kinetics, pol II pausing, functional consequences of enhancer regions, and transcription induction.

Besides wild type, full length human Rpb1, the authors use a cell line that expresses a truncated form of the CTD from 52 to 25 repeats, which has been introduced before (Boehning et al., 2018). They find that CTD shortening does not in general change RNA synthesis and processing, but that the kinetics of RNA synthesis in response to an external signal is delayed and compromised for the short CTD. Using U2OS osteosarcoma cells they find that the transcriptome of mRNA synthesis and degradation rates as well as mRNA half-lives were essentially unchanged in naïve cells compared to CTD shortened cells. Using TT-seq datasets of two biological replicates the authors identify that CTD shortening only very slightly increases the splicing ratio. A comprehensive list of additional experiments showed that CTD shortening did not affect alternative splicing of pre-mRNA and only slightly alters promoter-proximal pausing of RNA pol II as the pol II occupancy downstream of the transcription start site slightly decreases. Likewise, pol II velocity and termination were not changed upon CTD shorting. The only significant changes in transcription using the short CTD was seen in a reduced number of transcript putative enhancers by TT-seq as well as a strong reduction of signal-activated transcription activation and enhancer transcription. Analysing these transcriptomics data led to the identification of 63 genes whose transcription activation was significantly delayed in the CTD-shortened cells after 15 min, yet, recovered transcription after 30 min, indicating a delay but not a full stop in transcription activation.

Overall, the findings presented largely agree with a recent study from the Roeder lab (2020), showing-by the use of an inducible pol II degradation system-that the CTD is not essential for the post-initiation control of pol II activity. In the study from the Cramer lab the results reveal that the normal CTD length is required for efficient transcription activation and that the CTD is critical for rapid pol II recruitment to genes upon their activation.

Comments:
It has been shown that the very C-terminal 10 residues of the human CTD contribute to its stability and functional integrity (Chapman et al., 2004). Regarding the design of this study where the impact of the CTD length on transcription is analysed, it would have been desirable, if the human full length CTD of 52 repeats would have been compared to a shortened CTD of either the proximal 26 repeats plus or minus the C-term 10 residues, and the distal 26 repeats, that contain multiple sequence alterations, as e.g. the lysine residue at position 7.

The second paragraph of the Discussion section is not well prepared. While the first half describes how CTD shortening delays enhancer-dependent transcription, stating twice: "In particular, our results ..." ending with the phrase: "This is supported by the literature." without citing any literature,

the second half describes effects of the CTD in liquid-liquid phase separation. Both aspects of CTD function are not connected in this paragraph. The following paragraph again enlightens both of these aspects trying to combine the genome size with its possibly increased distance of enhancers to the TSS to the benefit of phase separation, which has been shown before to correlate with CTD length. These thoughts could have been explained much clearer.

The authors should list the genes whose transcription activation is delayed upon CTD shortening (Fig. 6E) in the supplementary material. Are there any housekeeping genes involved, what about HSP coding genes?

It is surprising that the work of Schwer and Shuman on the minimal length of the CTD is yeast and cell viability is not acknowledged in the manuscript (e.g., Schwer B, Sanchez AM, Shuman S. Proc Natl Acad Sci U S A. (2012), but there are many others).

Typos:
Fig. 2F: alternative 5' splicing

Referee #2:

This work investigates the role of the Pol II CTD in human transcription, pausing, RNA processing and gene induction using matched cell lines with either the normal 52 or truncated 25 CTD repeats. Using carefully done genome-wide analyses, the authors convincingly show very little change in transcription of most genes. Also showing little change were RNA splicing and termination with a small change observed in promoter proximal pause duration. Importantly, CTD length-dependent changes were observed in transcription of eRNAs and in the kinetics of gene induction upon TPA stimulation. Overall, the manuscript makes an important contribution toward understanding the role of the CTD and transcriptional regulation. Publication is recommended after addressing the following:

1. I'm not an expert on the topic of matching putative enhancers with regulated genes, but it seems like there is some disagreement in the literature as to the best approach for this. How do the authors enhancer-promoter assignments compare with other results in the literature based on 5C-type proximity measurements or other methods used for U2OS cells?

2. The discussion has a speculative section on phase separation and Pol II enhancer recruitment to enhancers as a model to explain their findings. There aren't direct results to support this model in the manuscript, but I think it's OK as it is clearly speculation. However, one aspect of this model is the proposal that the CTD is important for Pol II recruitment to enhancers - I presume they are implying that the CTD is more important at enhancers than promoters - is this what the authors mean? Is there any evidence to support an important role for the CTD and enhancer recruitment? If so please add. Overall, this section of the discussion should be clarified.

3. Figs 3F, H: What are the units on the Y-axis of these graphs - it's not clear in its present form. Can they be put into something more readily grasped such as pause duration (seconds) and elongation velocity (bp/min)?

4. Appendix Fig 4B. It's not clear what is being plotted in these graphs and the text and legend are

not adequate. Please expand the figure legend and perhaps clarify the text in the figure.

Referee #3:

In this manuscript Sawicka et al. provide a comprehensive analysis of both the transcriptome and the rate of transcription in two human cell lines with different length CTDs. This is a modern update of experiments done decades ago with far less precise and informative approaches. The experiments are very well designed and presented in a clear manner. Although the authors report only minor differences in transcription and processing in the strains with different length CTDs there are some interesting differences in enhancer driven expression and in the response to MAP kinase signaling. Given the renewed interest in CTD function involving phase separated domains this paper will be of general interest to the transcription field.

My main concern is with the interpretation of the data. The authors contend that the differences they observe are due to the length of the CTD but fail to consider the possibility that that lack of the non-consensus repeats that are deleted in the shorter CTD strain lead to some or all of these differences. The non-consensus repeats are highly conserved in mammals and have been shown to have different abilities (compared to consensus repeats) to phase separate or form hydrogels with transcription factors. This possibility should be mentioned in the discussion.

One of the most striking findings is that the short CTD strain seems to have fewer active enhancers. This is a striking result and coupled with the time delay in response to MAP kinase signaling argues for an important role for the CTD repeats lost in the truncated version. One interesting point not addressed by the authors is the distance between the enhancer and promoter in the different CTD strains. Does the shorter CTD preferentially lose contact with the more distal enhancers? This data should be accessible to the authors.

Minor concern: In the introduction the authors cite the Allison et al. 1988 reference for deletions of the CTD in different organisms. The correct reference for the mammalian deletion is Bartolomei et al 1988.

To Dr. Stefanie Boehm
The Editor of the EMBO Journal

24 November 2020

Dear Dr. Boehm,

We send here a revised version of our manuscript to be considered as a research article for The EMBO Journal: The length of the RNA polymerase II CTD controls transcription activation in human cells (EMBOJ-2020-107015).

We would like to thank all the reviewers for their valuable comments, which helped us to improve the manuscript. Below please find a detailed point-by-point response to all comments. The changes that we introduced to the manuscript according to reviewers' suggestions and comments are depicted in red in the manuscript file. Briefly, we have re-written the discussion section and included the appropriate references. We have also modified the figure legend for Fig EV3B (formerly Appendix Fig S4B) and added an additional table (Table EV2) with genes showing delayed induction kinetics in response to TPA stimulation upon CTD shortening.

The sequencing data available at the GEO database (GSE159092) under the reviewer token (inerwsgkxdgjfeb).

Thank you very much for your consideration.

With kind regards

Patrick Cramer

Referee #1:

The study by Sawicka et al. analyses the impact of the length of the C-terminal domain (CTD) of RNA polymerase II on transcription. The RNA pol II CTD is a unique structure of 52 hepta-repeats of the consensus sequence YSPTSPS that is required for transcriptional regulation and the recruitment of pre-mRNA processing factors, particularly for splicing. While factor binding to the CTD relies on the interaction with short regions comprising only 2 to 3 hepta-repeats, the length of the CTD is though to correlated with the complexity of the organism, ranging from 26 repeat in S.cerevisiae to 52 repeat in human. The authors perform a systematic multi-omics approach of two different CTD forms transfected in different cell types to investigate the effect of the CTD length on the transcriptome, pre-mRNA splicing kinetics, pol II pausing, functional consequences of enhancer regions, and transcription induction.

Besides wild type, full length human Rpb1, the authors use a cell line that expresses a truncated form of the CTD from 52 to 25 repeats, which has been introduced before (Boehning et al., 2018). They find that CTD shortening does not in general change RNA synthesis and processing, but that the kinetics of RNA synthesis in response to an external signal is delayed and compromised for the short CTD. Using U2OS osteosarcoma cells they find that the transcriptome of mRNA synthesis and degradation rates as well as mRNA half-lives were essentially unchanged in naïve cells compared to CTD shortened cells. Using TT-seq datasets of two biological replicates the authors identify that CTD shortening only very slightly increases the splicing ratio. A comprehensive list of additional experiments showed that CTD shortening did not affect alternative splicing of pre-mRNA and only slightly alters promoter-proximal pausing of RNA pol II as the pol II occupancy downstream of the transcription start site slightly decreases. Likewise, pol II velocity and termination were not changed upon CTD shorting. The only significant changes in transcription using the short CTD was seen in a reduced number of transcript putative enhancers by TT-seq as well as a strong reduction of signal-activated transcription activation and enhancer

transcription. Analysing these transcriptomics data led to the identification of 63 genes whose transcription activation was significantly delayed in the CTD-shortened cells after 15 min, yet, recovered transcription after 30 min, indicating a delay but not a full stop in transcription activation.

Overall, the findings presented largely agree with a recent study from the Roeder lab (2020), showing-by the use of an inducible pol II degradation system-that the CTD is not essential for the post-initiation control of pol II activity. In the study from the Cramer lab the results reveal that the normal CTD length is required for efficient transcription activation and that the CTD is critical for rapid pol II recruitment to genes upon their activation.

We thank the reviewer for the insightful analysis and comments, which helped us to improve the manuscript.

Comments:

1. It has been shown that the very C-terminal 10 residues of the human CTD contribute to its stability and functional integrity (Chapman et al., 2004). Regarding the design of this study where the impact of the CTD length on transcription is analysed, it would have been desirable, if the human full length CTD of 52 repeats would have been compared to a shortened CTD of either the proximal 26 repeats plus or minus the C-term 10 residues, and the distal 26 repeats, that contain multiple sequence alterations, as e.g. the lysine residue at position 7.

We thank the reviewer for pointing out the issue with CTD stability. Our manuscript is a follow-up study in which we used the same CTD constructs as in (Boehning et al, 2018) in order to determine if reduced Pol II cluster size observed in living cells expressing RPB1-25R CTD correlates with changes in RNA synthesis. The RPB1-25R CTD mutant was designed based on findings in (Chapman et al, 2004, 2005), as the reviewer pointed out. Our RPB1-25R CTD contains 10 C-terminal residues (which is marked as "CTD52" in Figure 1A) as well as N-terminal repeats 1-3 that are necessary for CTD stability. As the reviewer indicated, the proximal CTD repeats are canonical, whereas the distal repeats diverge from it.

We agree that it is desirable to prepare and investigate additional cell lines with other forms of CTD variants, but this is beyond the scope of our present study. However, given that noncanonical repeats are not present in our RPB1-25R CTD mutant, we cannot exclude the possibility that the lack of noncanonical repeats can contribute to the phenotype we observe. We have included a section in our discussion where we comment on this.

2. The second paragraph of the Discussion section is not well prepared. While the first half describes how CTD shortening delays enhancer-dependent transcription, stating twice: "In particular, our results ..." ending with the phrase: "This is supported by the literature." without citing any literature, the second half describes effects of the CTD in liquid-liquid phase separation. Both aspects of CTD function are not connected in this paragraph. The following paragraph again enlightens both of these aspects trying to combine the genome size with its possibly increased distance of enhancers to the TSS to the benefit of phase separation, which has been shown before to correlate with CTD length. These thoughts could have been explained much clearer.

We have rewritten our discussion accordingly and included the appropriate references. We trust the reviewer now finds it clearer and more complete.

3. The authors should list the genes whose transcription activation is delayed upon CTD shortening (Fig. 6E) in the supplementary material. Are there any housekeeping genes involved, what about HSP coding genes?

We have now listed the genes showing delayed activation upon CTD shortening in Appendix Table S2. Among these genes there are many transcription factors (e.g. MYC, FOSB, FOSL1, IER3, etc.). Interestingly, these genes encode transcription and regulatory factors

that control the expression of downstream, late-response target genes and thereby regulate the timing of the response (Herschman, 1991; Hargreaves et al, 2009). We hypothesize that delay in their activation contributes to the general phenotype we see upon gene activation in cells expression shortened CTD. We thank the reviewer for pointing this out.

4. It is surprising that the work of Schwer and Shuman on the minimal length of the CTD is yeast and cell viability is not acknowledged in the manuscript (e.g., Schwer B, Sanchez AM, Shuman S. Proc Natl Acad Sci U S A. (2012), but there are many others).

We apologize for the oversight and have added the reference to the manuscript.

5. Typos:
Fig. 2F: alternative 5' splicing

We have corrected the typo.

Referee #2:

This work investigates the role of the Pol II CTD in human transcription, pausing, RNA processing and gene induction using matched cell lines with either the normal 52 or truncated 25 CTD repeats. Using carefully done genome-wide analyses, the authors convincingly show very little change in transcription of most genes. Also showing little change were RNA splicing and termination with a small change observed in promoter proximal pause duration. Importantly, CTD length-dependent changes were observed in transcription of eRNAs and in the kinetics of gene induction upon TPA stimulation. Overall, the manuscript makes an important contribution toward understanding the role of the CTD and transcriptional regulation. Publication is recommended after addressing the following:

We would like to thank the reviewer for the support and useful comments that helped us to improve the manuscript.

1. I'm not an expert on the topic of matching putative enhancers with regulated genes, but it seems like there is some disagreement in the literature as to the best approach for this. How do the authors enhancer-promoter assignments compare with other results in the literature based on 5C-type proximity measurements or other methods used for U2OS cells?

Pairing putative enhancers with their promoters indeed remains challenging and can be done using several approaches (Hariprakash & Ferrari, 2019), and applicability depends on the data available in the cell line and condition used. There are no chromatin conformation capture-like based datasets available for U2OS cells, and such data are difficult to produce at high quality, so this is unfortunately beyond the scope of our work. Nevertheless, our used paring approach is generally accepted and robust, and certainly allows us to reach the presented conclusions. Briefly, we paired active gene promoters by searching for the gene TSS on either strand that is nearest to the eRNA TSS within a maximum distance of ±500 kb. This window roughly corresponds to a size of a TAD, which is the chromatin domain within which regulation generally occurs (Dixon et al, 2012). However, considering all possible enhancer-promoter pairs within +- 500kb would have resulted in a very high number of false-positive pairs. Thus, we decided to use an approach with higher precision at the cost of sensitivity (see Fig. 3a in (Fulco et al, 2019) for a partial benchmark comparing different E-P pairing approaches). Fulco et al, 2019 also shows in Fig. 3a that if no chromatin interaction and DHS (DNAse I hypersensitivity) data is available then pairing to the nearest active gene is the best approach considering a good balance between precision and sensitivity (other methods are either very sensitive, but include many false-positive predictions (e.g. pairing with each gene in a +- 500kb window) or are more precise, but miss many true-positive pairs (e.g. correlation based methods)).

2. The discussion has a speculative section on phase separation and Pol II enhancer

recruitment to enhancers as a model to explain their findings. There aren't direct results to support this model in the manuscript, but I think it's OK as it is clearly speculation. However, one aspect of this model is the proposal that the CTD is important for Pol II recruitment to enhancers - I presume they are implying that the CTD is more important at enhancers than promoters - is this what the authors mean? Is there any evidence to support an important role for the CTD and enhancer recruitment? If so please add. Overall, this section of the discussion should be clarified.

We did not wish to suggest that the CTD is more important at enhancers. To avoid such misunderstandings, we have clarified the text and carefully rewritten this part of the discussion. We trust the reviewer finds the new text clearer.

3. Figs 3F, H: What are the units on the Y-axis of these graphs - it's not clear in its present form. Can they be put into something more readily grasped such as pause duration (seconds) and elongation velocity (bp/min)?

Our experimental setup does not permit us to report elongation velocity and pausing duration on an absolute scale. As described in Methods, we estimate elongation velocity by dividing the TT-seq signal by the PRO-seq signal. This allows for comparisons between different transcribed units as well as across cell lines.

4. Appendix Fig 4B. It's not clear what is being plotted in these graphs and the text and legend are not adequate. Please expand the figure legend and perhaps clarify the text in the figure.

We expanded the figure legends and clarified the text. This figure summarizes the statistics of our enhancer-promoter pairing strategy. The left bar plot shows a fraction of annotated putative enhancers that could be paired to promoters using our strategy. 2490 putative enhancers in RPB1-52R cells (out of all 2954 enhancers we annotated in RPB1-52R cells) could be paired with promoters using our strategy. This shows that 84.29% of all the putative enhancers we annotated in RPB1-52R could be paired to promoters. In RPB1-25R cells, 83.83% of putative enhancers annotated could be paired to promoters. The right panel shows a fraction of active promoters in U2OS cells that could be paired to putative enhancers using our strategy. There are 17,319 active promoters in U2OS cells. In RPB1-52R cells, 1,755 of active promoters could be paired with putative enhancers (which constitute 10.13% of all active promoters in RPB1-52R cells). In the case of RPB1-25R cells, 1,241 of all active promoters could be paired with putative enhancers (which constitutes 7.17% of all active promoters).

Referee #3:

In this manuscript Sawicka et al. provide a comprehensive analysis of both the transcriptome and the rate of transcription in two human cell lines with different length CTDs. This is a modern update of experiments done decades ago with far less precise and informative approaches. The experiments are very well designed and presented in a clear manner. Although the authors report only minor differences in transcription and processing in the strains with different length CTDs there are some interesting differences in enhancer driven expression and in the response to MAP kinase signaling. Given the renewed interest in CTD function involving phase separated domains this paper will be of general interest to the transcription field.

We thank the reviewer for the helpful consideration and comments.

1. My main concern is with the interpretation of the data. The authors contend that the differences they observe are due to the length of the CTD but fail to consider the possibility that that lack of the non-consensus repeats that are deleted in the shorter CTD strain lead to some or all of these differences. The non-consensus repeats are highly conserved in mammals and have been shown to have different abilities

(compared to consensus repeats) to phase separate or form hydrogels with transcription factors. This possibility should be mentioned in the discussion.

This is an important point and we thank the reviewer for bringing this up. The distal part of the CTD indeed contains CTD repeats that diverge from the consensus sequence and they are absent in our RPB1-25R mutant. In particular, based on literature, we cannot exclude the possibility that the absence of non-consensus CTD repeats in the RPB1-25R variant contributes to the phenotype we observe. We have added a brief paragraph on this to the discussion.

2. One of the most striking findings is that the short CTD strain seems to have fewer active enhancers. This is a striking result and coupled with the time delay in response to MAP kinase signaling argues for an important role for the CTD repeats lost in the truncated version. One interesting point not addressed by the authors is the distance between the enhancer and promoter in the different CTD strains. Does the shorter CTD preferentially lose contact with the more distal enhancers? This data should be accessible to the authors.

This is a very interesting question. As described in the Methods section, we paired gene promoters by searching for the active gene TSS on either strand that is nearest to the eRNA TSS within a maximum distance of ±500 kb. Since we imposed distance restrictions in our pairing procedure, drawing conclusions about the distance between putative enhancers and their target promoters would suffer from a bias. If we however investigate this taking these limitations into consideration, there is no difference in the distance between promoters and their putative enhancers in RPB1-52R and RPB1-25R CTD cells (see figure below). We therefore did not change the text.

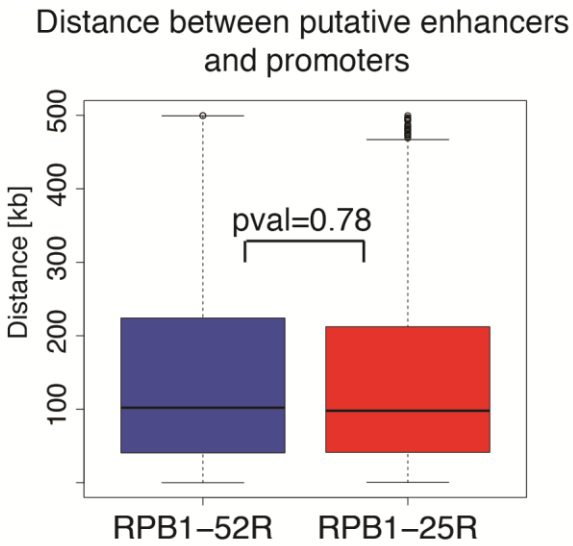

Figure: Bar plot showing distance between promoters and putative paired enhancers in RPB1-52R and RPB1-25R cells. p value = 0.78 (Mann-Whitney U test)

3. Minor concern: In the introduction the authors cite the Allison et al. 1988 reference for deletions of the CTD in different organisms. The correct reference for the mammalian deletion is Bartolomei et al 1988.

We added the reference to the text.

1st Revision - Editorial Decision 17th Dec 2020

17th Dec 2020

Re: EMBOJ-2020-107015R
The length of the RNA polymerase II CTD controls transcription activation in human cells

Dear Patrick,

Thank you for submitting your revised manuscript, we have now received the reports from the initial referees (see comments below). I am pleased to say that the referees find that their comments have been satisfactorily addressed and now support publication. I would like to therefore ask you to now address a number of editorial issues that are listed in detail below. Please make any changes to the manuscript text in the attached document only using the "track changes" option. Once these remaining issues are resolved, we will be happy to formally accept the manuscript for publication.

Thank you again for giving us the chance to consider your manuscript for The EMBO Journal. I look forward to receiving your final revision. Please feel free to contact me if you have further questions regarding the revision or any of the specific points listed below.

Kind regards,

Stefanie

Stefanie Boehm
Editor
The EMBO Journal

Referee #1:

In the revised version of the manuscript, the authors addressed all points raised from the first manuscript review. Particularly the Discussion section now reads much better and is more convincing with regard to the model of enhancer- and CTD length-dependent RNA pol II recruitment.

Referee #2:

The authors have done a good job of addressing concerns and revising the manuscript based on reviewer comments. I recommend publication.

Referee #3:

The authors have satisfied previous concerns.

2nd Authors' Response to Reviewers 4th Jan 2021

The authors performed the requested editorial changes.

2nd Revision - Final Editorial Decision 13th Jan 2021

Thank you again for submitting the final revised version of your manuscript for our consideration. I am pleased to inform you that we have now accepted it for publication in The EMBO Journal.

EMBO PRESS

YOU MUST COMPLETE ALL CELLS WITH A PINK BACKGROUND ↓

PLEASE NOTE THAT THIS CHECKLIST WILL BE PUBLISHED ALONGSIDE YOUR PAPER

Corresponding Author Name: Patrick Cramer
Journal Submitted to: The EMBO Journal
Manuscript Number: EMBOJ-2020-107015

Reporting Checklist For Life Sciences Articles (Rev. June 2017)

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. These guidelines are consistent with the Principles and Guidelines for Reporting Preclinical Research issued by the NIH in 2014. Please follow the journal's authorship guidelines in preparing your manuscript.

A- Figures

1. Data

The data shown in figures should satisfy the following conditions:

- ➔ the data were obtained and processed according to the field's best practice and are presented to reflect the results of the experiments in an accurate and unbiased manner.
- ➔ figure panels include only data points, measurements or observations that can be compared to each other in a scientifically meaningful way.
- ➔ graphs include clearly labeled error bars for independent experiments and sample sizes. Unless justified, error bars should not be shown for technical replicates.
- ➔ if n< 5, the individual data points from each experiment should be plotted and any statistical test employed should be justified
- ➔ Source Data should be included to report the data underlying graphs. Please follow the guidelines set out in the author ship guidelines on Data Presentation.

2. Captions

Each figure caption should contain the following information, for each panel where they are relevant:

- ➔ a specification of the experimental system investigated (eg cell line, species name).
- ➔ the assay(s) and method(s) used to carry out the reported observations and measurements
- ➔ an explicit mention of the biological and chemical entity(ies) that are being measured.
- ➔ an explicit mention of the biological and chemical entity(ies) that are altered/varied/perturbed in a controlled manner.

- ➔ the exact sample size (n) for each experimental group/condition, given as a number, not a range;
- ➔ a description of the sample collection allowing the reader to understand whether the samples represent technical or biological replicates (including how many animals, litters, cultures, etc.).
- ➔ a statement of how many times the experiment shown was independently replicated in the laboratory.
- ➔ definitions of statistical methods and measures:
  - common tests, such as t-test (please specify whether paired vs. unpaired), simple $\chi 2$ tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
  - are tests one-sided or two-sided?
  - are there adjustments for multiple comparisons?
  - exact statistical test results, e.g., P values = x but not P values < x;
  - definition of 'center values' as median or average;
  - definition of error bars as s.d. or s.e.m.

Any descriptions too long for the figure legend should be included in the methods section and/or with the source data.

In the pink boxes below, please ensure that the answers to the following questions are reported in the manuscript itself. Every question should be answered. If the question is not relevant to your research, please write NA (non applicable). We encourage you to include a specific subsection in the methods section for statistics, reagents, animal models and human subjects.

B- Statistics and general methods

Please fill out these boxes ↓ (Do not worry if you cannot see all your text once you press return)

1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size?	does not apply
1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.	does not apply
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	does not apply
3. Were any steps taken to minimize the effects of subjective bias when allocating animals/samples to treatment (e.g. randomization procedure)? If yes, please describe.	does not apply
For animal studies, include a statement about randomization even if no randomization was used.	does not apply
4.a. Were any steps taken to minimize the effects of subjective bias during group allocation or/and when assessing results (e.g. blinding of the investigator)? If yes please describe.	does not apply
4.b. For animal studies, include a statement about blinding even if no blinding was done	does not apply
5. For every figure, are statistical tests justified as appropriate?	The following test were used: Mann-Whitney U test (Figs: 2A, 2B, 2C, 2D, 3E, 3F, 3J, 3K, 4B, 4C, 5B, 5D, 6B and 6C); Fisher's exact test (Fig 2E); chi-squere test (Fig EV3C) and two-way ANOVA followed by Tukey multiple hypothesis testing (Fig EV1E).
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	For most of the test, no assumptions were made concerning the data distribution as Mann-Whitney U, chi-squere and Fisher's exact tests were applied, which are non-parametric. In Fig EV1E, statistical significance was estimated using two-way ANOVA, followed by the Tukey multiple comparisons test. To determine if normal distributung can be assumed in this case, Kolmogorov-Smirnov test was used.
Is there an estimate of variation within each group of data?	does not apply

USEFUL LINKS FOR COMPLETING THIS FORM

http://www.antibodypedia.com
http://1degreebio.org
http://www.equator-network.org/reporting-guidelines/improving-bioscience-research-repor

http://grants.nih.gov/grants/olaw/olaw.htm
http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Useofanimals/index.htm
http://ClinicalTrials.gov
http://www.consort-statement.org
http://www.consort-statement.org/checklists/view/32-consort/66-title

http://www.equator-network.org/reporting-guidelines/reporting-recommendations-for-tumo

http://datadryad.org

http://figshare.com

http://www.ncbi.nlm.nih.gov/gap

http://www.ebi.ac.uk/ega

http://biomodels.net/

http://biomodels.net/miriam/
http://jjj.biochem.sun.ac.za
http://oba.od.nih.gov/biosecurity/biosecurity_documents.html
http://www.selectagents.gov/

Is the variance similar between the groups that are being statistically compared?	does not apply

C- Reagents

6. To show that antibodies were profiled for use in the system under study (assay and species), provide a citation, catalog number and/or clone number, supplementary information or reference to an antibody validation profile. e.g., Antibodypedia (see link list at top right), 1DegreeBio (see link list at top right).	We used the following antibodies: α-HALO (dilution 1:1000, Promega, #G9211) or α−U1 snRP (dilution 1:200, Santa Cruz, #sc-39089).
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.	We used previously reported human osteosarcoma U2OS cells (Boehning et al, 2018). The parental U2OS cell line was authenticated by the UC Berkeley cell culture facility on 05/05/2017 by STR analysis. Drospohila Schneider-2 cells were obtained directly from Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ, #ACC130). All cell lines were regularly tested for Mycoplasma contamination using PlasmoTest Mycoplasma Detection Kit (InvivoGen, #rep-pt1).

* for all hyperlinks, please see the table at the top right of the document

D- Animal Models

8. Report species, strain, gender, age of animals and genetic modification status where applicable. Please detail housing and husbandry conditions and the source of animals.	does not apply
9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.	does not apply
10. We recommend consulting the ARRIVE guidelines (see link list at top right) (PLoS Biol. 8(6), e1000412, 2010) to ensure that other relevant aspects of animal studies are adequately reported. See author guidelines, under 'Reporting Guidelines'. See also: NIH (see link list at top right) and MRC (see link list at top right) recommendations. Please confirm compliance.	does not apply

E- Human Subjects

11. Identify the committee(s) approving the study protocol.	does not apply
12. Include a statement confirming that informed consent was obtained from all subjects and that the experiments conformed to the principles set out in the WMA Declaration of Helsinki and the Department of Health and Human Services Belmont Report.	does not apply
13. For publication of patient photos, include a statement confirming that consent to publish was obtained.	does not apply
14. Report any restrictions on the availability (and/or on the use) of human data or samples.	does not apply
15. Report the clinical trial registration number (at ClinicalTrials.gov or equivalent), where applicable.	does not apply
16. For phase II and III randomized controlled trials, please refer to the CONSORT flow diagram (see link list at top right) and submit the CONSORT checklist (see link list at top right) with your submission. See author guidelines, under 'Reporting Guidelines'. Please confirm you have submitted this list.	does not apply
17. For tumor marker prognostic studies, we recommend that you follow the REMARK reporting guidelines (see link list at top right). See author guidelines, under 'Reporting Guidelines'. Please confirm you have followed these guidelines.	does not apply

F- Data Accessibility

18: Provide a "Data Availability" section at the end of the Materials & Methods, listing the accession codes for data generated in this study and deposited in a public database (e.g. RNA-Seq data: Gene Expression Omnibus GSE39462, Proteomics data: PRIDE PXD000208 etc.) Please refer to our author guidelines for 'Data Deposition'. Data deposition in a public repository is mandatory for: a. Protein, DNA and RNA sequences b. Macromolecular structures c. Crystallographic data for small molecules d. Functional genomics data e. Proteomics and molecular interactions	The sequencing data available at the GEO database (GSE159092).
19. Deposition is strongly recommended for any datasets that are central and integral to the study; please consider the journal's data policy. If no structured public repository exists for a given data type, we encourage the provision of datasets in the manuscript as a Supplementary Document (see author guidelines under 'Expanded View' or in unstructured repositories such as Dryad (see link list at top right) or Figshare (see link list at top right).	does not apply
20. Access to human clinical and genomic datasets should be provided with as few restrictions as possible while respecting ethical obligations to the patients and relevant medical and legal issues. If practically possible and compatible with the individual consent agreement used in the study, such data should be deposited in one of the major public access-controlled repositories such as dbGAP (see link list at top right) or EGA (see link list at top right).	does not apply
21. Computational models that are central and integral to a study should be shared without restrictions and provided in a machine-readable form. The relevant accession numbers or links should be provided. When possible, standardized format (SBML, CellML) should be used instead of scripts (e.g. MATLAB). Authors are strongly encouraged to follow the MIRIAM guidelines (see link list at top right) and deposit their model in a public database such as Biomodels (see link list at top right) or JWS Online (see link list at top right). If computer source code is provided with the paper, it should be deposited in a public repository or included in supplementary information.	does not apply

G- Dual use research of concern

22. Could your study fall under dual use research restrictions? Please check biosecurity documents (see link list at top right) and list of select agents and toxins (APHIS/CDC) (see link list at top right). According to our biosecurity guidelines, provide a statement only if it could.	does not apply