EMBO reports

RNA Polymerase II CTD is dispensable for transcription and required for termination in human cells

Yousra Yahia, Alexia Pigeot, Amal Zine El Aabidine, Nilay Shah, Nezih Karasu, Ignasi Forne, Stefan Krebs, Helmut Blum, Cyril Esnault, Tom Sexton, Axel Imhof, Dirk Eick, and Jean-Christophe Andrau
DOI: 10.15252/embr.202256150

Corresponding author(s): Jean-Christophe Andrau (jean-christophe.andrau@igmm.cnrs.fr) , Dirk Eick (eick@helmholtz-muenchen.de)

Review Timeline:

Submission Date:	19th Sep 22
Editorial Decision:	6th Oct 22
Revision Received:	9th May 23
Editorial Decision:	14th Jun 23
Revision Received:	21st Jun 23
Accepted:	27th Jun 23

Editor: Esther Schnapp

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.)

Dear Dr. Andrau,

Thank you for the submission of your manuscript to EMBO reports. We have now received the full set of referee reports that is pasted below.

As you will see, the referees acknowledge that the findings are potentially interesting. However, they also raise some concerns and have several suggestions for how the study could be improved. I think all suggestions are good and should be addressed, but please let me know in case you disagree and we can discuss the revisions in more detail, also in a video chat, if you like.

I would thus like to invite you to revise your manuscript with the understanding that the referee concerns must be fully addressed and their suggestions taken on board. Please address all referee concerns in a complete point-by-point response. Acceptance of the manuscript will depend on a positive outcome of a second round of review. It is EMBO reports policy to allow a single round of major revision only and acceptance or rejection of the manuscript will therefore depend on the completeness of your responses included in the next, final version of the manuscript.

We realize that it is difficult to revise to a specific deadline. In the interest of protecting the conceptual advance provided by the work, we recommend a revision within 3 months (6th Jan 2023). Please discuss the revision progress ahead of this time with the editor if you require more time to complete the revisions.

IMPORTANT NOTE: we perform an initial quality control of all revised manuscripts before re-review. Your manuscript will FAIL this control and the handling will be DELAYED if the following APPLIES:
1) A data availability section providing access to data deposited in public databases is missing. If you have not deposited any data, please add a sentence to the data availability section that explains that.
2) Your manuscript contains statistics and error bars based on n=2. Please use scatter blots in these cases. No statistics should be calculated if n=2.

When submitting your revised manuscript, please carefully review the instructions that follow below. Failure to include requested items will delay the evaluation of your revision.

1) a .docx formatted version of the manuscript text (including legends for main figures, EV figures and tables). Please make sure that the changes are highlighted to be clearly visible.

2) individual production quality figure files as .eps, .tif, .jpg (one file per figure). See https://wol-prod-cdn.literatumonline.com/pb-assets/embo-site/EMBOPress_Figure_Guidelines_061115-1561436025777.pdf for more info on how to prepare your figures.

3) We replaced Supplementary Information with Expanded View (EV) Figures and Tables that are collapsible/expandable online. A maximum of 5 EV Figures can be typeset. EV Figures should be cited as 'Figure EV1, Figure EV2" etc... in the text and their respective legends should be included in the main text after the legends of regular figures.

- For the figures that you do NOT wish to display as Expanded View figures, they should be bundled together with their legends in a single PDF file called *Appendix*, which should start with a short Table of Content. Appendix figures should be referred to in the main text as: "Appendix Figure S1, Appendix Figure S2" etc. See detailed instructions regarding expanded view here: <https://www.embopress.org/page/journal/14693178/authorguide#expandedview>

- Additional Tables/Datasets should be labeled and referred to as Table EV1, Dataset EV1, etc. Legends have to be provided in a separate tab in case of .xls files. Alternatively, the legend can be supplied as a separate text file (README) and zipped together with the Table/Dataset file.

4) a .docx formatted letter INCLUDING the reviewers' reports and your detailed point-by-point responses to their comments. As part of the EMBO Press transparent editorial process, the point-by-point response is part of the Review Process File (RPF), which will be published alongside your paper.

5) a complete author checklist, which you can download from our author guidelines <https://www.embopress.org/page/journal/14693178/authorguide>. Please insert information in the checklist that is also reflected in the manuscript. The completed author checklist will also be part of the RPF.

6) Please note that all corresponding authors are required to supply an ORCID ID for their name upon submission of a revised manuscript (<https://orcid.org/>). Please find instructions on how to link your ORCID ID to your account in our manuscript tracking system in our Author guidelines <https://www.embopress.org/page/journal/14693178/authorguide#authorshipguidelines>

7) Before submitting your revision, primary datasets produced in this study need to be deposited in an appropriate public

database (see https://www.embopress.org/page/journal/14693178/authorguide#datadeposition). Please remember to provide a reviewer password if the datasets are not yet public. The accession numbers and database should be listed in a formal "Data Availability" section placed after Materials & Method (see also https://www.embopress.org/page/journal/14693178/authorguide#datadeposition). Please note that the Data Availability Section is restricted to new primary data that are part of this study. * Note - All links should resolve to a page where the data can be accessed. *
If your study has not produced novel datasets, please mention this fact in the Data Availability Section.

8) At EMBO Press we ask authors to provide source data for the main and EV figures. Our source data coordinator will contact you to discuss which figure panels we would need source data for and will also provide you with helpful tips on how to upload and organize the files.

9) Our journal also encourages inclusion of *data citations in the reference list* to directly cite datasets that were re-used and obtained from public databases. Data citations in the article text are distinct from normal bibliographical citations and should directly link to the database records from which the data can be accessed. In the main text, data citations are formatted as follows: "Data ref: Smith et al, 2001" or "Data ref: NCBI Sequence Read Archive PRJNA342805, 2017". In the Reference list, data citations must be labeled with "[DATASET]". A data reference must provide the database name, accession number/identifiers and a resolvable link to the landing page from which the data can be accessed at the end of the reference. Further instructions are available at https://www.embopress.org/page/journal/14693178/authorguide#referencesformat

10) Regarding data quantification (see Figure Legends: https://www.embopress.org/page/journal/14693178/authorguide#figureformat)

The following points must be specified in each figure legend:

- the name of the statistical test used to generate error bars and P values,

- the number (n) of independent experiments (please specify technical or biological replicates) underlying each data point,

- the nature of the bars and error bars (s.d., s.e.m.),

- If the data are obtained from n {less than or equal to} 2, use scatter blots showing the individual data points.

Discussion of statistical methodology can be reported in the materials and methods section, but figure legends should contain a basic description of n, P and the test applied.

- Please also include scale bars in all microscopy images.

11) The journal requires a statement specifying whether or not authors have competing interests (defined as all potential or actual interests that could be perceived to influence the presentation or interpretation of an article). In case of competing interests, this must be specified in your disclosure statement. Further information: https://www.embopress.org/competing-interests

We would also welcome the submission of cover suggestions, or motifs to be used by our Graphics Illustrator in designing a cover.

As part of the EMBO publication's Transparent Editorial Process, EMBO reports publishes online a Review Process File (RPF) to accompany accepted manuscripts. This File will be published in conjunction with your paper and will include the referee reports, your point-by-point response and all pertinent correspondence relating to the manuscript.

You are able to opt out of this by letting the editorial office know (emboreports@embo.org). If you do opt out, the Review Process File link will point to the following statement: "No Review Process File is available with this article, as the authors have chosen not to make the review process public in this case."

I look forward to seeing a revised form of your manuscript when it is ready. Please use this link to submit your revision: https://embor.msubmit.net/cgi-bin/main.plex

Best regards,
Esther

Esther Schnapp, PhD
Senior Editor
EMBO reports

Referee #1:

Yahia et al. report a thorough analysis of an RNA polymerase II mutant harboring a large deletion in the Rpb1 CTD (CTD-Δ5). Using established approaches to assess the functionality of Pol II mutants, the effects of the CTD-Δ5 Pol II on nascent and poly(A) RNA, on Pol II occupancy, on epigenetic marks and on the Pol II interactome were determined 24 h after addition of alpha-amanitin to inhibit endogenous Pol II. The data are of high quality, they provide interesting new insights into CTD function, and they are consistent with the proposed model.

Clarifying/addressing the following points will, in my opinion, improve the manuscript.

1. Given that prior studies indicated greatly reduced transcription in the CTD-Δ5 mutant and that the data (wig files) are rescaled to normalize enrichment scores to the depth of sequencing, it is important to (at a minimum) comment on the "RNA levels" (e.g. total read numbers per sample in each of the replicates) in the text (and ideally in a table). Knowing the sequencing depth might help the reader appreciate why the authors cannot comment on splicing or capping. For example, it is clear that remaining cytoplasmic RNA complicates conclusions regarding poly(A) RNA seq, but this should, in my opinion, not be a confounding issue for nascent RNA-seq.
2. While the findings from Δ5 and YFFF regarding the mutants effect on read through transcription are strong (in large part because the experiments were done in parallel under the same conditions), other comparisons (e.g. that Δ5 shows a less marked phenotype than YFFF in poly(A) seq) should mention that the data from YFFF had been obtained 48 h after addition of alpha-amanitin to avoid any confusion.
3. Out of curiosity - any thoughts of whether CTD-Δ5 is unstable or seizes to be expressed 48 h after alpha-amanitin addition? [it had been readily detected 2 days after alpha-amanitin addition in a prior study with a seemingly comparable experimental setup (Meininghaus et al 2000)]
4. It would be nice if the authors expanded on the last sentence in the discussion (or excise it). I am not sure of what the missing standard refers to. Is it to acknowledge that interpretation of all mutational effects need to take into account that they could 'simply' be a consequence of a reduced number of intact repeats rather than be specific to certain mutations?
5. Experimental Procedures:
• Viability curve - what do the error bars indicate?
• PolyA RNA-seq - the modifications to the protocol are puzzling. The authors say that the fragmented (~150 nucleotides) RNA was purified using the RNeasy Mini Kit, which excludes short RNAs ({less than or equal to} ~ 200 nucleotides in length) - was the flowthrough used for library prep??
• ChrRNA-seq - what does 'rigorous treatment' with TurboDNase mean? i.e. how do the authors ensure that the treatment is complete ? (e.g. qPCR for tubulin prior to RT?)

Referee #2:

Yahia et al have investigated the transcriptional properties of Pol II without most of the CTD using alpha-amanitin versions of Pol II with wild-type or mutant CTDs. They have carried out chromatin RNA-seq followed by DESeq analysis. The delta 5 mutant appears to cause an increase of transcription across the genome, which appears to be largely due to readthrough and loss of Pol II at the ends of genes (although see below). In addition, the readthrough can cause interference of downstream genes. Interestingly, transcription of putative enhancers is less affected. They have also carried out Pol II ChIP-seq, which indicates that transcription by the delta 5 mutant is very low, and perhaps gives more readthrough, although I can't see this clearly on Figure 3. There is an effect of CTD deletion on polyadenylated RNA, but I did not really understand the authors' comments about this and they seem to have used total RNA rather than nuclear or chromatin-associated RNA.
Importantly, the authors have investigated the interactome of the delta 5. Interaction with many Mediator and Integrator subunits is lost. Quite unexpectedly, interactions with the spliceosome and CPA factors is maintained. Loss of most of the Pol II CTD also affects the level of chromatin-associated RNA from histone and snRNA genes (but see below). Finally, the authors have investigated whether readthrough transcription caused by loss of the CTD affects TAD borders or structure and conclude that it doesn't.
The authors conclude from their data that the delta 5 mutant can still carry out long-range transcription in the absence of most of the CTD but termination, and perhaps initiation and polyadenylation, is impaired.
There are some interesting findings here. For example, the interactome data is very interesting and novel. In addition, the results of the DESeq data clearly indicate a readthrough phenotype. However, I have some comments about some of the data and the conclusions drawn.
Comments
Chromatin RNA-seq is not a direct measure of transcription, as it represents a mix of transcription and transcript retention on the chromatin. If the delta 5 Pol II produces poorly-processed transcripts, this will affect transcript retention on chromatin and therefore the chromatin RNA-seq signal independent of any transcriptional changes. The ChIP-seq data clearly shows that the delta 5 mutant is severely impaired in transcription but the readthrough phenotype should be more clearly pointed out in Figure 3

with more examples. Given the low level of transcription by the Pol II without most of the CTD, I don't understand the conclusion that "The CTD appears also dispensable for the correct recruitment of Pol II at promoters and the process of transcription per se."
Also, from the Methods section, it appears that the chromatin RNA-seq and ChIP-seq reads have not been normalised to a spike-in control. As the delta 5 pol II has a global effect on transcription, spike-in normalization is necessary. In addition, it appears from the GEO submission that Input controls were not sequenced, which could affect the conclusions drawn from the ChIP-seq data.
A higher transcription elongation rate is also known to promote transcriptional readthrough. There is therefore the possibility that that the delta 5 mutant has a higher elongation rate than WT Pol II. From the proteomics data, are the elongation factors still binding to the delta 5 Pol II (DISF, STP6, SEC, PAF1c)?
Is the delta 5 Pol II known to be still phosphorylated on the five heptapeptides left?
In addition, some attempt at analysing de novo splicing and polyadenylation could have been made as a defect in splicing would lead to readthrough transcription. If I understood well, the authors used total RNA for their poly(A) analysis, whereas it would have been better to analyse nuclear or chromatin-associated RNA or even RNA from inducible genes. Nascent techniques (TT-seq, POINT-seq, or others) would also have been useful.
Also, why is there no single gene ChIP validation and why analyse the chromatin RNA-seq for snRNA genes and histone genes but not the Pol II ChIP-seq, which should be more informative about transcription? If the spliceosome is still associated with chromatin, surely the mature snRNA will be picked up in the chromatin RNA-seq?
The TAD analysis is well done and the conclusions are justified but this data doesn't add very much and the low level of transcription by the delta 5 might well not impact these.
In conclusion, the authors have presented some intriguing and interesting data, particularly the results of the interactome analysis but some more detailed analysis of what is happening to elongation and RNA processing is missing.
Minor points
Is the delta 5 the same as the Bentley delta 5?
The English needs editing by a native English speaker.

Referee #3:

In this manuscript Yahia et al. present the results describing the transcriptional properties of an RNA polymerase II mutant with a minimal length CTD. Transcription by the delta5 Pol II is examined using an established protocol in which an amanitin resistant large subunit gene is induced with doxycycline while the endogenous Pol II is inhibited with amanitin. This enables the examination of transcription by delta5 Pol II in a narrow time window before the cells die from the effects of the CTD mutation. One of the nice features here is the inclusion of a Tyr mutants that has previously been shown to alter Pol II termination. The main conclusion of the authors is that truncation of the CTD has a greater effect on termination and in fact increased transcription of some genes compared to WT. The other major conclusion is that delta5 Pol II transcribes more widely than previously observed. The delta5 interactome indicates reduced interaction with the mediator raising the question of how delta5 can initiate. These are interesting observations that lay the groundwork for further studies on the role of the CTD.

There are several minor concerns that I think could be addressed to improve the paper:

1. One problem with the interpretation of the results is that the RNAseq data are normalized to the depth of reads. Does this type of normalization over-emphasize the level of transcription in the delta5 strain? I would think that a spike-in control would provide a more accurate normalization.

2. Another question is whether the observed global changes in transcription are due to CTD truncation directly or due to changes in expression of a gene or genes that may alter transcription. While not addressable given the experimental approach, this should be mentioned in the text.

3. There seems to be more delta5 upstream of coding regions in the sense direction. Is this readthrough from upstream genes? This explanation would not be favored as the set of genes used in the analysis has had nearby genes removed. Could this indicate promiscuous initiation? Perhaps at nearby enhancers? If the delta5 polymerase can initiate at any open chromatin site then this could explain the "pervasive" nature of the delta5 occupancy. Perhaps long-read RNA sequence data could determine whether delta5 transcribes long RNAs or whether it is simply initiating promiscuously at multiple sites in open chromatin.

4. The capping machinery is known to bind the CTD raising to question of whether the transcripts synthesized by delta5 capped? If not, are these transcripts subjected to nuclear quality control mechanisms?

5. At the end of the Discussion there is a typographical error referring to a "3' cap".

Answer to reviewers:
We thank all three reviewers for their constructive comments that have helped us to improve the manuscript. We hope our answers and changes in the manuscript will fulfill their expectations.

Referee #1:

Yahia et al. report a thorough analysis of an RNA polymerase II mutant harboring a large deletion in the Rpb1 CTD (CTD-Δ5). Using established approaches to assess the functionality of Pol II mutants, the effects of the CTD-Δ5 Pol II on nascent and poly(A) RNA, on Pol II occupancy, on epigenetic marks and on the Pol II interactome were determined 24 h after addition of alpha-amanitin to inhibit endogenous Pol II. The data are of high quality, they provide interesting new insights into CTD function, and they are consistent with the proposed model.

Clarifying/addressing the following points will, in my opinion, improve the manuscript.

We thank the referee for his positive assessment and will try to clarify the points raised below.

1. Given that prior studies indicated greatly reduced transcription in the CTD-Δ5 mutant and that the data (wig files) are rescaled to normalize enrichment scores to the depth of sequencing, it is important to (at a minimum) comment on the "RNA levels" (e.g. total read numbers per sample in each of the replicates) in the text (and ideally in a table). Knowing the sequencing depth might help the reader appreciate why the authors cannot comment on splicing or capping. For example, it is clear that remaining cytoplasmic RNA complicates conclusions regarding poly(A) RNA seq, but this should, in my opinion, not be a confounding issue for nascent RNA-seq.

#R.1 To answer the referee's question, we have now included a table that relates the sequencing depth for each RNA-seq (and ChIP-seq) experiments performed in this article (Table EV4). However, the sequencing depth of ChrRNA-seq per se is not informative on neither (1) the absolute number of RNA produced in a cell, (2) splicing levels nor (3) capping as detailed below.

(1) In the case of the RNA levels, an ideal experiment would include spike-in controls optimized for cell number normalization, which remains a difficult task. The existing spike-in controls proposed for RNA library preparation are essentially the ERCC (synthetic RNA spike-in) which are added to constant levels of total RNA in a given experiment (Risso et al., 2014). However, those are not satisfactory since they do not reflect specifically the amounts of Pol II transcripts but rather the total amounts of transcripts produced (which quantitatively are a majority of Pol I transcripts). So given that ERCC are to the best technical controls, we chose not to use them for normalization (see also our answer to referee 2 #R2.4).

(2) concerning the splicing levels, they are difficult to assess because the major population in ChrRNA-seq experiment is unprocessed transcripts. We have however now included a new analysis of the fraction of exons and introns in ChrRNA in the rWT and CTD-Δ5 mutant (Fig. 2E) and describe it in the text (lines *170-173*): '*We also observed that intronic reads generally tend to increase in the mutant context as compared to exonic reads but this difference becomes reduced when considering genes with significant units of transcripts, suggesting a slight level of splicing defect in the CTD-Δ5 context*'. They indeed suggest a splicing defect for the Δ5 mutant but this defect becomes reduced when genes with a minimal level of expression (RPKM>1) are considered.

(3) the capping levels requires a more complex set-up that is not covered by Chr-RNA-seq (CAGE-seq or other nascent specific approach such as PRO-CAP).

2. While the findings from Δ5 and YFFF regarding the mutant effect on read through transcription are strong (in large part because the experiments were done in parallel under the same conditions), other comparisons (e.g. that Δ5 shows a less marked phenotype than YFFF in poly(A)

seq) should mention that the data from YFFF had been obtained 48 h after addition of alpha-amanitin to avoid any confusion.
#R1.2 We have included this precision in the revised manuscript (line 228-231):' *The fact that RT of CTD-Δ5 mutant PolyA RNA is less pronounced than that of the YFFF mutant could be explained by the time of collection as those from CTD-Δ5 are extracted after 24h of α-amanitin addition while those from YFFF after 48h.'*

3. Out of curiosity - any thoughts of whether CTD-Δ5 is unstable or seizes to be expressed 48 h after alpha-amanitin addition? [it had been readily detected 2 days after alpha-amanitin addition in a prior study with a seemingly comparable experimental setup (Meininghaus et al 2000)]
#R1.3 We do not know why this mutant seizes expression but this observation is reproducible in our hands. We speculate that there might slight changes in the procedure in 20 years' time that we did not identify.

4. It would be nice if the authors expanded on the last sentence in the discussion (or excise it). I am not sure of what the missing standard refers to. Is it to acknowledge that interpretation of all mutational effects needs to take into account that they could 'simply' be a consequence of a reduced number of intact repeats rather than be specific to certain mutations?
#R1.4 What we meant here was that the knowledge of the CTD impact on transcription in mammalian cells requires a reference in which the CTD is essentially absent to monitor globally the extent of its contribution. We have now implemented this sentence (*line 382-385*): "*With this paper, we provide a missing standard for all past and future CTD manipulation studies, i.e that a complete defect of the CTD results primarily, from a transcriptional point, in a massive pervasive phenotype. Thus, such effect observed for other mutations such as for the YFFF relate to one of the major functions of the domain.*"

5. Experimental Procedures:
• Viability curve - what do the error bars indicate?
#R1.5.1 The error bars indicate the variations over 3 independent experiments. This was now included in the revised manuscript (see Fig. 1B legend).

• PolyA RNA-seq - the modifications to the protocol are puzzling. The authors say that the fragmented (~150 nucleotides) RNA was purified using the RNeasy Mini Kit, which excludes short RNAs (~ 200 nucleotides in length) - was the flowthrough used for library prep??
#R1.5.2 We agree with the referee that in principle this kit is designed to exclude short RNAs. In practice, a slight modification to the protocol we previously introduced in (Shah et al., 2018) allows to get a good recovery of fragments between 100 and 200 nt. This was now added in the method section (*lines 498-502*): '*Fragmentation reaction was stopped by adding 90µl nuclease-free water and quickly adding 350µl RLT buffer from RNeasy Mini Kit (QIAGEN, Germany) followed by purification of fragmented RNA using RNA Cleanup Protocol from this kit. However, to enhance the recovery of smaller fragments, we added 500µl ethanol instead of recommended 250µl.'*

• ChrRNA-seq - what does 'rigorous treatment' with TurboDNase mean?

#R1.5.3 The rigorous treatment is the one described in the TURBO DNA-free™procedure guidelines. We now summarized this in the method as follows *(lines 484-488)*: '*after incubation with Turbo DNase for 37°C for 30 minutes samples are not directly subjected to the Stop reagent but instead re-extracted with Trizol, precipitated and after resuspension digested again with Turbo DNAse, stopped with Stop reagent, extracted and finally resuspended in 20ul water.'*

i.e. how do the authors ensure that the treatment is complete? (e.g. qPCR for tubulin prior to RT?)
#R1.5.4 We assessed this by performing qPCR on GAPDH genomic DNA sequences in the presence or absence of reverse transcriptase. These controls showed no DNA amplification in the absence of RT reaction (while amplification was found in its presence). This was now added in the method section (*lines 517-520*). '*To confirm the absence of remaining contaminant DNA, a*

qPCR on a coding region of the GAPDH gene was performed prior and after reverse transcription for each sample. (Forward primer: 5'-ATTTGGTCGTATTGGGCGC-3' and reverse primer: 5'-TGAAGGGGTCATTGATGGC-3').

Referee #2:

Yahia et al have investigated the transcriptional properties of Pol II without most of the CTD using alpha-amanitin versions of Pol II with wild-type or mutant CTDs. They have carried out chromatin RNA-seq followed by DESeq analysis. The delta 5 mutant appears to cause an increase of transcription across the genome, which appears to be largely due to readthrough and loss of Pol II at the ends of genes (although see below). In addition, the readthrough can cause interference of downstream genes. Interestingly, transcription of putative enhancers is less affected. They have also carried out Pol II ChIP-seq, which indicates that transcription by the delta 5 mutant is very low, and perhaps gives more readthrough, although I can't see this clearly on Figure 3.
There is an effect of CTD deletion on polyadenylated RNA, but I did not really understand the authors' comments about this and they seem to have used total RNA rather than nuclear or chromatin-associated RNA.
Importantly, the authors have investigated the interactome of the delta 5. Interaction with many Mediator and Integrator subunits is lost. Quite unexpectedly, interactions with the spliceosome and CPA factors is maintained. Loss of most of the Pol II CTD also affects the level of chromatin-associated RNA from histone and snRNA genes (but see below). Finally, the authors have investigated whether readthrough transcription caused by loss of the CTD affects TAD borders or structure and conclude that it doesn't.
The authors conclude from their data that the delta 5 mutant can still carry out long-range transcription in the absence of most of the CTD but termination, and perhaps initiation and polyadenylation, is impaired.
There are some interesting findings here. For example, the interactome data is very interesting and novel. In addition, the results of the DESeq data clearly indicate a readthrough phenotype. However, I have some comments about some of the data and the conclusions drawn.

We thank the referee for his/her in depth reviewing of our manuscript. We have tried to best address his/her comments below.

Comments
Chromatin RNA-seq is not a direct measure of transcription, as it represents a mix of transcription and transcript retention on the chromatin. If the delta 5 Pol II produces poorly-processed transcripts, this will affect transcript retention on chromatin and therefore the chromatin RNA-seq signal independent of any transcriptional changes.
#R2.1 We agree that ChrRNA-seq is not per se a direct measurement of nascent transcription. However, based on our past experience of analysis of published data (Nojima et al., 2015), for example of NET-seq, this procedure yields a read-out extremely close to that of the nascent procedure, to the exception of short 5' transcripts, which are not optimally covered by ChrRNA-seq. We also agree that the CTD mutant might well have a retention defect because of processing issue, but this could not explain by itself a phenotype of global readthrough over hundreds of kb. If this would be the case, such readthrough would already exist in the context of our rWT control cells or any human cells, which to the best of our knowledge has never been described, for example using read-out such as GRO-seq, PRO-seq, NET-seq or TT-seq. Thus, we assume that while a certain level of read-through does exit in natural cellular context a few kb after the termination signals, the readthrough described for the CTD-Δ5 mutant (up to several hundreds of kb) could not be explained by an increased amount of poorly-processed transcripts. We also note that previously published work performed in transcription assay in vitro with Pol II devoid of CTD indicate a read-through that is consistent with our observations (Tsao et al., 2012). Nevertheless, we have now included this possibility in the article (*lines 167-170*): '*Although we cannot completely rule out that the chrRNA procedure could lead to transcripts retention, we believe this hypothesis unlikely since it would imply that WT cells produce such transcripts to a massive level, which has never been described*.' Further on, as also mentioned in the article, no

interaction defect was observed between splicing/processing factors, detected in the rWT, and the CTD-Δ5 mutant.

The ChIP-seq data clearly shows that the delta 5 mutant is severely impaired in transcription but the readthrough phenotype should be more clearly pointed out in Figure 3 with more examples.
#R2.2 More ChIP-seq examples are now included in Fig. EV3B. The region shown in the grey square (left panel) and that is highly transcribed in the rWT shows more Pol II accumulation in intergenic regions associated to transcriptional read-through in CTD-Δ5 as compared to the adjacent regions that are less transcribed.

Given the low level of transcription by the Pol II without most of the CTD, I don't understand the conclusion that "The CTD appears also dispensable for the correct recruitment of Pol II at promoters and the process of transcription per se."
#R2.3 The major point of our article is that Pol II CTD-Δ5 can still transcribe and that even though its transcription is likely lower, Pol II accumulation is still detected at promoters, which remains the most likely recruitment areas, possibly because they have more open chromatin structures. We have nevertheless modified this statement by '*The CTD appears also dispensable for the recruitment of Pol II, most likely at promoters, and the process of transcription per se*' (*line 205*).

Also, from the Methods section, it appears that the chromatin RNA-seq and ChIP-seq reads have not been normalised to a spike-in control. As the delta 5 pol II has a global effect on transcription, spike-in normalization is necessary.
#R2.4 We agree with the referee that spike-in normalization would improve the quality of absolute quantitative statements in our article, which was not our main intention. In the case of the ChIP-seq, we have initially tried to apply such controls by adding drosophila chromatin to the samples but failed (see below). However, and importantly our statements relate to the qualitative read-through phenotype that monitor the level of transcripts (or Pol II) over the gene bodies compared to that of the adjacent intergenic regions. The statements relating to our observations would not change if the data would be spike-in-normalized but would essentially give an information about the differences for RNA/Pol II levels in the rWT and CTD-Δ5 cells. Thus, we did not draw strong conclusion on the absolute amount of RNA (nor polymerase molecules) produced but rather on the relative levels of the transcripts at variable locations of the genome. In the quantitative differential analysis of ChrRNA-seq in Fig. 1, we carefully use the term 'apparent' up-regulation of gene transcription' for describing the CTD-Δ5 feature. We have now added the following sentence at the end of the first result section to make it even more clear (*lines 142-144*): '*However, we note that this conclusion should be moderated given that these data are not spike-in normalized and that previous work have shown that the CTD-Δ5 shows overall less transcription activity.*'

Coming back on the spike-in procedures mostly used in RNA-seq data, they include ERCC synthetic transcripts that are added based on the amount of the total RNA extracted, but not the number of cells and are aimed at correcting technical variations linked to sequencing libraries preparation (Risso et al., 2014). Because most of the quantified transcripts are Pol I transcripts, this can introduce a strong bias in the data if the Pol I and Pol II transcripts are not regulated in the same manner, or may not display the same stability, which is often the case in many perturbations applied to the cells. Thus, optimizing a procedure based on the number of cells would probably be the best option but also present issues relating on the reliability of the precise count of cell number, a question we are dealing with in our laboratory. We considered this question out the frame of the present study because, again, we did not intend to compare the absolute but rather the relative number of transcripts.

Concerning the ChIP-seq experiment, we do often apply spike-in in our lab by using drosophila chromatin as proposed previously. However, this procedure can be robust essentially if the epitopes are conserved between human and drosophila. Since, we did not hold a cell line tagged for RPB1-HA in flies (using S2 cells, our standard procedure), we have tried nevertheless to add a small amount of drosophila chromatin to the human ChIP samples, hoping that the HA epitopes would reproducibly pull-down non-specific signal from the nucleus to normalize the ChIP. This was not the case since we recovered less HA signal originating from fly than from human cells, as

compared to the input DNA (see table below). Furthermore, the fraction of drosophila reads recovered was not consistent in our 2 biological replicates. However, and as shown on the table below, we could clearly see that in both replicates, there were more drosophila reads in CTD-Δ5 ChIP experiments as compared to rWT, suggesting that less Pol II molecules on chromatin were recovered overall in the mutant context. Because these quantitative trials were not robust enough, and since our main point was not to describe the absolute, but rather the relative, amounts of Pol II, we did not include their description in the manuscript.

	Sample name	Biological replicate	# Human reads	# Drosophila reads	% Drosophila reads
ChIP Pol-HA	rWT	AP1408 (rep1)	140727.594	2299400	1,6
		YY1171 (rep2)	72720098	2154964	2,9
	Δ5	AP1426 (rep1)	114617312	3073854	2,6
		YY1175 (rep2)	70417950	3723730	5,0
Input	rWT	AP1407 (rep1)	82930316	14909506	15,2
	Δ5	AP1425 (rep1)	62852000	13801670	18,0

In addition, it appears from the GEO submission that Input controls were not sequenced, which could affect the conclusions drawn from the ChIP-seq data.
#R2.5 We apologize for this omission. This experiment was in fact performed. Our GEO has now been updated with the missing input data and an example of this input is shown on Fig. EV3B. The amount of tags sequences for input is displayed in the table above.

A higher transcription elongation rate is also known to promote transcriptional readthrough. There is therefore the possibility that that the delta 5 mutant has a higher elongation rate than WT Pol II. From the proteomics data, are the elongation factors still binding to the delta 5 Pol II (DISF, STP6, SEC, PAF1c)?
#R2.6 We thank the reviewer for raising this point. Out of the mentioned factors, only SPT6 (SUPT6H) yields statistically significant result that can be interpretated (Table EV1) in our data. Interestingly SPT6 is lost in the context of the CTD-Δ5 mutant (p-value = $3.7x10^{-3}$). It has been indeed described that SPT6 influences positively the elongation rate (Narain et al., 2021), while it generally decreases its processivity. In the same work, the authors describe a readthrough phenotype that is restricted to long genes, while our defect is rather global. We have now described this SPT6 loss in our manuscript (*lines 252-256*): '*Interestingly, we also observed a loss of SPT6 in the mutant context (Table EV1), whose loss has been previously linked to a termination defect (Narain et al., 2021). This impairment is consistent with our observations even though the level of RT transcription is more global that the one previously described, that is not restricted to short genes.*'

Is the delta 5 Pol II known to be still phosphorylated on the five heptapeptides left?
#R2.7 We do not know the phosphorylation pattern of the CTD-Δ5. But on the western blot performed with the HA blot in Fig. 1C, less hyperphosphorylated band/smear can be observed, as compared with rWT and YFFF. So, we would speculate that it tends to be hypo-phosphorylated. Assessment of the Δ5 phosphorylation is difficult given the presence of remaining α-amanitin-blocked endogenous Pol II (Fig. 1C) that would impair the interpretation of the signals.

In addition, some attempt at analysing de novo splicing and polyadenylation could have been made as a defect in splicing would lead to readthrough transcription. If I understood well, the authors used total RNA for their poly(A) analysis, whereas it would have been better to analyse nuclear or chromatin-associated RNA or even RNA from inducible genes. Nascent techniques (TT-seq, POINT-seq, or others) would also have been useful.
#R2.8 We now provide an analysis to monitor the relative level of intronic vs exonic reads in the mutants using ChrRNA-seq data (Fig. 2E) suggesting a moderate splicing defect in the mutant.

We agree with the reviewer that other nascent techniques would have useful, especially for splicing outputs, but our past analyses and published data (Nojima et al., 2015) have shown that ChrRNA-seq yields very close read-outs as compared to NET-seq (to the exception of short transcripts). We will certainly push efforts to apply these procedures in the future. Concerning polyA RNA-seq analysis, we agree with the referee's comment as we indeed used total RNA and cannot rule out that the majority of transcripts isolated come from stabilized mRNA produced by endogenous Pol-II prior addition of α-amanitin.

Also, why is there no single gene ChIP validation and why analyse the chromatin RNA-seq for snRNA genes and histone genes but not the Pol II ChIP-seq, which should be more informative about transcription? If the spliceosome is still associated with chromatin, surely the mature snRNA will be picked up in the chromatin RNA-seq?
#R2.9 We have indeed performed single ChIP on model genes prior library preparation for ChIP-seq (shown below). But since this is a routine experiment, we did not include it originally in the manuscript. It is now presented as Fig. EV3A in the revised manuscript associated to the following sentence (*lines 185-187*): '. *After checking the HA ChIP enriched for Pol II on highly expressed target in both backgrounds (Figure EV3A), we prepared libraries for further high-throughput sequencing.'*
We have now profiled Pol II ChIP-seq data at snRNA and histone gene units in the manuscript that are presented in the new Fig. EV5E-F. These data indicate little apparent recruitment of Pol II at these locations. Overall, and since these results are in contrast to the ChRNA-seq, this suggests that snRNA and histone transcripts are stabilized (and/or poorly processed in the mutant). Given the observation on snRNAs, we have proposed the referee's suggestion as a possible hypothesis relating to the spliceosome in the manuscript (*lines 288-289*) and given previous description of its association to nuclear fraction in human cells (Girard et al., 2012):' *In the case of snRNA, one could thus speculate that spliceosome retention in chromatin fraction of the nucleus could assist this process'.*

The TAD analysis is well done and the conclusions are justified but this data doesn't add very much and the low level of transcription by the delta 5 might well not impact these.
#R2.10 We agree that this result is essentially negative with no effect of the readthrough and/or low transcription of the CTD-Δ5 on 3D genome organization. However, and since the question of the impact of transcription and readthrough transcription is still debated in the field, we believe this result is of interest for the community.

In conclusion, the authors have presented some intriguing and interesting data, particularly the results of the interactome analysis but some more detailed analysis of what is happening to elongation and RNA processing is missing.

Minor points
Is the delta 5 the same as the Bentley delta 5?
The CTD-Δ5 was originally described by the Corden lab in 1995 (Gerber et al., 1995) but was also used in further studies in the Bentley lab (McCracken et al., 1997) and others, including the Eick lab (Meininghaus et al., 2000; Meininghaus and Eick, 1999).

The English needs editing by a native English speaker.
We will proceed with English editing following the review process, should our article be accepted, once all scientific modifications of the manuscript will be integrated.

Referee #3:

In this manuscript Yahia et al. present the results describing the transcriptional properties of an RNA polymerase II mutant with a minimal length CTD. Transcription by the delta5 Pol II is examined using an established protocol in which an amanitin resistant large subunit gene is induced with doxycycline while the endogenous Pol II is inhibited with amanitin. This enables the

examination of transcription by delta5 Pol II in a narrow time window before the cells die from the effects of the CTD mutation. One of the nice features here is the inclusion of a Tyr mutants that has previously been shown to alter Pol II termination. The main conclusion of the authors is that truncation of the CTD has a greater effect on termination and in fact increased transcription of some genes compared to WT. The other major conclusion is that delta5 Pol II transcribes more widely than previously observed. The delta5 interactome indicates reduced interaction with the mediator raising the question of how delta5 can initiate. These are interesting observations that lay the groundwork for further studies on the role of the CTD.

We thank the reviewer for its positive input, below are the answer to his/her questions.

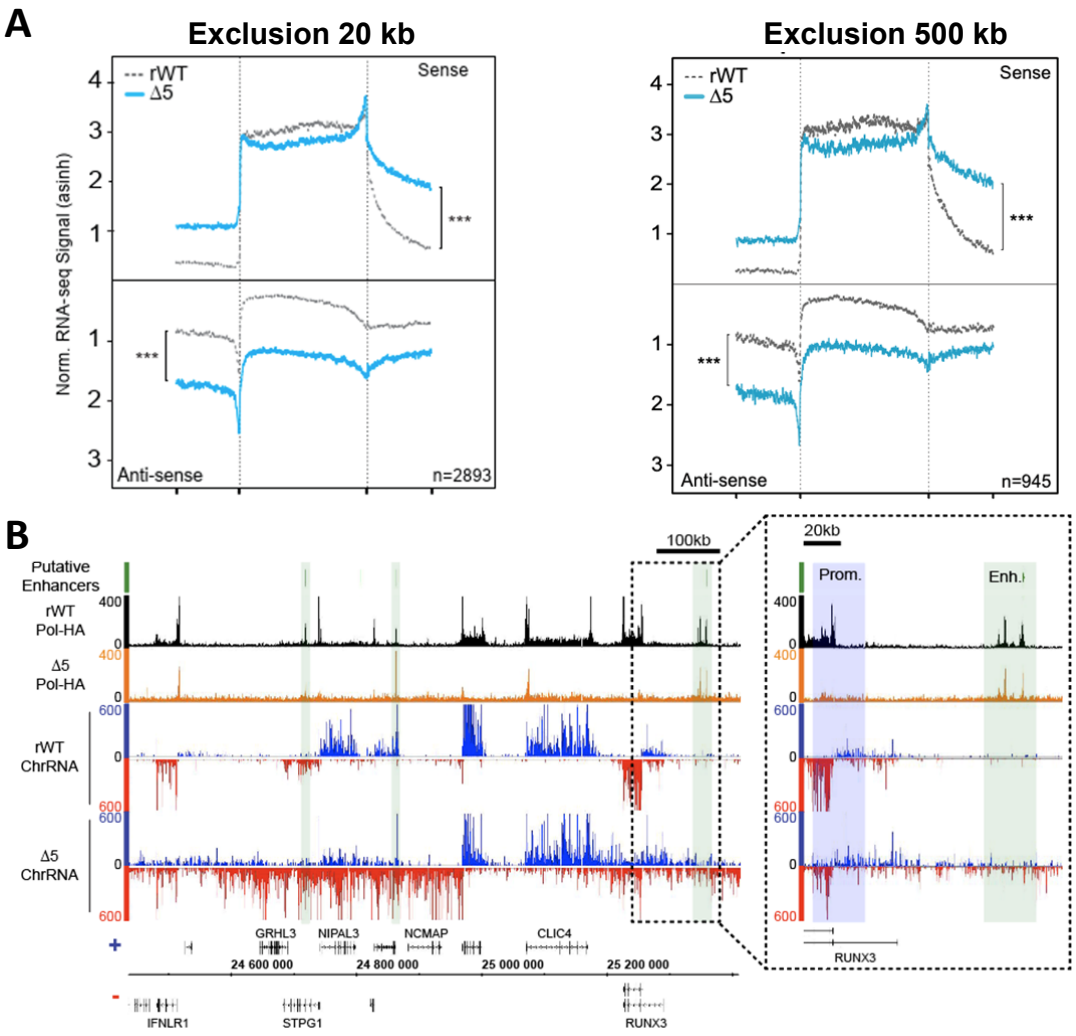
There are several minor concerns that I think could be addressed to improve the paper:

1. One problem with the interpretation of the results is that the RNAseq data are normalized to the depth of reads. Does this type of normalization over-emphasize the level of transcription in the delta5 strain? I would think that a spike-in control would provide a more accurate normalization.
#R3.1 As discussed in details in the answer of Referee 2 (#R2.4), we chose not to use spike-in since the existing standard procedures based on ERCC are not satisfactory. ERRC RNAs are to be added to the same RNA levels in between 2 given conditions. This strategy assumes that ribosomal RNA levels, which represent the majority of RNA quantified, are the same. This assessment is certainly incorrect in many cellular perturbations, especially in the case of Pol mutations. Unless a rigorous cell normalization-based method is used, these controls are not useful. In the quantitative differential analysis of ChrRNA-seq in Fig. 1, we carefully use the term 'apparent' up-regulation of gene transcription' for describing the CTD-Δ5 feature. We have now added the following sentence at the end of the first result section to make it even more clear (*lines 142-144*): *'However, we note that this conclusion should be moderated given that these data are not spike-in normalized and that previous work have shown that the CTD-Δ5 shows overall less transcription activity.*'

2. Another question is whether the observed global changes in transcription are due to CTD truncation directly or due to changes in expression of a gene or genes that may alter transcription. While not addressable given the experimental approach, this should be mentioned in the text.
#R3.2 We agree that gene expression might explain the defect of the CTD-Δ5 mutant. Indeed, Mediator and Integrator are the more striking proteins and protein complexes that lose their association to Pol II. As mentioned in the discussion (*lines 341-351*), the termination is fully consistent with recently observed features of mutant of these complexes and thus is the most likely explanation of the transcription defect. However, we have now included the possibility of gene expression impairment as an explanation (*lines 352-355*): '*Given these described features of Med and Int, and although we can't formally rule out that other gene(s) with impaired expression might contribute the transcriptional phenotype of CTD-D5, we propose that the loss of these two complexes essentially explain the termination defect in absence of the CTD repeats.*'

3. There seems to be more delta5 upstream of coding regions in the sense direction. Is this readthrough from upstream genes? This explanation would not be favored as the set of genes used in the analysis has had nearby genes removed. Could this indicate promiscuous initiation? Perhaps at nearby enhancers? If the delta5 polymerase can initiate at any open chromatin site then this could explain the "pervasive" nature of the delta5 occupancy. Perhaps long-read RNA sequence data could determine whether delta5 transcribes long RNAs or whether it is simply initiating promiscuously at multiple sites in open chromatin.

#R3.3 We favor the idea that the high ChrRNA-seq essentially originates from coding region readthrough signal. As shown in Fig. 1E and 2A, the readthrough phenotype of the CTD-Δ5 can span over hundreds of kb, so adjacent genes, even at long distance, could produce this signal. As shown below, the level of upstream signal decreases if the exclusion distance is increased to 500 kb, indicating that the level of readthrough depends at least partially on the distance (Referee



Referee Figure: A- Upstream readthrough RNA signal depends on the distance to adjacent genes. Left, ChrRNA-seq average gene profile, as described in Fig. 2B of the manuscript, by excluding adjacent genes in 20 kb areas. Right, as on the left panel by excluding adjacent genes within 500 kb. Note that readthrough signal is higher in the right panel suggesting that upstream readthrough depends on the distance to adjacent genes. **B- Enhancers do not elongate more transcripts than adjacent promoters.** Representative example of a Pol II-bound enhancer as defined in the article. While more Pol II remain bound at the enhancer location (green area) in the CTD-Δ5 mutant, less transcripts are elongated as compared to the Runx3 promoter (blue area). The ChrRNA transcript level around the enhancer is in the range of that detected in intergenic areas.

Fig. A below). Interestingly, in some cases this signal can result in apparent transcriptional interference as in the examples shown in Fig. 2F and EV2F. We disfavor the idea that enhancers (as described in Fig. 3C) are responsible for this phenotype in the CTD-Δ5 because even though significant Pol II is recruited at these locations, little nascent ChrRNA-seq signal accumulates as compared to promoters (Referee Fig. B below). This data suggests that Pol II can be recruited efficiently at enhancers and possibly initiate transcription but is unable to efficiently elongate these transcripts as it does at promoters. Please note that while ChrRNA-seq is a good readout

for nascent transcription, it does not score optimally for small RNAs accumulating at promoters and enhancers.

We have tried in the past to apply long-read sequencing for polyA RNAs of the CTD-Δ5, since it was not possible to do it for nascent transcripts given the very high amount of starting material required for sequencing. As expected, these experiments were highly biased for very stable and spliced transcripts and it was not possible to conclude on any defects relating to transcription.

4. The capping machinery is known to bind the CTD raising to question of whether the transcripts synthesized by delta5 capped? If not, are these transcripts subjected to nuclear quality control mechanisms?
In our proteomic analysis (Table EV1) we could detect significant loss of capping factors, it is thus difficult to conclude based on this data. We do not rule out that nuclear quality might be at play, possibly through the integrator loss of function as proposed recently (Lykke-Andersen et al., 2021). This specific reference has now been added in the discussion section (*line 347*).

5. At the end of the Discussion there is a typographical error referring to a "3' cap".
This was corrected.

Gerber, H.P., Hagmann, M., Seipel, K., Georgiev, O., West, M.A., Litingtung, Y., Schaffner, W., and Corden, J.L. (1995). RNA polymerase II C-terminal domain required for enhancer-driven transcription. Nature *374*, 660-662.
Girard, C., Will, C.L., Peng, J., Makarov, E.M., Kastner, B., Lemm, I., Urlaub, H., Hartmuth, K., and Lührmann, R. (2012). Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. Nature communications *3*, 994.
Lykke-Andersen, S., Žumer, K., Molska, E., Rouvière, J.O., Wu, G., Demel, C., Schwalb, B., Schmid, M., Cramer, P., and Jensen, T.H. (2021). Integrator is a genome-wide attenuator of non-productive transcription. Molecular cell *81*, 514-529.e516.
McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., and Bentley, D.L. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature *385*, 357-361.
Meininghaus, M., Chapman, R.D., Horndasch, M., and Eick, D. (2000). Conditional expression of RNA polymerase II in mammalian cells. Deletion of the carboxyl-terminal domain of the large subunit affects early steps in transcription. J Biol Chem *275*, 24375-24382.
Meininghaus, M., and Eick, D. (1999). Requirement of the carboxy-terminal domain of RNA polymerase II for the transcriptional activation of chromosomal c-fos and hsp70A genes. FEBS Lett *446*, 173-176.
Narain, A., Bhandare, P., Adhikari, B., Backes, S., Eilers, M., Dölken, L., Schlosser, A., Erhard, F., Baluapuri, A., and Wolf, E. (2021). Targeted protein degradation reveals a direct role of SPT6 in RNAPII elongation and termination. Mol Cell *81*, 3110-3127.e3114.
Nojima, T., Gomes, T., Grosso, A.R., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, N.J. (2015). Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. Cell *161*, 526-540.
Risso, D., Ngai, J., Speed, T.P., and Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. Nature biotechnology *32*, 896-902.
Shah, N., Maqbool, M.A., Yahia, Y., El Aabidine, A.Z., Esnault, C., Forne, I., Decker, T.M., Martin, D., Schuller, R., Krebs, S.*, et al.* (2018). Tyrosine-1 of RNA Polymerase II CTD Controls Global Termination of Gene Transcription in Mammals. Molecular cell *69*, 48-61 e46.

Tsao, D.C., Park, N.J., Nag, A., and Martinson, H.G. (2012). Prolonged α-amanitin treatment of cells for studying mutated polymerases causes degradation of DSIF160 and other proteins. Rna *18*, 222-229.

Dear Dr. Andrau,

Thank you for the submission of your revised manuscript. We have now received the enclosed reports from the referees, and I am happy to say that all support the publication of your work now. Referees 2 and 3 still have a few more suggestions regarding the manuscript text that I would like you to incorporate before we can proceed with the official acceptance of your manuscript.

A few editorial requests will also need to be addressed:

- Please move the Data Availability Section to the end of the materials and methods.

- Please correct the conflict of interest subheading to "Disclosure and Competing Interest Statement"

- The funding info in our online submission system and in the ms file do not match, please correct when you upload the final ms.

- DATASET EV LEGENDS: The EV tables should be renamed as Dataset EV# files, correct the callouts. Remove the colour from Table EV1.

- I attach to this email a related ms file with comments by our data editors. Please address all comments in the final ms.

EMBO press papers are accompanied online by A) a short (1-2 sentences) summary of the findings and their significance, B) 2-3 bullet points highlighting key results and C) a synopsis image that is exactly 550 pixels wide and 200-600 pixels high (the height is variable). You can either show a model or key data in the synopsis image. Please note that text needs to be readable at the final size. Please send us this information along with the final manuscript.

I would like to suggest some minor changes to the title and abstract that needs to be written in present tense. Please let me know whether you agree with these changes, and please also specify the kind of cells used in your study (mammalian, or else). The term "in vivo" is usually used for animal studies, so the title should be modified:

RNA Polymerase II CTD is dispensable for transcription and required for termination

The largest subunit of RNA polymerase (Pol) II harbors an evolutionary conserved C-Terminal-Domain (CTD), composed of a repetition of heptapeptides, central in the transcriptional process. Here, we analyze the transcriptional phenotypes of a CTD-delta5 mutant that carries a large CTD truncation. Our data show that this mutant can transcribe genes in living cells but displays a pervasive phenotype with impaired termination, similar to but more severe than previously characterized mutations of CTD tyrosine residues. The CTD-delta5 mutant does not interact with the Mediator and Integrator complexes [OK?] involved in activation of transcription and maturation of RNAs. Examination of long-distance interactions and CTCF binding patterns in CTD-delta5 mutant cells reveals no changes in TAD domains or borders. Our data demonstrate that the CTD is largely dispensable for the act of transcription in living cells. We propose a model in which CTD-depleted Pol II has a lower entry rate into DNA but becomes pervasive once engaged in transcription, resulting in a loss of termination.

I look forward to seeing a final version of your manuscript as soon as possible. Please use this link to submit your revision: https://embor.msubmit.net/cgi-bin/main.plex

Best regards,
Esther

Esther Schnapp, PhD
Senior Editor
EMBO reports

Referee #1:

The authors have adequately addressed the points raised during the previous review. The revised manuscript is significantly improved.

Referee #2:

The authors have done a very good job to revise the manuscript according to the reviewers' comments. However, I don't understand what the authors mean exactly by these added clarifying statements-

"Although we cannot completely rule out that the chrRNA procedure could lead to transcripts retention, we believe this hypothesis unlikely since it would imply that WT cells produce such transcripts to a massive level, which has never been described."
and
"In the case of snRNA, one could thus speculate that spliceosome retention in chromatin fraction of the nucleus could assist this process."
and therefore suggest that they ask a native English speaker to help craft clearer statements when they are going over the manuscript to polish the English.

Referee #3:

In this revised manuscript the authors have addressed most concerns raised by reviewers. The paper resents an interesting and comprehensive analysis of transcription by Pol II lacking most of the CTD (CTDdelta5). This is a challenging task as the CTD performs essential functions and the experimental setup is looking at a narrow window following inhibition of the WT Pol II and cell death due to the lack of CTD repeats on the remaining mutant Pol II. Global changes in transcription are expected but how to measure these changes has been difficult. The authors have examined transcripts bound to chromatin by RNA-seq (ChrRNA) as a proxy for nascent transcripts with the caveat that some transcripts may not be nascent but retained due to defects in processing. The main observation from the ChrRNA analysis is that there is an increase in transcription of many genes and extensive readthrough of Pol II termination sites. The problem with this analysis is that the data are normalized to read depth. If there is much lower transcription in CTDdelta5 cells, then subtle changes will be exaggerated by normalization. The authors argue that spike in controls are not appropriate due to possible changes in the bulk of Pol I transcripts. This may be true, although it seems no less warranted than normalizing to read depth. Having an orthologous control would be helpful. The authors argue that WT cells do not produce readthrough transcripts at a "massive" level. This is probably true but in some cases the WT readthrough may be amplified by the normalization process. This possibility is still not clear from the revised text. Despite this shortcoming I feel that on balance this manuscript provides substantial new and interesting data that will certainly provoke deeper examination of the role of the CTD in all aspects of the Pol II transcription cycle.

Referee #1:

The authors have adequately addressed the points raised during the previous review. The revised manuscript is significantly improved.

We thank the referee for his comments.

Referee #2:

The authors have done a very good job to revise the manuscript according to the reviewers' comments. However, I don't understand what the authors mean exactly by these added clarifying statements-
"Although we cannot completely rule out that the chrRNA procedure could lead to transcripts retention, we believe this hypothesis unlikely since it would imply that WT cells produce such transcripts to a massive level, which has never been described."

We have now changed this sentence by:
Although we cannot completely rule out that the apparent read-through phenotype of the CTD-Δ5 originates from transcript retention on the chromatin, we believe this hypothesis unlikely since it would imply that WT cells produce such transcripts to a massive level, which has never been described.

and
"In the case of snRNA, one could thus speculate that spliceosome retention in chromatin fraction of the nucleus could assist this process."

We have changed this sentence by:
In the case of snRNA genes, one could speculate that spliceosome retention in the chromatin fraction of the nucleus could cause this defect.

and therefore suggest that they ask a native English speaker to help craft clearer statements when they are going over the manuscript to polish the English.

We have asked a native English speaker (Tom Sexton, one of the authors) to proofread the manuscript and these corrections were integrated in the revised manuscript.

Referee #3:

In this revised manuscript the authors have addressed most concerns raised by reviewers. The paper resents an interesting and comprehensive analysis of transcription by Pol II lacking most of the CTD (CTDdelta5). This is a challenging task as the CTD performs essential functions and the experimental setup is looking at a narrow window following inhibition of the WT Pol II and cell death due to the lack of CTD repeats on the remaining mutant Pol II. Global changes in transcription are expected but how to measure these changes has been difficult. The authors have examined transcripts bound to chromatin by RNA-seq (ChrRNA) as a proxy for nascent transcripts with the caveat that some transcripts may not be nascent but retained due to defects in processing. The main observation from the ChrRNA analysis is that there is an increase in transcription of many genes and extensive readthrough of Pol II termination sites. The problem with this analysis is that the data are normalized to read depth. If there is much lower transcription in CTDdelta5 cells, then subtle changes will be exaggerated by normalization. The authors argue that spike in controls are

not appropriate due to possible changes in the bulk of Pol I transcripts. This may be true, although it seems no less warranted than normalizing to read depth. Having an orthologous control would be helpful. The authors argue that WT cells do not produce readthrough transcripts at a "massive" level. This is probably true but in some cases the WT readthrough may be amplified by the normalization process.

While we agree that the normalization might change how the data looks at the end, we have used 2 modes of normalization shown in the main and supplementary figures (for example, in Figure 2B, EV2B and EV2C). One in normalized on the number of counts, the other based on the signals over the gene bodies. Both indicate the same result. Furthermore, we have inserted in the revised manuscript the following sentence, which we believe to address this question (line 139):
However, we note that this conclusion should be moderated given that these data are not spike-in normalized and that previous work have shown that the CTD-D5 shows overall less transcription activity.

This possibility is still not clear from the revised text. Despite this shortcoming I feel that on balance this manuscript provides substantial new and interesting data that will certainly provoke deeper examination of the role of the CTD in all aspects of the Pol II transcription cycle.

We thank the referee for his positive assessment of the manuscript.

Dr. Jean-Christophe Andrau
Institut de Génétique Moléculaire de Montpellier (IGMM)
Transcription and Epigenomics in developing T-cells
UMR5535 CNRS
1919 Route de Mende
Montpellier Cedex 5, 34293
France

Dear Dr. Andrau,

I am very pleased to accept your manuscript for publication in the next available issue of EMBO reports. Thank you for your contribution to our journal.

At the end of this email I include important information about how to proceed. Please ensure that you take the time to read the information and complete and return the necessary forms to allow us to publish your manuscript as quickly as possible.

As part of the EMBO publication's Transparent Editorial Process, EMBO reports publishes online a Review Process File to accompany accepted manuscripts. As you are aware, this File will be published in conjunction with your paper and will include the referee reports, your point-by-point response and all pertinent correspondence relating to the manuscript.

If you do NOT want this File to be published, please inform the editorial office within 2 days, if you have not done so already, otherwise the File will be published by default [contact: emboreports@embo.org]. If you do opt out, the Review Process File link will point to the following statement: "No Review Process File is available with this article, as the authors have chosen not to make the review process public in this case."

Thank you again for your contribution to EMBO reports and congratulations on a successful publication. Please consider us again in the future for your most exciting work.

Yours sincerely,

Esther Schnapp, PhD
Senior Editor
EMBO reports

****

THINGS TO DO NOW:

Please note that you will be contacted by Wiley Author Services to complete licensing and payment information. The required 'Page Charges Authorization Form' is available here: https://www.embopress.org/pb-assets/embo-site/er_apc.pdf - please download and complete the form and return to embopressproduction@wiley.com

EMBO Press participates in many Publish and Read agreements that allow authors to publish Open Access with reduced/no publication charges. Check your eligibility: https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/affiliation-policies-payments/index.html

You will receive proofs by e-mail approximately 2-3 weeks after all relevant files have been sent to our Production Office; you should return your corrections within 2 days of receiving the proofs.

Please inform us if there is likely to be any difficulty in reaching you at the above address at that time. Failure to meet our deadlines may result in a delay of publication, or publication without your corrections.

All further communications concerning your paper should quote reference number EMBOR-2022-56150V3 and be addressed to emboreports@wiley.com.

Should you be planning a Press Release on your article, please get in contact with emboreports@wiley.com as early as possible, in order to coordinate publication and release dates.

EMBO Press Author Checklist

Corresponding Author Name: Jean-Christophe ANDRAU
Journal Submitted to: EMBO Reports
Manuscript Number: 56150V2

USEFUL LINKS FOR COMPLETING THIS FORM
The EMBO Journal - Author Guidelines
EMBO Reports - Author Guidelines
Molecular Systems Biology - Author Guidelines
EMBO Molecular Medicine - Author Guidelines

Reporting Checklist for Life Science Articles (updated January 2022)
This checklist is adapted from Materials Design Analysis Reporting (MDAR) Checklist for Authors. MDAR establishes a minimum set of requirements in transparent reporting in the life sciences (see Statement of Task: 10.31222/osf.io/9sm4x). Please follow the journal's guidelines in preparing your manuscript.
Please note that a copy of this checklist will be published alongside your article.

Abridged guidelines for 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.
- ideally, figure panels should include only measurements that are directly comparable to each other and obtained with the same assay.
- plots 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. Any statistical test employed should be justified.
- Source Data should be included to report the data underlying figures according to the guidelines set out in the authorship 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 χ2 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.

Please complete ALL of the questions below.
Select "Not Applicable" only when the requested information is not relevant for your study.

Materials

Newly Created Materials	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
New materials and reagents need to be available; do any restrictions apply?	Not Applicable	

Antibodies	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
For **antibodies** provide the following information: - Commercial antibodies: RRID (if possible) or supplier name, catalogue number and or/clone number - Non-commercial: RRID or citation	Yes	Materials and methods

DNA and RNA sequences	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Short novel DNA or RNA including primers, probes: provide the sequences.	Yes	Materials and methods

Cell materials	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Cell lines: Provide species information, strain. Provide accession number in repository **OR** supplier name, catalog number, clone number, and/**OR** RRID.	Yes	Materials and methods/Data availability
Primary cultures: Provide species, strain, sex of origin, genetic modification status.	Not Applicable	
Report if the cell lines were recently **authenticated** (e.g., by STR profiling) and tested for mycoplasma contamination.	Yes	Materials and methods

Experimental animals	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Laboratory animals or Model organisms: Provide species, strain, sex, age, genetic modification status. Provide accession number in repository **OR** supplier name, catalog number, clone number, **OR** RRID.	Not Applicable	
Animal observed in or captured from the field: Provide species, sex, and age where possible.	Not Applicable	
Please detail **housing and husbandry conditions**.	Not Applicable	

Plants and microbes	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Plants: provide species and strain, ecotype and cultivar where relevant, unique accession number if available, and source (including location for collected wild specimens).	Not Applicable	
Microbes: provide species and strain, unique accession number if available, and source.	Not Applicable	

Human research participants	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
If collected and within the bounds of privacy constraints report on age, sex and gender or ethnicity for all study participants.	Not Applicable	

Core facilities	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
If your work benefited from core facilities, was their service mentioned in the acknowledgments section?	Not Applicable	

Design

Study protocol	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
If study protocol has been **pre-registered, provide DOI in the manuscript**. For clinical trials, provide the trial registration number **OR** cite DOI.	Not Applicable	
Report the **clinical trial registration number** (at ClinicalTrials.gov or equivalent), where applicable.	Not Applicable	

Laboratory protocol	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Provide DOI OR other citation details if **external detailed step-by-step protocols** are available.	Yes	Materials and methods/References

Experimental study design and statistics	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Include a statement about **sample size** estimate even if no statistical methods were used.	Yes	Figures
Were any steps taken to minimize the effects of subjective bias when allocating animals/samples to treatment (e.g. **randomization procedure**)? If yes, have they been described?	Not Applicable	
Include a statement about **blinding** even if no blinding was done.	Not Applicable	
Describe **inclusion/exclusion criteria** if samples or animals were excluded from the analysis. Were the criteria pre-established? If sample or data points were omitted from analysis, report if this was due to attrition or intentional exclusion and provide justification.	Yes	Materials and methods
For every figure, are **statistical tests** justified as appropriate? Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it. Is there an estimate of variation within each group of data? Is the variance similar between the groups that are being statistically compared?	Yes	Materials and methods

Sample definition and in-laboratory replication	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
In the figure legends: state number of times the experiment was **replicated** in laboratory.	Yes	Figures/Data availability
In the figure legends: define whether data describe **technical or biological replicates**.	Yes	Figures/Data availability

Ethics

Ethics	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Studies involving **human participants**: State details of **authority granting ethics approval** (IRB or equivalent committee(s), provide reference number for approval.	Not Applicable	
Studies involving **human participants**: 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.	Not Applicable	
Studies involving **human participants:** For publication of **patient photos**, include a statement confirming that consent to publish was obtained.	Not Applicable	
Studies involving experimental **animals**: State details of **authority granting ethics approval** (IRB or equivalent committee(s), provide reference number for approval. Include a statement of compliance with ethical regulations.	Not Applicable	
Studies involving **specimen and field samples:** State if relevant **permits** obtained, provide details of authority approving study; if none were required, explain why.	Not Applicable	

Dual Use Research of Concern (DURC)	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Could your study fall under dual use research restrictions? Please check biosecurity documents and list of **select agents and toxins** (CDC): https://www.selectagents.gov/sat/list.htm	Not Applicable	
If you used a select agent, is the security level of the lab appropriate and reported in the manuscript?	Not Applicable	
If a study is subject to dual use research of concern regulations, is the name of the **authority granting approval and reference number** for the regulatory approval provided in the manuscript?	Not Applicable	

Reporting

The MDAR framework recommends adoption of discipline-specific guidelines, established and endorsed through community initiatives. Journals have their own policy about requiring specific guidelines and recommendations to complement MDAR.

Adherence to community standards	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
State if relevant guidelines or checklists (e.g., **ICMJE, MIBBI, ARRIVE, PRISMA**) have been followed or provided.	Not Applicable	
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.	Not Applicable	
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.	Not Applicable	

Data Availability

Data availability	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Have **primary datasets** been deposited according to the journal's guidelines (see 'Data Deposition' section) and the respective accession numbers provided in the Data Availability Section?	Yes	Data availability
Were **human clinical and genomic datasets** deposited in a public access-controlled repository in accordance to ethical obligations to the patients and to the applicable consent agreement?	Not Applicable	
Are **computational models** that are central and integral to a study available without restrictions in a machine-readable form? Were the relevant accession numbers or links provided?	Not Applicable	
If publicly available data were reused, provide the respective **data citations in the reference list**.	Not Applicable	