THE
EMBO
JOURNAL

High-resolution transcription maps reveal the widespread impact of roadblock termination in yeast

Tito Candelli, Drice Challal, Jean-Baptiste Briand, Jocelyne Boulay, Odil Porrua, Jessie Colin, Domenico Libri

Review timeline:

Submission date:	2 June 2017
Editorial Decision:	19 July 2017
Revision received:	19 October 2017
Editorial Decision:	9 November 2017
Revision received:	14 December 2017
Accepted:	15 December 2017

Editor: Anne Nielsen

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. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision 19 July 2017

Thank you for submitting your manuscript for consideration by the EMBO Journal. It has now been seen by three referees whose comments are shown below.

As you will see from the reports, all referees express interest in the findings reported in your manuscript, although they also raise a number of points that will have to be addressed before they can support publication of your manuscript here. Referees #1 and #2 are the more positive and mainly ask for additional discussions/clarifications. On the other hand, ref #3 is concerned that the resolution obtained by CRAC is not high enough to support the Rap1 roadblock model and that this should be supplemented using other techniques. I realise that setting up a whole new round of NET-Seq or PAR-CLIP may be outside the scope of a revision but I would encourage you to address this point experimentally as far as possible. Referee #3 also asks you to elaborate on the generality of roadblock termination, both in terms of the number of sites using this mechanism and the amplitude of regulation at those sites. From our side, these are relevant points that would clearly strengthen the overall conclusiveness and impact of the revised manuscript.

Given the referees' positive recommendations, I would like to invite you to submit a revised version of the manuscript, addressing the comments of all three reviewers.

Thank you for the opportunity to consider your work for publication. I look forward to your revision.

REFEREE REPORTS

Referee #1:

This paper presents convincing evidence that the inhibition of transcription elongation, followed by termination of the stalled polymerase is a common feature associated with DNA binding factors.

The data add to our understanding of transcription patterns and the interplay between genes, particularly within compact fungal genomes. The work appears to have been well performed and will be of wide interest in the field of transcription. I have only minor comments.

1) Some of the apparent peaks seem surprisingly far upstream. Can the authors relate the site of RNAPII CRAC signals to the protein binding sites in terms of the RNAPII and RP protein foot prints

2) Fig. 4: Without Rap1 see decreased pausing at RB and decreased TX downstream. Without Rna15 see see increased signal at RB and increased TX downstream. Looks like effects of both RB and readthrough on downstream genes are quite different depending on presence of pA site? Can the authors comment.

3) The start of the results section might be better placed in the Introduction as it describes a previously reported system and is similar to the start of the Colin et al. 2014 MS, including Fig. 1B.

Referee #2:

This lab has previously published the existence of a new mechanism of transcription termination called "roadblock" termination, in which the RNA polII elongation is literally blocked by a DNA-bound general transcription factor Reb1 (Collin et al., 2014). The present manuscript extends this observation, showing that "roadblock termination" can be mediated by other DNA-bound factors, in particular other general transcription terminators. One of the main messages of the manuscript is that a significant background of transcriptional read-through occurs at canonical terminators (CPF-CF and NNS types) and that roadblock termination plays an important role as a backup of "foolproof " mechanism to protect neighbouring genes from this natural source of pervasive transcription. This observation is obviously of "general interest" for the scientific community interested in gene expression.

Overall, the data adequately support the conclusions of the manuscript.

The conclusions presented in this manuscript substantially overlap the results recently published by another group (Roy et al., Genome Research, 2016). This paper was mainly based on RNA 3'-end mapping, while the experimental approach of the present manuscript is mainly relying on an improved CLIP-seq method to precisely map the elongating PolII. In a sense, the two papers complement each other rather nicely. There are discrepancies in some of the interpretations. The main one, which is discussed a bit lengthily in the present manuscript, is that the Roy paper proposes that the NNS and roadblock mechanisms cooperate to promote efficient transcription termination, while the present manuscript rather favours a "foolproof" mechanism, by which the roadblock termination acts as a backup for polymerases that would have escaped classical termination.
In my opinion, the two explanations are not mutually exclusive and there differences might appear in part semantic.

Minor comments:

I think that a few modifications could improve the manuscript significantly.

Fig. EV3B and EV3C have been inversed in the text.
Also, I am not sure that the Malabat et al. paper is the most appropriate dataset to document the point discussed there. The TIFF-seq manuscript (Pelechano et al, Nature 2013 - doi: 10.1038/nature12121) would seem much more appropriate.

Importantly, for all meta-gene analyses, the number of loci included in the analysis must be included in the figure. Even better, in addition to that, a table giving the identity of all genes used in each of these figures should be provided.

There is one thing I do not understand at all. In Figure 4 A and B, the yellow bar is supposed to represent an "operationally defined region of termination". I don't understand what that means. The

region is located upstream of the polyA sites. That transcription termination would terminate before the polyA site is obviously impossible. Numerous data in the literature shows that it occurs in yeast at something between 50 and 150 nucleotides downstream of the polyA site. Also, similar experiments reported in the literature (mainly relying on the PAR-CLIP method; see for example figure 7C in Creamer et al. PLoS Genetics 2016 - doi:10.1371/journal.pgen.1002329), show an inverse profile, with polII accumulating around the polyA site. The authors should thus discuss this difference (is it due to there modified CRAC method?) and explain what they mean by an "operationally defined region of termination", which in any case cannot represent the actual sequence region where transcription actually terminates.

Referee #3:

This paper provides a nice follow up to the earlier paper by Colin et al. that demonstrated the ability of Reb1 to block Pol II elongation both in vivo and in vitro. In this paper Candelli et al provide evidence that Rap1 and several other transcription factors can act as roadblocks (RBs) to Pol II elongation. The use of a UV cross-linking approach potentially expands this analysis genome wide.

My main concerns with this manuscript are the experimental design, the generality of the RB mechanism and the lack of high-resolution data showing precisely where Pol II pauses or terminates.

1. My first major concern has to do with the experimental design. The authors are using UV irradiation to cross-link proteins to nucleic acids. The CRAC procedure isolates RNA that is cross-linked to RNA polymerase thus mapping the position of the polymerase at the time of formation of the crosslink. Irradiation with 254 nm UV will also cross-link proteins to DNA and this may confound the interpretation of these experiments. If DNA-binding proteins like Rap1 are efficiently cross-linked then they may provide an artificial barrier to Pol II elongation. During the 50 second period of irradiation Pol II could elongate several thousand bases. If Rap1 is more efficiently cross-linked compared to Pol II then Pol II may accumulate at the cross-linked Rap1 sites. The results shown in Figure 1 show that there is a Rap1 dependent termination or pause in the absence of UV but is this the case genome wide? The authors should provide alternative data from NET-seq or PAR-CLIP results showing peaks of Pol II adjacent to downstream Rap1 binding sites in the absence of UV.

2. A second major concern is in the generality of the RB phenomenon. The authors show several metasite analyses of Pol II occupancy aligned with Rap1 (or other transcription factor) binding sites. In none of these plots are we told how many binding sites were included in the analysis. In Materials and Methods the authors state that the analysis was restricted to Rap1 sites within 300 bp of a poly(A) site. How many examples of this orientation were analyzed? The same concern applies to the centromere and tRNA analysis. The authors should include in the supplemental material a list of the sites used in each metasite analysis.

In addition to knowing how many RBs were analyzed in the metasite analysis it would be nice to know how many actually demonstrate the RB phenomenon. The authors explain that any values more than 5 SD above the mean were excluded. How many sites were excluded? It would also be nice to know how many were below the mean. Do all Rap1 binding sites act as RBs? Are there some that do not? Does the ability to act as a RB correlate with the different types of Rap1-dependent promoters characterized by the Shore lab? Does the RB ability of Rap1 correlate with the occupancy determined by Rhee and Pugh? Answering these questions could go some distance toward supporting the generality of the RB phenomenon.

3. A third problem with the paper is that the authors do not show any high-resolution data demonstrating a Rap1 or other RB. In figure 1 we are shown a Northern of a small RNA "of a size compatible with termination occurring immediately upstream of the Rap1 site". Without size markers on the blot and an indication of the distance between the TSS and Rap1 site it is difficult to judge the accuracy of this claim. Have the authors performed 3' RACE or RNA protection assays to more precisely map the 3'-ends of this short RNA? Including this data would offer stronger support for the specificity of the RB mechanism.

In figure 2 the Pol II occupancy maps are at too low a resolution to see precisely where transcription is blocked. A nucleotide level map would be helpful in this regard. The reason for this concern is based on the metasite analysis shown in figure 3A (left panel). In this figure it looks like the peak of Pol II extends beyond the upstream edge of the Rap1 binding site. Does this argue that some RBs are not steric in nature? This same concern applies to the downstream edge of the centromere where the metasite peak of Pol II extends into the CBF3 binding site Figure 7A). In the individual CEN14 map Pol II pauses well upstream of the CBF3 binding site. Is CEN14 an outlier?

Other comments.

4. In Figure 1A how many Rap1 terminators were used to make the logo? Do all of these synthetic Rap1 sites act as RBs?

5. The Pol II occupancy maps have no quantification on the Y axis. This information is necessary for the reader to compare the various tracks. Is the readthrough after anchor away at the same scale as the no rapamycin track? The authors should add units to the maps or state in the legend that all maps are at the same scale.

6. In the text on page 9 I believe the authors have referred incorrectly to figures EV3C and EV3B.

7. In Figure 7 it is not clear which data is PAR-CLIP and which is CRAC. Panel 7A is labeled Pol II occupancy but the data is not specified. The origin of the data in 7C is also unclear.

This paper is potentially very important if the authors can strengthen their claim that the RB phenomenon occurs genome wide. Addressing the concerns outlined here will make this paper more convincing. Detailed maps of the RBs described in this paper will provide an important resource for researchers interested in probing the roles of ncRNAs.

1st Revision - authors' response 19 October 2017

Answers to referees' comments (bold)

referee #1:

This paper presents convincing evidence that the inhibition of transcription elongation, followed by termination of the stalled polymerase is a common feature associated with DNA binding factors. The data add to our understanding of transcription patterns and the interplay between genes, particularly within compact fungal genomes. The work appears to have been well performed and will be of wide interest in the field of transcription. I have only minor comments.

We thank the reviewer for appreciating our work.

1) Some of the apparent peaks seem surprisingly far upstream. Can the authors relate the site of RNAPII CRAC signals to the protein binding sites in terms of the RNAPII and RP protein foot prints

Unfortunately, to our knowledge, the precise footprinting *in vivo* of many of the RB factors studied here is not known. The footprinting of the polymerase is around 35nt, which fits very nicely with the 15-20nt distance of the catalytic center of the roadblocked polymerase from the site of Reb1 and Rap1 binding. For tRNAs we have shown in figure 7 the footprints of TFIIIB and TFIIIC and, again, the data fit with the notion of uncompressible collision between the leading edge of the polymerase and the

edge of the DNA-bound factor (e.g. the RB peak is at position -15 relative to TFIIIB binding site). For some transcription factors (e.g. figure EV3) the RB peak is wider, and sometimes located a bit upstream, possibly consistent with the notion of other factors bound upstream of the site of alignment. Discussing the topology of these sites of collision would be quite speculative, and studying them would be probably beyond the scope of the study.

2) Fig. 4: Without Rap1 see decreased pausing at RB and decreased TX downstream. Without Rna15 see increased signal at RB and increased TX downstream. Looks like effects of both RB and readthrough on downstream genes are quite different depending on presence of pA site? Can the authors comment.

The effects of Rap1 depletion or Rna15 mutation are indeed different. When Rap1 is depleted, transcription goes through the roadblock and the pausing peak decreases. The increase in transcription downstream is not clearly seen in the metagene analysis, probably because the signal due to readthrough polymerases combines with that of downstream transcription, which generally decreases because of the absence of Rap1 (lack of activation or protection, or both). But increase in the levels of readthrough transcription is clearly seen in the individual cases shown in figure 2. When Rna15 is mutated, polymerases reading through the poly(A) signal from the upstream gene accumulate at the roadblock and to some extent go through it (the efficiency of the block is probably similar, but the flow is increased), invading the downstream promoter. This leads to some levels of transcription interference, and a general decrease in the expression of downstream genes. Therefore the signal increases in the region immediately downstream of the roadblock, as noticed by the referee, but decreases later (this is better appreciated in figure 3 for both Reb1 and Rap1). Although we did not analyse this in detail, we do not think that the presence of a poly(A) signal affects the behaviour of polymerases that are blocked downstream (in figure 2B a RB peak is observed clearly, and the difference in intensity relative to fig 2A is most likely correlated with the different temperature at which the assay was done).

3) The start of the results section might be better placed in the Introduction as it describes a previously reported system and is similar to the start of the Colin et al. 2014 MS, including Fig. 1B.

We described the general results of our previous work in the introduction. In the results section we chose to briefly describe the experimental system, which in our opinion belongs to the results section. We do not feel strongly about this, maybe the editor could advise if the paper is accepted.

Referee #2:

This lab has previously published the existence of a new mechanism of transcription termination called "roadblock" termination, in which the RNA polII elongation is literally blocked by a DNA-bound general transcription factor Reb1 (Collin et al., 2014). The present manuscript extends this observation, showing that "roadblock termination" can be mediated by other DNA-bound factors, in particular other general transcription terminators. One of the main messages of the manuscript is that a significant background of transcriptional read-through occurs at canonical terminators (CPF-CF and NNS types) and that roadblock termination plays an important role as a backup of "foolproof " mechanism to protect neighbouring genes from this natural source of pervasive transcription. This observation is obviously of "general interest" for the scientific community interested in gene expression.

© European Molecular Biology Organization

Overall, the data adequately support the conclusions of the manuscript.

The conclusions presented in this manuscript substantially overlap the results recently published by another group (Roy et al., Genome Research, 2016). This paper was mainly based on RNA 3'-end mapping, while the experimental approach of the present manuscript is mainly relying on an improved CLIP-seq method to precisely map the elongating PolII. In a sense, the two papers complement each other rather nicely. There are discrepancies in some of the interpretations. The main one, which is discussed a bit lengthily in the present manuscript, is that the Roy paper proposes that the NNS and roadblock mechanisms cooperate to promote efficient transcription termination, while the present manuscript rather favours a "foolproof" mechanism, by which the roadblock termination acts as a backup for polymerases that would have escaped classical termination.
In my opinion, the two explanations are not mutually exclusive and there differences might appear in part semantic.

We also thank this referee for appreciating our work and for the suggestions to improve it.

Minor comments:

I think that a few modifications could improve the manuscript significantly.

Fig. EV3B and EV3C have been inversed in the text.
Also, I am not sure that the Malabat et al. paper is the most appropriate dataset to document the point discussed there. The TIFF-seq manuscript (Pelechano et al, Nature 2013 - doi: 10.1038/nature12121) would seem much more appropriate.

We thank the referee for pointing this out. We inversed panels B and C in figure EV3 (now EV2). We preferred using the data from Malabat et al. as this set is deeper for the analysis of 5' ends of transcripts and also contains data on unstable transcripts (stabilized in *rrp6*Δ and *upf1*Δ*)*. Because the scope of this analysis is to detect additional initiation sites in intergenic regions, which we generally failed to identify, it was important to use the largest possible dataset in terms of coverage and RNA abundance.

Importantly, for all meta-gene analyses, the number of loci included in the analysis must be included in the figure. Even better, in addition to that, a table giving the identity of all genes used in each of these figures should be provided.

We apologize, a table with the number of sites for each figure had been prepared but not uploaded by mistake. We have now extended this table with all the coordinates of the sites used for every figure.

There is one thing I do not understand at all. In Figure 4 A and B, the yellow bar is supposed to represent an "operationally defined region of termination". I don't understand what that means. The region is located upstream of the polyA sites. That transcription termination would terminate before the polyA site is obviously impossible. Numerous data in the literature shows that it occurs in yeast at something between 50 and 150 nucleotides downstream of the polyA site. Also, similar experiments reported in the literature (mainly relying on the PAR-CLIP method; see for example figure 7C in Creamer et al. PLoS Genetics 2016 - doi:10.1371/journal.pgen.1002329), show an inverse profile, with polII accumulating around the polyA site. The authors should thus discuss this difference (is it due to there modified CRAC method?) and explain what they mean by an "operationally defined region of termination", which in any case cannot represent the actual sequence region where transcription actually terminates.

© European Molecular Biology Organization

This is a very important point and we agree with the referee that additional clarification is required. We clearly observed, as reported in the results section, a decrease in RNAPII occupancy in the region around the pA site, often starting before the main site. This is also observed in the CRAC data from the Tollervey lab (data not shown), indicating that the observation is not due to the modifications we introduced.

The most direct and high resolution analyses of termination have been performed by ParCLIP (Corden lab) and NetSeq (Churchman and Weissman labs). We compared our data to the latest NetSeq data from the Churchman lab (Harlen et al, 2016), and the ParCLIP data from the Corden lab (Schaughency et al., 2014) (see attached file for metasite comparison and a couple of representative examples). To facilitate comparison with the NETseq data, we only used the 3'-end of reads from our datase (as in NETseq). Comparison with the ParCLIP was done using the whole read sequence.

NetSeq and CRAC profiles are very similar but markedly different from the ParCLIP profile in the 3' and of genes and after the pA site, where a strong RNAPII peak is often observed. The difference between NetSeq and ParCLIP was already noticed by Creamer et al., who suggested that crosslinking might trap a labile intermediate in the termination reaction that would be lost in NetSeq. This explanation appears now unlikely, because the marked average increase in RNAPII occupancy observed in ParCLIP is not observed by CRAC, a technique that also relies on crosslinking.
The reason for these differences are not clear, but this often massive increase in RNAPII occupancy downstream of the pA site only observed in ParCLIP might have led to a downstream shift in the estimate of the site of termination.

Therefore it is unclear which dataset is more appropriate for defining the site of termination without additional experiments.

The decrease in RNAPII occupancy in the termination region observed by NET-Seq and CRAC might be the results of several components, including speeding of the polymerase in this region and cryptic or normal termination at many sites around the main pA site. However, these data clearly indicate that using ParCLIP data to define the site of termination might not be accurate and call for additional experiments that fall beyond the scope of this study.

Therefore, we used a mutant that is clearly impaired in termination (*rna15-2*) to visualize the behaviour of the RNAPII signal in the presence of a real termination defect. Regardless of its possibly multiple components, the RNAPII CRAC signal in the CPF termination mutant turned out to be markedly different from the signal observed in the absence of Rap1, indicating that the absence of the roadblock does not provoke termination defects at CPF sites.

We did our best to synthetically clarify this point in the revised version of the article, and this answer to referees might also be used for additional details.

Referee #3:

This paper provides a nice follow up to the earlier paper by Colin et al. that demonstrated the ability of Reb1 to block Pol II elongation both in vivo and in vitro. In this paper Candelli et al provide evidence that Rap1 and several other transcription factors can act as roadblocks (RBs) to Pol II elongation. The use of a UV cross-linking approach potentially expands this analysis genome wide.

My main concerns with this manuscript are the experimental design, the generality of the RB mechanism and the lack of high-resolution data showing precisely where Pol II pauses or terminates.

© European Molecular Biology Organization

1. My first major concern has to do with the experimental design. The authors are using UV irradiation to cross-link proteins to nucleic acids. The CRAC procedure isolates RNA that is cross-linked to RNA polymerase thus mapping the position of the polymerase at the time of formation of the crosslink. Irradiation with 254 nm UV will also cross-link proteins to DNA and this may confound the interpretation of these experiments. If DNA-binding proteins like Rap1 are efficiently cross-linked then they may provide an artificial barrier to Pol II elongation. During the 50 second period of irradiation Pol II could elongate several thousand bases. If Rap1 is more efficiently cross-linked compared to Pol II then Pol II may accumulate at the cross-linked Rap1 sites. The results shown in Figure 1 show that there is a Rap1 dependent termination or pause in the absence of UV but is this the case genome wide? The authors should provide alternative data from NET-seq or PAR-CLIP results showing peaks of Pol II adjacent to downstream Rap1 binding sites in the absence of UV.

We understand the concern of the referee but it would be quite unlikely that revealing a RB event would systematically (or almost) require at least two crosslinking events, one on the polymerase and one on Rap1. We know that crosslinking to the polymerase cannot be very efficient (at least in these conditions), because several crosslinking events on the same molecule might prevent efficient cDNA synthesis by the reverse transcriptase. We have indeed optimized the time of crosslinking (50") to observe a large majority of single crosslinking events as detected by deletions/mutation analyses. Still, it remains conceivable that Rap1 crosslinks to the DNA with higher efficiency than RNAPII. To comply with the referee's suggestions we performed the same metasite analyses with data from NetSeq (that does not involve crosslinking) and PAR-CLIP that only involves RNA crosslinking, and in both cases we saw evidence for RB.
As requested by the referee we have now been included in appendix Fig S2 a metagene analysis with NET-Seq data and a snapshot of the roadblock at the *HYP2* site (Fig 2B) in which the CRAC signal is presented in parallel with PAR-CLIP and NET-Seq data. PAR-CLIP data where already included in the paper for the metagene analyses of tRNA, centromeres and many transcription factors.

To further support the occurrence of termination, we have also provided metasite analyses showing the aggregate distribution of RNA 3'-ends in wt and *rrp6Δ* cells around sites of Rap1, Reb1, tRNA and centromeres roadblocks (appendix Fig S2 C-D and EV5). The distribution of RNA 3' ends shows peaks that coincide with roadblock sites and are generally higher in *rrp6Δ* cells, consistent with the notion that at least some of these termination events generate unstable transcripts.

2. A second major concern is in the generality of the RB phenomenon. The authors show several metasite analyses of Pol II occupancy aligned with Rap1 (or other transcription factor) binding sites. In none of these plots are we told how many binding sites were included in the analysis. In Materials and Methods the authors state that the analysis was restricted to Rap1 sites within 300 bp of a poly(A) site. How many examples of this orientation were analyzed? The same concern applies to the centromere and tRNA analysis. The authors should include in the supplemental material a list of the sites used in each metasite analysis.

We added these data in the revised version, together with the coordinates of the sites used for all metasite analyses. Please note that the analysis of Rap1 sites downstream of CPF terminators (i.e. within 300bp of an mRNA pA site) was only done for supporting the notion of constitutive RT at CPF sites that are limited by the Rap1 RB.

In addition to knowing how many RBs were analyzed in the metasite analysis it would be

nice to know how many actually demonstrate the RB phenomenon. The authors explain that any values more than 5 SD above the mean were excluded. How many sites were excluded?

We apologize for not having been sufficiently clear on this point. We did not exclude any Rap1 sites based on the 5 SD threshold. As we wrote in the Material and Methods section, when using the mean to summarize the data at every position, we excluded any values above 5 SD of the mean to avoid the risk that rare but strong outliers would dominate in the mean values plotted. This renders our analysis more stringent, but does not result in excluding any Rap1 sites.

It would also be nice to know how many were below the mean. Do all Rap1 binding sites act as RBs? Are there some that do not? Does the ability to act as a RB correlate with the different types of Rap1-dependent promoters characterized by the Shore lab? Does the RB ability of Rap1 correlate with the occupancy determined by Rhee and Pugh? Answering these questions could go some distance toward supporting the generality of the RB phenomenon.

The questions asked by this referee converge on the general question of what are the elements that favour or disfavour the occurrence of a roadblock depending on orientation, genomic location or kind of Rap1 promoters. These are important questions that are extremely difficult to address. The levels of detectable roadblock (i.e. pausing peak of the polymerase) depend on three parameters that cannot be easily measured for every site. The first is the level of incoming transcription, which is generally very low because Rap1 sites are located in intergenic regions, often between two divergent genes, or downstream of genes the transcription of which is independently terminated. The second parameter is linked to the efficiency of roadblock, which is expected to depend on the measured Rap1 occupancy but could also be influenced by many other factors *in vivo*, including the different topology of Rap1-dependent promoters, the competition by nucleosomes or other TF, etc. The third factor is the efficiency of clearance of the roadblocked polymerase by the Rsp5-dependent pathway, the kinetics of which – and its local specificities – are poorly understood. In other terms, not seeing a roadblock in a given position might be due to an inefficient roadblock by Rap1 (or any other RB factors), but also to poor incoming transcription, to efficient clearance of the roadblocked polymerase, or a combination of these factors. These uncertainties prevent a reliable evaluation of the extent of roadblock at a given position and the consequential reliable assessment of correlations.
Nonetheless, we have shown that the insertion of a core Rap1 site in the middle of a highly transcribed sequence is sufficient to induce a potent roadblock in different sequence contexts, in some cases selected for termination (i.e. in all the artificial Rap1-dependent terminators isolated) but also in regions of elongation without any selective pressure for termination (i.e. the insertion of the sole Rap1 site in the *HSP104* coding background). Interestingly, such a strong transcription context requires that the Rap1 site be inserted in the C-rich orientation, because its reverse complement would not be able to induce a roadblock. This strong directionality observed in this context is lost in a genomic context in which the levels of transcription are much lower. To illustrate these points we have mapped the sites of termination for 6 selected clones by RNaseH cleavage followed by high-resolution northern blot (Fig EV1A, see below). The sequences of the selected region of all the clones is also indicated.
The fact that in different heterologous context the sole Rap1 site can induce termination *in vivo* strongly supports the notion that Rap1 can induce a robust roadblock even under conditions of strong transcription as the ones determined by the Tet promoter.

3. A third problem with the paper is that the authors do not show any high-resolution data demonstrating a Rap1 or other RB. In figure 1 we are shown a Northern of a small RNA "of a size compatible with termination occurring immediately upstream of the Rap1 site". Without size markers on the blot and an indication of the distance between the TSS and Rap1 site it is difficult to judge the accuracy of this claim. Have the authors performed 3'

RACE or RNA protection assays to more precisely map the 3'-ends of this short RNA? Including this data would offer stronger support for the specificity of the RB mechanism.

We disagree with the referee. The technique we used to study roadblock termination allows—as NET-Seq, PAR-CLIP, and GRO-seq—to detect position of the RNAPII at near-single nucleotide resolution. However, we agree with the referee that we failed to provide sufficient details about the precise position of the site of pausing and termination. We have corrected this in the revised version of the manuscript as detailed in the following answers.
Concerning figure 1, to comply with the referee's request we have mapped at high resolution the 3' ends of the RNAs produced by six selected clones using a combination of RNaseH cleavage and PAGE (Fig EV1A). This experiment indicates that all RNA fragments produced terminate at most a few nucleotides upstream of the RB site.

In figure 2 the Pol II occupancy maps are at too low a resolution to see precisely where transcription is blocked. A nucleotide level map would be helpful in this regard. The reason for this concern is based on the metasite analysis shown in figure 3A (left panel). In this figure it looks like the peak of Pol II extends beyond the upstream edge of the Rap1 binding site. Does this argue that some RBs are not steric in nature? This same concern applies to the downstream edge of the centromere where the metasite peak of Pol II extends into the CBF3 binding site Figure 7A). In the individual CEN14 map Pol II pauses well upstream of the CBF3 binding site. Is CEN14 an outlier?

We have included in the legend of figure 2 the approximate position of the RNAPII peak relative to the first nucleotide of the first Rap1 binding site downstream of the peak. The approximation is linked to stochastic variability in the position of RNAPII during pausing and backtracking.
Concerning the metasite analysis in figure 3, the reason why part of the RNAPII peak bleeds over the average Rap1 site is also due to the uncertainty with which the latter is defined. Indeed, we used the Rap1 occupancy sites determined by ChIP-Seq by the Shore laboratory and used the coordinates of the maximum of each peak to align the RNAPII data. The position of the maximum is of course subject to stochastic variation, which cannot be easily corrected using the sequence of the site because the latter is not as clearly defined as a Reb1 site. Because of this uncertainty in the alignment, the metagene peak partially bleeds on the Rap1 metasite. That RNAPII pausing and termination occurs before Rap1 binding is however clearly shown in the experiments shown in figure EV1 and in the snapshots presented in figure 2.

Finally, concerning the downstream edge of the centromere (according to our alignment), it is true that the average RNAPII peak is located within CDEIII, but this likely reflects heterogeneity at different centromeres, most likely because termination is not always induced by CBF3 and in some instances occurs at the border of CDEII. Consistent with this notion, we observed that the RNA 3' ends distribute in two distinct peaks, one upstream of CDEIII and another within this region, both peaks essentially representing unstable RNAs that are degraded by the exosome. These data are now presented in Fig EV5.

Other comments.

4. In Figure 1A how many Rap1 terminators were used to make the logo? Do all of these synthetic Rap1 sites act as RBs?

All of the selected terminators containing Rap1 sites do terminate transcription, which is the basis for our selection. We have now precisely mapped the 3' ends of the RNAs produced by 6 of these clones (Fig EV1).

5. The Pol II occupancy maps have no quantification on the Y axis. This information is necessary for the reader to compare the various tracks. Is the readthrough after anchor away at the same scale as the no rapamycin track? The authors should add units to the maps or state in the legend that all maps are at the same scale.

We apologize, the scales are obviously the same, this should have been clearly stated; it is in the revised version.

6. In the text on page 9 I believe the authors have referred incorrectly to figures EV3C and EV3B.

This was corrected

7. In Figure 7 it is not clear which data is PAR-CLIP and which is CRAC. Panel 7A is labeled Pol II occupancy but the data is not specified. The origin of the data in 7C is also unclear.

We used PAR-CLIP data for the metagene analyses and CRAC data for the individual snapshots. This is now stated in the legend.

This paper is potentially very important if the authors can strengthen their claim that the RB phenomenon occurs genome wide. Addressing the concerns outlined here will make this paper more convincing. Detailed maps of the RBs described in this paper will provide an important resource for researchers interested in probing the roles of ncRNAs.

We thank the referee for appreciating the interest of our work.

2nd Editorial Decision 9 November 2017

Thank you for submitting a revised version of your manuscript to The EMBO Journal. It has now been seen by two of the original referees whose comments are shown below. As you will see they both find that all criticisms have been sufficiently addressed and they recommend the manuscript for publication. However, before we can go on to officially accept the manuscript there are a few editorial issues concerning text and figures that I need you to address in a final revision:

-> Please reformat the reference list to show 20 authors + et al rather than the current 10 authors + et al in accordance with the journal guidelines.

-> Please make sure the callouts in the text fit with the nomenclature in the figures/tables; we noticed the following issues:
- Callouts on p.22 mention Table EV1 and EV2, these should presumably be Appendix Table S1 and Appendix S2? (there are no EV tables).
- A callout is missing for Dataset EV1 (mandatory).
- Appendix table S2 has been called out but needs to be updated from 'Supplemental table 2' -> Appendix table S2 on p.22 and p.24.
- Callouts missing for Appendix Table S1 and S3 (mandatory).

-> We generally encourage the publication of source data, particularly for electrophoretic gels and blots, with the aim of making primary data more accessible and transparent to the reader. We would need 1 file per figure (which can be a composite of source data from several panels) in jpg, gif or PDF format, uploaded as "Source data files". The gels should be labelled with the appropriate figure/panel number, and should have molecular weight markers; further annotation would clearly be useful but is not essential. These files will be published online with the article as a supplementary "Source Data". Please let me know if you have any questions about this policy.

-> Papers published in The EMBO Journal include a 'Synopsis' to further enhance discoverability. Synopses are displayed on the html version of the paper and are freely accessible to all readers. The synopsis includes a short standfirst - written by the handling editor - as well as 2-5 one sentence bullet points that summarise the paper and are provided by the authors. I would therefore ask you to include your suggestions for bullet points.

-> In addition, I would encourage you to provide an image for the synopsis. This image should provide a rapid overview of the question addressed in the study but still needs to be kept fairly modest since the image size cannot exceed 550x400 pixels.

Thank you again for giving us the chance to consider your manuscript for The EMBO Journal, I look forward to receiving your final revision.

REFEREE REPORTS

Referee #1:

I supported publication of the original MS. The responses to my comments are fairly minimal but the MS has been improved overall and I am happy to recommend publication.

Referee #3:

In this revised manuscript Candelli et al. have addressed the points of concern in the previous submission. In particular, the metagene analysis shown in the Appendix Figure S2 allays concern that the RB phenomenon was UV-dependent. Parts A and B of this figure should be included in supplemental material. I now feel that the manuscript makes a strong contribution to our understanding of termination mechanisms and will serve as a resource for further investigations.

2nd Revision - authors' response 14 December 2017

The authors made the requested changes and submitted the final version of their manuscript.

© European Molecular Biology Organization

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: Libri
Journal Submitted to: EMBO Journal
Manuscript Number: EMBOJ-2017-97490R

USEFUL LINKS FOR COMPLETING THIS FORM

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

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 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.

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)

Question	Answer
1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size?	Sample sizes have been dictated by availability of data fulfilling a given criteria. This is explicitated in the manuscript for each experiment.
1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.	NA
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	Inclusion/exclusion criteria have been explicitated in the manuscript for each experiment
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.	NA
For animal studies, include a statement about randomization even if no randomization was used.	NA
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.	NA
4.b. For animal studies, include a statement about blinding even if no blinding was done	NA
5. For every figure, are statistical tests justified as appropriate?	usage of the paired t-test in figure 4 was justified in the methods section.
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	the data to which the paired t-test was applied are approximately normally distributed. This was assessed through inspection of a quantile-quantile plot (qqplot).
Is there an estimate of variation within each group of data?	the distributions in fig 4 are represented by boxplots that show data variation the form of inter-quartile ranges.

Is the variance similar between the groups that are being statistically compared?	similar variance was detected in all groups compared

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).	NA
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.	NA

* 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.	NA
9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.	NA
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.	NA

E- Human Subjects

11. Identify the committee(s) approving the study protocol.	NA
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.	NA
13. For publication of patient photos, include a statement confirming that consent to publish was obtained.	NA
14. Report any restrictions on the availability (and/or on the use) of human data or samples.	NA
15. Report the clinical trial registration number (at ClinicalTrials.gov or equivalent), where applicable.	NA
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.	NA
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.	NA

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	All data generated have been deposited and a GEO access number has been provided.
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).	NA
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).	NA
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.	NA

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.	NA