Thank you again for your patience regarding your pending manuscript submission with us. I have now had the chance to go through your manuscript file and the referee comments in light of the additional input from our expert advisor - and the outcome is that this is not a straightforward decision. All three reports are included below and as you will see both refs and advisor find your manuscript interesting in principle and praise the first part of the work addressing dynamic CTCF recruitment on hmC sites in CD45 (i.e. figures 1-6).

However, at the same time they do raise rather serious concerns about the genome-wide binding data - and the conclusions based on this - that had been added in the new manuscript before submission. As you will see, this relates both to the presentation and discussion of the data - as well as to the difficulties arising from not having direct CTCF ChIP seq data in T cells.

Since the inclusion of the genome-wide data was an initial criteria from our side to expand the generality of the regulatory process beyond CD45, I would like to consult with you on the concerns raised by the referees before I go on to make an official decision for this manuscript. In addition, I would be curious to know how much data you would be able to provide in response to the request from ref #1 for functional characterization of additional targets (as a possible alternative to expand generality). My suggestion would therefore be that you provide an outline of the experiments/reanalysis that could be included to address the concerns of the referees (and possibly allow the inclusion of the genome-wide data).

I would like to once more apologize for the duration of the process in this case, but I do hope that a
discussion of the reports at this stage will help us to come to the right decision for your study.

I look forward to hearing from you.

Author response 15 February 2015

Thank you for the reviews. I am very glad to see that all three reviewers found the CD45 work to be of high interest, creative and well designed. I am also relieved to see that two of the reviewers felt that the CD45 work stood on its own merits and is of sufficient caliber for publication in EMBO J. I felt the same way when I submitted the initial manuscript in late August.

I believe that some confusion in the review of this manuscript derives from a distinction between the chromatin field and the “alternative” splicing field. By definition, the mechanisms that promote alternative splicing are highly selective and alter a discrete number of events in the transcriptome. As 5hmC and 5mC are pervasive in the transcribed genome, our goal is to systematically dissect discrete regulatory factors that may rely on differential methylation to promote or inhibit splicing. To accomplish this, a reliable model gene, such as CD45, is key. The model gene approach is widely accepted in the splicing field, as is evident through the reviews. Otherwise, the signals that promote constitutive splicing tend to occlude the specific signals that contribute to alternative splicing.

With that said, I was actually quite impressed to see that we were able to identify a relationship between differential 5hmC and alternative splicing at CTCF binding sites in the genome-wide data. I understand your Editorial decision as well as Reviewer 1’s interest in this data as a comforting “proof of principle.” With an eye toward satisfying both your desire to include genome-wide data for publication in EMBO and the two reviewers who believe it should be excluded, I suggest combining the genome-wide analysis into one succinct Figure focused on just CTCF sites. This will further allow for some explanation of aspects of the analysis that were not clear in this version. For example, the CTCF ChIP-seq shown in Figure 7E are from our CD4+ T cells. I briefly describe potential adjustments to the manuscript and offer my responses to reviewer comments below.

Reviewer 1:
Figure 4: Some data in the literature claims that CTCF recruits TET proteins (see for example PMID 25183525). Panel B seems to combine cell lines having integrated WT or Mut CTCF binding sites. Segregating the WTs and the Muts may provide information on the impact of a WT CTCF binding site on the conversion from 5mC to 5hmC (one would expect that if CTCF recruits TETs, Muts would be less efficiently converted than WTs).

We had the same thought and examined this prior to submission. However, we did not see a clear connection between the CTCF binding site and 5hmC. While a minor increase in 5hmC was detected at the WT minigenes, it is not statistically significant. We can include this in the supplemental in a revised manuscript.

Figures 7-9: the model should be validated on a series of genes identified in the genome-wide studies.

We can identify candidates through the genome-wide analysis and include this as validation in the newly CTCF focused figure.

Reviewer 2:
Comments were technical and addressable.

Reviewer 3:
1) Figure 7D, shows encouraging results but the CTCF sites that the authors have selected due to the lack of an efficient CTCF ChIP-Seq cannot directly be related to the system being studied, i.e. activation of CD4+ T-cells.

See response to below.

2. Figure 7E, the authors mention CTCF sites based on ChIP-Seq. This is confusing and makes one wonder why do they not use those sites for the analyses resulting in figure 7D?
The data in 7D were from ENCODE defined sites, whereas the data in 7E are from our empirically
determined CTCF ChIP-seq data in naïve and activated CD4+ T cells. While we were unable to
determine differentially bound sites between these two highly distinct cell types, we were
successfully able to call peaks in the CD4+ T cell data. These were used in the Fig. 7E analysis. I am
very sorry that this wasn’t clear.

To clarify further, determining differential ChIP is a major issue in the chromatin field and
programs designed for this sort of analysis are lacking. In our case, this was additionally
complicated by the fact that naïve and activated T cells are very distinct at the chromatin level and
yield distinct levels of IgG and input reads. As a result, we did not feel confident in calling
differential sites, and instead focused on locations of CTCF binding. We were very happy to see (7E
top) that our alternatively spliced exons were enriched for downstream CTCF/5hmC. In focusing the
genome-wide analysis on CTCF/5hmC sites and providing additional examples, I believe we can
strengthen and clarify this relationship.

3. The overlap between hydroxymethylated sites, CTCF binding sites and differentially spliced
exons is quite low. The authors tend to overstate the generality of their model.
-As mentioned in the introduction, I agree with this point and think this is one of the drawbacks of
examining “alternative” splicing in a field trending towards default genome-wide associations.

4. The authors state that an increase in hmC associates with an increased inclusion of exons near
CTCF binding sites but as mentioned by referee #2, figure 8C shows the opposite trend genome-
wide and the authors fail to address this in their discussion.
-Figure 7 focuses on CTCF sites, whereas figure 8 examines all 5hmC sites in distinct exon
categories. The rationale behind this analysis derives from our belief (and that of others in the field)
that CTCF is not the only factor regulating methyl-sensitive alternative splicing. For example, the
laboratory of Keji Zhao showed that MECP2 binds to methylated intragenic DNA to promote
inclusion. At those sites, the presence of hmC could reflect exclusion, consistent with the elevated
5hmC that we observe in excluded exons. Nonetheless, this panel could be removed from the
condensed genome-wide figure.

5. The cleanest results emanating from the genome-wide data is the influence of CpG density on the
direction of hmC variations. Although noteworthy, this is to me not essential to this story nor does it
directly support any of the earlier claims. It is, at this stage at least, an interesting but preliminary
and independent observation.
-This would be removed in a condensed genome-wide figure.

To summarize, we propose to condense the genome-wide data into one figure focused on CTCF. We
will further show additional examples of CTCF regulated exons that are found through the genome-
wide data. In addition, we will make it clear that the data that are included in the current figure 7E
are indeed CTCF ChIP-seq from our samples.

1st Editorial Decision 16 February 2015

Thank you for sending a response to the comments of our referees and advisor - as well as an outline
of what additional data and clarifications could be included in a revised manuscript.

Based on this response - and in light of the largely positive assessment from our referees - I would
invite you to submit a revised version of the manuscript, addressing the comments of all three
referees. In line with your outlined response, I would particularly ask you to focus your efforts on
the following points:

-> Please expand the study with functional characterization of additional target loci as requested by
ref #1

-> Please clarify the nature and origin of the genome-wide ChIP-seq data for CTCF, as well as the
consistency with the mechanism delineated for the isolated CD45 locus (ref #2 and #3)
Please re-organise/refocus the genome-wide data to strengthen the conclusions for CTCF as a regulator of alternative splicing.

I would like to emphasize that these points - as well as all minor concerns raised - will have to be addressed to the full satisfaction of the referees in the revised manuscript. It is EMBO Journal policy to allow only a single round of revision, and acceptance or rejection of your manuscript will therefore depend on the completeness of your responses in this revised version.

When preparing your letter of response to the referees' comments, please bear in mind that this will form part of the Review Process File, and will therefore be available online to the community. For more details on our Transparent Editorial Process, please visit our website: http://emboj.embopress.org/about#Transparent_Process

We generally allow three months as standard revision time. As a matter of policy, competing manuscripts published during this period will not negatively impact on our assessment of the conceptual advance presented by your study. However, we request that you contact the editor as soon as possible upon publication of any related work, to discuss how to proceed. Should you foresee a problem in meeting this three-month deadline, please let us know in advance and we may be able to grant an extension.

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

REFEREE REPORTS

Referee #1:

In an earlier study, this lab had shown that binding of CTCF to CD45 exon 5 favors the inclusion of this alternative exon in the mature CD45 transcript. As CTCF binding is sensitive to DNA methylation, this study provided evidence for an impact of 5mC on alternative splicing. In the present manuscript, the authors show that inclusion of the CD45 exon is observed not only in cell lines showing no methylation overlapping the CTCF binding site of exon 5, but also in cell lines where this binding site is hydroxymethylated (5hmC). Consistent with this, they show that activation of human CD4+ T cells results in decreased accumulation of TET1, 2, and 3, decreased levels of 5hmC on exon 5, and decreased recruitment of CTCF to that exon. Likewise, depletion of either TET1 or TET2 with siRNAs in a BL41-derived cell line efficiently including exon 5, caused reduced inclusion of exon 5 and reduced levels of 5hmC on the DNA encoding exon 5. CHO-derived cells having stably integrated in vitro-methylated reporter constructs either WT or with a mutant CTCF-binding site in Exon 5 further show that CTCF-binding rather than 5hmC is important for exon 5-inclusion and Pol II accumulation. In vitro experiments allow to further show that CTCF binding to DNA is not affected by hydroxymethylation. Finally, a genome-wide transcriptome analysis comparing naïve and activated CD4+ T cells suggests that 5hmC is enriched at CTCF-regulated alternatively spliced exons.

Enrichment of 5hmC at CTCF binding sites and the anti-correlation between 5hmC and 5mC at these sites has been documented in several model systems (see for example, pmid 23352666). Thus, an effect of 5hmC and TET proteins on alternative splicing of CTCF-binding exons is, to some extent, an expected result. Yet, the paper is very thorough and technically creative, and especially the last part of the manuscript decoding the mechanisms linking levels of 5hmC to exon usage will most likely be of great interest for the field. I recommend strengthening this section by validating the model on some of the genes identified in the genome-wide study.

Specific comments
- Figure 4: Some data in the literature claims that CTCF recruits TET proteins (see for example PMID 25183525). Panel B seems to combine cell lines having integrated WT or Mut CTCF binding sites. Segregating the WTs and the Muts may provide information on the impact of a WT CTCF binding site on the conversion from 5mC to 5hmC (one would expect that if CTCF recruits TETs, Muts would be less efficiently converted than WTs).
- Figure 6E requires a control showing that in the conditions used, DNA methylation interferes with
CTCF binding.
- Figures 7-9: the model should be validated on a series of genes identified in the genome-wide studies.

Minor comment
Some of the studies linking CTCF sites to 5hmC enrichment should be cited.

Referee #2:

The paper by Bailly et al addresses the interesting issue of how DNA methylation within genes affects their expression by regulating alternative splicing. Important previous work by the Oberdoerffer lab showed that binding of the methyl sensitive DNA-binding factor CTCF enhances inclusion of the alternative exon 5 of CD45 and a suggested mechanism was through induction of RNA polymerase pausing. This paper provides interesting new evidence that down-regulated oxidation of 5meC to 5hmC in activated T lymphocytes, where TET1 and TET2 expression is low, enhances skipping of exon 5 by inhibiting CTCF binding. The correlative evidence in support of this model provided in Figure 2 is strong and is supported by the direct demonstration that CTCF can bind to 5hmC modified binding site (Fig. 6D, E).

The paper become rather muddled when the authors try to argue that the model they have developed for CD45 exon 5 applies generally to control of alternative splicing in activated T cells. If the CD45 exon 5 model has relevance generally then the clear prediction is that alternative exons (n=633, Fig. 7C) that are more excluded in activated T cells will have reduced 5hmC and increased 5meC like CD45 exon 5. Conversely it is predicted that exons which are more included (n=462) will have increased 5hmC and greater CTCF binding. I could not find a direct test of these predictions in the data presented but the results in Figure 8A do not conform with the model and those in Fig. 8C seem to contradict it. This important discrepancy is not directly addressed in the discussion and how the model can be reconciled with the genome-wide data is left quite obscure. As a result, the paper does not have a coherent take home message. My suggestion is to present the CD45 data on its own merits without attempting to argue that it applies generally to regulated alternative splicing during T cell activation.

Specific points:
1. The CTCF binding experiments in Fig. 6E would be strengthened significantly by including a negative control probe that is methylated on CpG's.
2. The flow of the paper would be improved by showing the CTCF binding to 5hmC modified DNA much earlier, say after Fig. 2 rather than keeping the reader guessing about this critical issue.
3. It was unclear how levels of 5hmC were quantified over exon 5 in the WT- and Mut-hmC cell lines in Fig. 4B. This is important because the authors argue (p. 12 end of para 1) that CTCF binding not 5hmC per se is responsible for the differences in exon 5 inclusion between these clones (Fig. 4C).
4. Figure 4G which argues for pol II pausing over exon 5 in the WT cells relies on a single PCR amplicon in this region and would be strengthened by adding some additional amplicons.
5. The arguments (p. 19) for "bookmarking" by 5hmC and that 5hmC is "susceptible to modulation" (how?) were vague and failed to clarify the overall picture for me.

Referee #3/Arbitrating advisor:

The authors reported previously that the binding of CTCF in the 5th exon of the CD45 gene favors exon inclusion in the mature transcript. This CTCF binding site contains CpG nucleotide which are sensitive to DNA methylation (mC). In this earlier study, the authors provided evidence for an impact of mC on alternative splicing. Indeed, methylation of CpGs in the CTCF binding site impairs CTCF binding. The first part of the present work by Bailly et al. (i.e; first 14 pages and first 6 figures) shows that inclusion of the CD45 exon is observed not only in cell lines harboring no methylation overlapping CTCF binding site of exon 5, but also in cell lines in which this binding site is hydroxymethylated (hmC). Consistent with this observation, they show that activation of human CD4+ T cells results in decreased accumulation of TET enzymes, reduced levels of hmC on exon 5, and decreased recruitment of CTCF. These data support the idea that TET enzymes can influence exon inclusion rate via oxidation of mC to hmC. The authors pursue their investigation by
showing that reduced inclusion of exon 5 associates with decreased levels of 5hmC. They further show that CTCF-binding, rather than hmC, is important for exon 5-inclusion.

The first part of the article, focusing on the CD45 gene system, is indeed quite interesting, well designed and the data support the author's conclusions. Hence, I find this part of the manuscript suitable for EMBO J., providing that referees' comments are fully addressed.

Concerning the contradictory opinions of the 2 referees, I would agree with referee 2 on his/her suggestion not to include the second part regarding genome-wide data and this, for the following reasons:

1. Figure 7D, shows encouraging results but the CTCF sites that the authors have selected due to the lack of an efficient CTCF ChIP-Seq cannot directly be related to the system being studied, i.e. activation of CD4+ T-cells.
2. Figure 7E, the authors mention CTCF sites based on ChIP-Seq. This is confusing and makes one wonder why do they not use those sites for the analyses resulting in figure 7D?
3. The overlap between hydroxymethylated sites, CTCF binding sites and differentially spliced exons is quite low. The authors tend to overstate the generality of their model.
4. The authors state that an increase in hmC associates with an increased inclusion of exons near CTCF binding sites but as mentioned by referee #2, figure 8C shows the opposite trend genome-wide and the authors fail to address this in their discussion.
5. The cleanest results emanating from the genome-wide data is the influence of CpG density on the direction of hmC variations. Although noteworthy, this is to me not essential to this story nor does it directly support any of the earlier claims. It is, at this stage at least, an interesting but preliminary and independent observation.

1st Revision - authors' response 24 May 2015

Response to Referee #1

We would like to thank the reviewer for the detailed response. We are glad that the reviewer found our study to be “thorough and technically creative.” The Reviewer’s main concerns related to an extension of citations, additional controls and validation of the genome-wide data. We have addressed all these concerns, as outlined in the below point-by-point response.

In response to the specific comments:

- Figure 4: Some data in the literature claims that CTCF recruits TET proteins (see for example PMID 25183525). Panel B seems to combine cell lines having integrated WT or Mut CTCF binding sites. Segregating the WTs and the Muts may provide information on the impact of a WT CTCF binding site on the conversion from 5mC to 5hmC (one would expect that if CTCF recruits TETs, Muts would be less efficiently converted than WTs).

-We had the same thought and examined this possibility prior to initial submission. However, we did not see a clear connection between a competent CTCF binding site and 5hmC levels. While a minor increase in 5hmC is detected at the wildtype minigenes, the result is not statistically significant. This is included as Supplementary Fig S5A in this revised manuscript.

- Figure 6E requires a control showing that in the conditions used, DNA methylation interferes with CTCF binding.

This is an important point that we agree is a necessary demonstration for our model (i.e. that CTCF interacts with 5hmC but not 5mC-containing CD45 DNA). Unfortunately, despite extensive efforts, we are unable to find clear evidence that CTCF is able to discriminate between 5hmC and 5mC in the context of CD45 exon 5 in vitro. By all measures, we find equivalent interaction between CTCF and unmethylated or CpG methylated CD45 oligos. Our goal for including the 5hmC EMSA in the initial submission was to merely state that CTCF “can” interact with 5hmC-containing DNA in the context of CD45 exon 5. Nonetheless, noting that this is an important point, we have expanded our efforts to address this issue during the revision period, as detailed below.
As shown in the new **Fig 6E**, we performed EMSA with purified CTCF and fully CpG methylated exon 5 duplexes and determined clear and specific binding to the 5mC containing probes. Methylation was validated through reduced digestion with PvuRts1l (5hmC preferred) and BsaHI (5hmC/5mC inhibited) (new **Fig 6D**). We further confirmed through Western blot that CTCF/5mC interaction in EMSA was not attributed to copurification of the 5mC binding CTCF paralog BORIS/CTCFL (new Supplementary **Fig S6F**). Note that interaction of CTCF with the 5mC oligos is not due to binding at a secondary site, as mutation of the CTCF binding site fully ablated binding (**Fig 6C**). In addition, these results are not an artifact of the extended probe size, as the shortest probe that we are able to see robust binding to unmethylated DNA shows a similar pattern for methylated DNA (new Supplementary **Fig S6E**). In addition, CTCF/5mC interaction cannot be attributed to an artifact of probe synthesis as *in vitro* methylation with the M.Sssl CpG methyltransferase also results in a probe that is able to interact with CTCF (new Supplementary **Fig S6D**).

Acknowledging the extensive body of literature that shows CTCF interaction with DNA is inhibited by overlapping 5mC, including our own previous publication, we performed additional analysis into this issue. We explored several possibilities to reconcile our *in vitro* results from the *in vivo* observations and in each case were unable to explain the discrepancy, as outlined below:

1) **Differential strength of interaction?**

We reasoned that perhaps CTCF interaction with 5mC-containing CD45 DNA is less stable than with unmethylated DNA. To test this, we performed titration with cold unmethylated competitor in EMSA. The resulting autoradiograph showed comparable binding of CTCF to unmethylated or methylated exon 5 DNA (**Reviewer Fig 1**).

![CD45 Exon 5 EMSA Titration](Image)

**Reviewer Fig 1.** EMSA with purified CTCF and cold competition with increasing amounts of unlabeled wildtype probe.

2) **Competition with other binding factors *in vivo***?

We reasoned that perhaps another factor, such as the methyl-binding paralog CTCFL, occupies the CTCF binding site when methylated *in vivo*. To test this, we incubated biotinylated exon 5 duplexes with BL-E5(+) cell nuclear lysates. Following streptavidin affinity purification, CTCF Western blot was performed to assess the extent of copurification. The resulting immunoblot showed efficient interaction of CTCF with unmethylated, 5hmC and 5mC-containing CD45 duplexes in cellular lysates (new Supplementary **Fig S6C**).

3) **Altered nucleosome positioning?**

As genome-wide studies of CTCF binding have shown that CTCF interacts with nucleosome free DNA, we reasoned that perhaps 5hmC and 5mC containing DNA differ in nucleosome assembly in a manner that could impact CTCF interaction *in vivo*. We thus performed MNase profiling of CD45 DNA in isogenic conditions of reciprocal 5hmC and 5mC (BL41 sub-populations and CD4+ T cells, shown for BL41 here). Primer efficiencies were calculated through standard curves to allow for direct comparison of adjacent regions. The resulting analysis revealed no distinction between MNase digestion patterns in the 5hmC and 5mC contexts and further suggest that CTCF is able to interact with nucleosomal DNA when hydroxymethylated (**Reviewer Fig 2**).
4) Differential interaction with RNA?

Interaction of CTCF with nucleosomal DNA led us to question whether CTCF is recruited to 5hmC sites through its ability to interact with RNA. To address this possibility, we examined CTCF interaction with CD45 DNA in the presence or absence of pol II transcription inhibition through DRB treatment. We found no difference in CTCF binding to CD45 DNA in hydroxymethylated BL-E5(+) cells, as assessed through CTCF ChIP and qPCR, thereby demonstrating that interaction with 5hmC sites is unlikely to be via indirect interaction with nascent RNA (new Fig 6F).

From the sum of these and additional data that are not shown here for the sake of brevity, we can only conclude that CTCF is technically able to interact with 5hmC and 5mC-containing CD45 DNA, but that other determinants disfavor 5mC binding in vivo. Notably, the CD45 CTCF binding site contains CpGs at the locations that were linked to methylation-associated differential occupancy in a genome-wide survey (representing less than 30% of CTCF binding sites) (Wang et al, 2012). Conversely, several other reports, including a recent quantitative mass spectrometry analysis, showed equivalent CTCF preference for unmethylated versus 5mC-containing probes (Spruijt et al, 2013). While outside the scope of the current manuscript, we noted two observations that support context dependency in our EMSA studies:

1) When we generate PCR products in the presence of 5hmC or 5mC NTPs such that all cytidines are modified (not just CpGs), CTCF binding to both 5mC and 5hmC-containing DNA was fully ablated (Reviewer Fig 3), revealing currently unknown structural requirements or quantitative burdens in CTCF methyl-sensitivity.
2) We considered whether CTCF interaction with 5mC-containing DNA is influenced by post-translational modifications (PTMs). Our rationale derived from the fact that CTCF and the paralog CTCFL have near identical zincfingers and share many common proteinprotein interactions, but CTCFL preferentially binds at methylated CTCF binding sites in vivo. Mass spectrometry for CTCF and CTCFL PTMs identified a number of locations of phosphorylation (not shown). We thus examined CTCF interaction with CD45 DNA following protein dephosphorylation. To our surprise, dephosphorylation ablated CTCF interaction with unmethylated CD45 DNA, revealing an unknown role for PTMs in CTCF binding preferences (Reviewer Fig 4).

While we realize that these results are not conclusive, and do not bring us closer to resolving the distinction between CTCF binding preferences for CD45 DNA in vitro versus in vivo, these data are shown to assure the reviewers that we have gone to great lengths to dissect the association of CTCF with 5hmC versus 5mC-containing CD45 DNA. We would further like to point out that we, and the
observational field in general, are unable to truly characterize CTCF binding to CD45 DNA in vivo. At best, we can say that CTCF is inhibited by 5mC-containing DNA (as evidenced through TET1/2 knockdown) but the source of this depletion is not definitive. Given the large body of evidence that supports CTCF interaction with 5hmC relative to 5mC-containing DNA in cellular conditions (at CD45 and the genome-wide results), we hope that the Reviewer will not dwell on this inconsistency. We maintain that the in vitro data demonstrate that CTCF can bind to 5hmC-containing DNA in the context of exon 5 and the in vivo data show that this is conducive to CTCF binding with a net impact on alternative splicing.

-Figures 7-9: the model should be validated on a series of genes identified in the genome-wide studies.

Thank you for the suggestion. We have included these data in the revised manuscript (new Supplementary Fig 8) and have focused the genome-wide analysis to specifically examine CTCF and 5hmC in naïve and activated CD4+ T cells.

Minor comment
Some of the studies linking CTCF sites to 5hmC enrichment should be cited.

We regret the omission and have included several citations in the revised manuscript.

Response to Referee #2

We would like to thank the reviewer for the generally positive review. We are glad that the reviewer found the CD45 data to stand on its own merit. The Reviewer’s main concern related to the presented genome-wide analysis. It was suggested that this data be removed from the manuscript altogether. We agree on the strength of the CD45 data and our initial submission to EMBOJ focused solely on this aspect of the manuscript. However, inclusion of genome-wide generalization of our model was an Editorial condition of review. Considering the Editorial policy and the fact that the genome-wide data was positively received by Referee #1, we have tried to satisfy all parties in this revised manuscript through a thorough reanalysis of the primary data from naïve and activated CD4+ T cells. Through highly restrictive filters, we identified a number of alternatively spliced exons with proximal differentially methylated CTCF binding sites in naïve versus activated cells. Importantly, alternative exons with downstream CTCF showed strong adherence to the CD45 model, wherein decreased 5hmC and increased 5mC at the CTCF binding site was associated with upstream exon exclusion (and vice versa). We further performed a thorough survey at two of the newly identified sites (CTCF ChIP, 5hmC MedIP, 5mC MedIP and RNA) with new donor samples and were able to fully validate the model in two additional genes (APOPT1 and HAPLN3, new Supplementary Fig 8). We hope that the Referee will find this data to be a valuable extension of the CD45 model.

In response to the specific comments:

1. The CTCF binding experiments in Fig. 6E would be strengthened significantly by including a negative control probe that is methylated on CpG’s.

We now show EMSA with commercially synthesized probes that are uniformly methylated at every CpG and perplexingly also find robust interaction between 5mC-containing CD45 DNA and CTCF. We have included this data in the revised manuscript and have performed a number of additional analyses to attempt to determine the source of the discrepancy between the in vitro data showing that CTCF is able to bind CD45 DNA in the context of CpG 5mC and the in vivo data showing clear depletion of CTCF at 5mC-containing sites. Please see the extensive response to Referee 1 above for a detailed description of these investigations. We regret that we were unable to come to a clearer conclusion on this point, but feel confident that our data show that CTCF is at least “able” to bind to 5hmC-containing DNA in the context of CD45 exon 5 sequence.

2. The flow of the paper would be improved by showing the CTCF binding to 5hmC modified DNA much earlier, say after Fig. 2 rather than keeping the reader guessing about this critical issue.
We thank the Reviewer for this valuable suggestion. However, given the scope of the newly presented data, we have chosen to leave the EMSA studies at the end of the CD45 analysis so as to not distract from the manuscript flow.

3. It was unclear how levels of 5hmC were quantified over exon 5 in the WT- and MuthmC cell lines in Fig. 4B. This is important because the authors argue (p. 12 end of para 1) that CTCF binding not 5hmC per se is responsible for the differences in exon 5 inclusion between these clones (Fig. 4C).

The method of calculation was detailed in the Materials and Methods of the original manuscript. We have kept the formula in the same location in the current version, but also reference the method in the figure legend of the main text.

4. Figure 4G which argues for pol II pausing over exon 5 in the WT cells relies on a single PCR amplicon in this region and would be strengthened by adding some additional amplicons.

Unfortunately, we used the entirety of the pol II ChIP material for the shown primer sets and are unable to return to the panel of clones due to methyl-drift. Consistent with a role for transcription in the recruitment of DNMT3B to gene bodies through interaction with H3K36me3 (Baubec et al, 2015), after several weeks in culture the wildtype unmethylated clone showed evidence of the development of de novo methylation within the CD45 minigene body, which was also evident after freeze-thaw of an early passage. As a result, additional data points would require us to go back to the very beginning of stable clone generation. As we were cognizant of the occurrence of methyl-drift in cell culture, all materials used for the figures were harvested in parallel to be certain of methyl-status at the time of analysis.

As our goal was to demonstrate that 5hmC is not directly responsible for pol II pausing at exon 5, we hope the Reviewer will agree that the distinction between pol II occupancy in the wildtype and mutant hydroxymethylated clones clearly makes this point. We should also mention that all ChIPs were performed in triplicate and our primer efficiencies were calculated against a standard curve and were uniformly determined to approach the maximum efficiency of 2, thereby reducing the possibility for artifacts (response to Referee 1, Reviewer Fig 2).

With that said, the Supplementary Figure contains information for an additional downstream primer set (exon 6, also present in the original version). Exon 6 was omitted from the main figure as we found that pol II phasing was lost downstream of the CTCF site, as would be predicted. In addition, we felt that exon 6 detracts from the main message of the figure. Judging by increased pol II at the 3’ end of the hydroxymethylated minigenes (the exon 6 primer set, WT and Mut) as compared to the unmethylated minigene, it is possible that 5hmC has a direct role on 3’ end pol II pausing. This is consistent with the described global accumulation of 5hmC at the 3’ ends of genes. While this is outside of the scope of the manuscript, we show it here for the sake of the Reviewers (Reviewer Fig 5).
5. The arguments (p. 19) for "bookmarking" by 5hmC and that 5hmC is "susceptible to modulation" (how?) were vague and failed to clarify the overall picture for me.

We apologize for the confusion and have removed this text from the revised version of the manuscript.

Response to Referee #3

We would like to thank the Referee for these valuable comments. We are pleased that the Reviewer found the CD45 data to be interesting and well-designed. Like Referee #2, this Reviewer felt that the CD45 data was sufficient for publication on its own merit, and that the genome-wide data should be omitted. As described in the response to Referee #2, the genome-wide analysis was an Editorial requirement for review and was favored by Referee #1. In an attempt to satisfy all parties, we have conducted a thorough reanalysis of the CD4+ T cell data in this revised version. We have also expanded on methodologies in the text to clear up confusions stemming from the original text (for example, the CTCF-ChIP-seq data was indeed derived from our primary CD4+ T cell samples). We hope that the Reviewer will find this revised version, including the genome-wide analysis, acceptable for publication in EMBOJ.

In response to the specific comments:

1. **Figure 7D**, shows encouraging results but the CTCF sites that the authors have selected due to the lack of an efficient CTCF ChIP-Seq cannot directly be related to the system being studied, i.e. activation of CD4+ T-cells.

Please see the response to point 2 below.

2. **Figure 7E**, the authors mention CTCF sites based on ChIP-Seq. This is confusing and makes one wonder why do they not use those sites for the analyses resulting in figure 7D?

The data in Fig 7E of the original manuscript were indeed from our empirically determined CTCF ChIP-seq data in naïve and activated CD4+ T cells. In an attempt to explain why we did not filter for differentially bound CTCF binding sites in naïve and activated cells, we inadvertently caused this confusion. We have remedied this in the revised manuscript by instead filtering for differential methylation at CTCF binding sites. Briefly, as shown in the new Supplementary Fig S7A-B, the majority of CTCF binding sites were detected in naïve and activated cells and were not differential. While there are clearly a number of sites that fall outside the overlap between naïve and activated cells, computational identification of these sites is complicated by the overall higher signal in the entire population of active site data (overall shift toward active axis). This is an artifact of more efficient ChIP in active cells due to more accessible chromatin. We thus did not feel confident in performing a differential CTCF filter and instead focused on examining 5hmC overlap at CTCF sites that were proximal to alternative versus constitutive exons in the original manuscript. In retrospect, this approach limited our ability to examine the net impact on alternative splicing, which was the ultimate goal.

In this revised manuscript, we instead focus on CTCF binding sites with clear evidence of reciprocal changes in 5hmC and 5mC in response to T cell activation. As MedIP involves immunoprecipitation of purified, histone-free DNA, it is not subject to artifacts related to altered chromatin accessibility. In this manner, we focused on a subset of CTCF binding sites with differential methylation and were able to determine the impact on proximal alternative splicing with high confidence. The resulting analysis found robust associations between methyl-sensitive CTCF locations downstream of alternative exons that conform to the established model.

3. The overlap between hydroxymethylated sites, CTCF binding sites and differentially spliced exons is quite low. The authors tend to overstate the generality of their model.

We fully agree with this point and think this is one of the drawbacks of examining “alternative” splicing in a field trending towards default genome-wide associations. By definition, the mechanisms that promote alternative splicing are highly selective and alter a discrete number of
events in the transcriptome. As 5hmC and 5mC are pervasive in the transcribed genome, our goal is to systematically dissect discrete regulatory factors that may rely on differential methylation to promote or inhibit splicing. To accomplish this, a reliable model gene, such as CD45, is key.

With that said, we are impressed that we were able to identify a relationship between differential 5hmC and alternative splicing at CTCF binding sites in the genome-wide data. While the number of events remaining after high-stringency filtering was small, the resulting associations at downstream CTCF binding sites were robust. Notably, our analysis pipeline certainly underestimates the association between CTCF and splicing, as we purposely omitted complex splicing patterns that could obscure the directional relationship to CTCF, such as CD45. CTCF binding sites are often found in locations with multiple alternative exons and many of these would be depleted from our analysis.

In addition, we have purposely filtered away genes that showed differential gene expression as well as other splicing patterns, such as mutually exclusive splicing and intron retention. In other words, in an attempt to apply high stringency to this analysis, we are quite certain that we have grossly underestimated CTCF dependent events.

4. The authors state that an increase in hmC associates with an increased inclusion of exons near CTCF binding sites but as mentioned by referee #2, figure 8C shows the opposite trend genome-wide and the authors fail to address this in their discussion.

The data in Figure 8 of the previous submission examined exonic 5hmC and 5mC levels independent of association to CTCF. As we show in the new Fig 7C, CTCF overlap at exons represents approximately 0.5% of all exons. Thus the contribution of CTCF dependent events would not have been visible at the all exon scale. These exonic data that are not related to CTCF have been removed from the revised manuscript to avoid additional confusion. Instead, we specifically and solely query the impact of differential 5hmC on CTCF dependent splicing. Briefly, we find that while the number of events is limited, 100% of activation-excluded exons with downstream CTCF showed decreased 5hmC and increased 5mC at the binding site upon CD4+ T cell activation. Conversely, 100% of CTCF binding sites downstream of alternative exons that showed increased 5hmC and decreased mC were associated with upstream exon inclusion.

5. The cleanest results emanating from the genome-wide data is the influence of CpG density on the direction of hmC variations. Although noteworthy, this is to me not essential to this story nor does it directly support any of the earlier claims. It is, at this stage at least, an interesting but preliminary and independent observation.

We thank the Reviewer for this perspective and we agree. As the analysis of overall methylation levels at alternative versus constitutive exons has been removed from the current manuscript, we have also removed the relationship between CpG content and directional changes in 5hmC levels in this revised version.

References:


2nd Editorial Decision 06 July 2015

Thank you for submitting your revised manuscript for consideration by The EMBO Journal and my apologies for the slight delay in the reviewing process (one of the referees had asked for additional time and we thought it important to get the opinion from all three referees in this case).

Your study has now been seen by all three original referees and we have received their comments (included below). As you will see, while referee #3 is more positive about the revised manuscript I am afraid that ref#1 and #2 both express serious concerns about the overall conclusiveness of the study in light of the experiments added during revision; consequently, they cannot support publication in The EMBO Journal.

All three referees clearly respect the extensive work that has gone into both the original and the revised version of your manuscript, but at the same time they share skepticism about the validity of the model given the identical binding pattern seen for CTCF on 5meC and 5hmC-containing DNA in vitro. I realize that ref #3 is being more positive on the implications of this finding, but although the conflicting in vitro and in vivo data may not make the proposed model obsolete (as stated by the ref) it also does little to conclusively support it.

Given these negative opinions from the referees, I am afraid we have to come to the conclusion that we are unable to offer further steps towards publication in The EMBO Journal.

Thank you in any case for the opportunity to consider this manuscript. I am truly sorry we cannot be more positive on this occasion, but I do wish you all the best for rapid publication of this work elsewhere.

REFEREE REPORTS

Referee #1:

The fact that CTCF binding to CD45 Exon 5 DNA is not affected by CpG methylation (mC) in EMSA is problematic because it breaks the mechanistic chain of events. The author can only conclude that DNA methylation causes "something" to interfere with CTCF binding and Pol II accumulation, and someone will have to look at this again to really understand the mechanism. Only mentioning this issue in Figure 6 is also very misleading for the reader.

Referee #2:

I did not find that the revisions went very far to clarify the complex relationship between 5hmC and alternative splicing. I am afraid that if anything the relationship has become more obscure as a result of the finding reported in the rebuttal that, contrary to the widely held view, 5meC does not in fact markedly inhibit CTCF binding to exon 5 sequences in vitro. It is therefore a bit misleading that this dogma, which now seems not to apply in all cases, was repeated in the abstract of the revised manuscript. As a result the correlation between hi 5hmC/lo 5meC and enhanced CTCF binding to CD45 exon 5 does not at present have a satisfactory explanation, and might be due to unknown indirect effects. It therefore seems premature to draw firm conclusions about a molecular mechanism linking TET activity, 5hmC, CoTC binding, pol II pausing and exon skipping. The authors conclusion about genome wide effects of 5hmC on alternative splicing was not strengthened by the revision. The conclusion based on Figure 7F that "In sum, through high-stringency iterative filtering in naïve and activated CD4+ T cells genome-wide data, we provide evidence that oxidation of methylcytosine at CTCF binding sites globally regulates upstream exon inclusion" is not justified as it applies to only 41 (25+16) splicing events where a large difference in exon inclusion correlated with reciprocal changes in 5meC/5hmC. A secondary point is that as I stated in my original review, the relation between 5hmC and pol II pausing in exon 5 remains a bit flimsy as it still only supported by data for a single PCR amplicon rather than multiple amplicons or higher resolution pol II ChIP-seq analysis.
Referee #3:

This revised manuscript by Bailly et al. shows substantial improvements. Authors appropriately addressed most of my points. While their amendment of the manuscript has led to a clearer version overall, the changes brought to the genome-wide part are satisfactory. First, they have discarded, as suggested, "unrelated" observations which help the flow of the manuscript. Second, they have pushed their analysis in the right direction and managed to observe patterns allowing them to support their model and better understand CTCF-mediated alternative splicing and its modulation by DNA epigenetic modifications. Furthermore, the additional work they have executed and their transparent reporting of additional experiments, which sometimes, do not bring additional support to their model, shows the authors' good faith. These results, while slightly disappointing, are well discussed by the authors and do not render the proposed model obsolete. In conclusion, in view of the modifications brought to the earlier version I would recommend the manuscript for publication.

Appeal to resubmit 23 September 2015

I am sorry for the delay in responding to this decision. We were disappointed by the response to the revised manuscript, but were even more perplexed by our inability to demonstrate CTCF discrimination between mC and hmC in the requested EMSA assays. After additional experimentation during these past two months, we believe we have uncovered the source of the discrepancy between the in vitro and in vivo data: CTCF does not interact with hmC, but rather binds strongly to the oxidation derivative 5caC. By using smaller probes that only contain the modified residues in the CTCF binding site, we established EMSA conditions that reproducibly show CTCF binding to unmethylated, but not hmC or mC containing DNA. These conditions clearly demonstrate that CTCF is also not able to bind to 5fC, but strongly interacts with 5caC. Binding to 5caC can be competed with both unmethylated and 5caC containing competitor, and produces a robust super-shift when pre-incubated with anti-CTCF antibody. These data are attached as a PDF for your reference. We further demonstrate that like 5hmC, 5caC levels are reduced at CD45 DNA following T cell activation. These results are consistent with a model in which TET-catalyzed 5caC (through a 5hmC intermediate) promotes CTCF binding through direct interaction.

As we focused our entire first round of revisions on the requested genome-wide data, I wondered if you would allow an additional revision in this case. As you are aware, two of the reviewers were essentially happy with the CD45 work in the first place. We now fully roundup the CD45 data and further establish, for the first time, a role for CTCF-5caC interaction in the genome. These results are in line with the results of Spruijt et al (Cell 2013), in which they identify CTCF as a 5caC specific reader (supplemental data). More broadly, this demonstration reveals an intriguing potential link to CTCF-mediated pol II pausing at 5caC-rich intragenic DNA. As I am sure you are aware, He, Fu and Wang recently described a relationship between 5caC and reduced pol II elongation (Nature 523, 621–625, 30 July 2015). Our results raise the possibility that some of the observed effects may be mediated through interaction with CTCF at intragenic locations.

We are currently preparing the manuscript containing the 5caC data for submission. As we have already established a relationship with EMBO J, we hope that you will consider taking another look. As a caveat, we do not have 5caC genome-wide data and would have to remove the 5hmC results from the revised version. I realize this was an important element for Editorial consideration, but I would like to remind you that two of the reviewers did not feel that the genome-wide perspective was required, and we have now devoted a year to this ultimately unfruitful effort. We hope that you will reconsider this stance. As an expert in the field of methylation and splicing, I maintain that the CTCF/5caC connection is a conceptual advance that broadens our concept of the intragenic epigenome to include an adaptor platform of methylcytosine derivatives and distinct methyl-sensitive readers.

EDITOR’S RESPONSE
Thank you again for contacting us about the possibility to resubmit your manuscript including the new EMSA data supporting that CTCF does discriminate between 5hmC and mC in vitro. I have now had the chance to go through the previous submissions of your work as well as the reports from the referees and to discuss this in some depth with our chief editor.

We are glad to hear that you have been able to clarify the discrepancy between the in vivo and in vitro data for CTCF binding to mC. We would therefore offer to send the manuscript back to referees #1 and #3 at this stage. We would choose these two referees in order to speed up the process and to ensure the focus is on the new data in hand and not the perceived overall interest. While we realize that the absence of global binding data for 5caC makes it difficult to say how much of the 5hmC CTCF ChIP signal could reflect 5caC binding, we would still recommend that you leave the available genome-wide data in the manuscript at this stage so that the referees can comment whether it adds to the study per se. I would like to emphasise that the final decision and outcome for your new manuscript will depend on the feedback from the referees. Please be aware that we cannot therefore predict the outcome of this consultation and thus we cannot commit at this time to offer publication of the revised manuscript.

We would again like to thank the Referee for his/her initial assessment that our study was “thorough and technically creative” particularly with respect to the CD45 work. As detailed in the previous rebuttal, we responded to the Referee’s earlier concerns in a detailed point-by-point. However, one of the provided figures for the revision raised a new concern related to a CTCF EMSA showing interaction with CpG methylated CD45 DNA (5mA). As there is precedence in the literature showing that methylation does not necessarily block CTCF interaction in vitro (see Discussion), we perhaps misjudged the importance of resolving this issue for the current manuscript. Our overall goal was to show that TET-catalyzed oxidation of 5mA can facilitate CTCF interaction, which then promotes weak exon inclusion. We felt, and still feel, that the provided in vivo evidence clearly demonstrates that CTCF is antagonized by 5mA in cells. Nonetheless, as Referees 1 and 2 raised substantial concerns about the CTCF EMSA, we have since re-approached the biochemistry from a more stringent perspective anchored in first determining conditions that mimic the observed in vivo results (i.e. wherein CTCF binds to cytosine but not methylcytosine-containing DNA). The resulting analysis revealed an intriguing connection to 5-carboxyctosine (5caC), as detailed below.

Response to Specific Comments

_The fact that CTCF binding to CD45 Exon 5 DNA is not affected by CpG methylation (mC) in EMSA is problematic because it breaks the mechanistic chain of events._

We thank the Referee for highlighting the 5mA CD45/CTCF EMSA presented in the revised manuscript as a significant concern. We have since addressed this issue in greater detail and believe the manuscript to be significantly improved as a result. Before going into the details, however, I would like to highlight that our intent in showing the initial 5mA CD45 EMSA was to merely show that CTCF ‘can’ interact with oxidized cytosine species in the context of CD45 DNA. We did not mean to imply that this was definitive of CTCF binding _in vivo_. CTCF is an exceptionally complicated molecule that has essentially defied “rule” assignment. A number of research groups are focused on defining CTCF binding preferences, including the murky relationship to DNA methylation. The consensus from these studies is that CTCF interaction with methylated DNA is not as clear-cut as the early literature would have you believe. Certainly, Schubeler and colleagues have clearly shown CTCF interaction with bisulfit-resistant DNA at low methylated regions (LMRs). Considering the genome-wide enrichment of 5mA at CTCF binding sites and direct evidence that CTCF interacts with TET1, the resistant species most likely reflects 5mA. However, the functional relevance of such an association had not been defined. Noting that LMRs and differentially methylated regions (DMRs) are both enriched at transcribed sequences, these observations raised the possibility of a functional link between 5mA oxidation and CTCF in the regulation of pre-mRNA processing, as we show here. Our goal with the initial EMSA was to show that CD45 DNA represents an LMR that is technically competent to bind methylated DNA in the highly artificial context of _in vitro_ biochemistry. However, we do realize that the lack of observed preference for
5hmC over 5mC in the queried linear templates, tempered enthusiasm for our presented model suggesting that 5mC oxidation promotes CTCF interaction.

We now provide evidence that CTCF does indeed preferentially bind to oxidated methylcytosine species, as shown for 5caC. The observed 5caC binding is in fact stronger than to unmethylated DNA. We further show that 5caC is enriched at CD45 exon 5 in naïve CD4+ T cells. Given that several studies have established 5caC as extremely transient in DNA, such that mapping is impaired in the absence of TDG knockdown (see Discussion), ready detection of 5caC at CD45 DNA in the absence of any manipulation is significant in and of itself. This observation raises the possibility that efficient CTCF binding to 5caC-containing DNA “protects” against further oxidation. Whether 5caC is a preferred CTCF binding substrate in the context of nucleosomal DNA in vivo is not clear at this time. However, due to the highly charged nature of 5caC it is possible that its presence in double-stranded DNA may favor CTCF binding through structural or electrophysical means. It is further worth noting that CTCF does in fact interact with nucleosome-free DNA in vivo and 5mC appears to favor nucleosome assembly. Notably, our study is supported by an unbiased mass spectrometry survey that identified CTCF as a 5caC-specific reader. Together, the sum of our results provide a rationale for and evidence supporting a role for the TET proteins in CTCF-dependent alternative splicing through oxidation of overlapping 5mC.

Response to Referee #2:

I would like to thank the Reviewer for the helpful comments throughout the revision process. The Reviewer raised a new concern related to the CTCF/5mC EMSA performed for the primary review. We have revisited the EMSAs in the current submission, and hope that the Reviewer will find this issue adequately resolved. However, I am sorry to find that the Reviewer still has concerns related to the small number of genes identified in the genome-wide analysis. We respond to both these points below.

Response to Specific Comments

I am afraid that if anything the relationship has become more obscure as a result of the finding reported in the rebuttal that, contrary to the widely held view, 5meC does not in fact markedly inhibit CTCF binding to exon 3 sequences in vitro.

We thank the Reviewer for highlighting the presented EMSA as a confounding issue that required a clear resolution. We have now repeated the EMSA experiments under conditions that discriminate between unmethylated and 5mC-containing CD45 DNA, consistent with the in vivo observations. In this setting, we find that CTCF fails to bind to 5mC and 5hmC-containing DNA, but instead shows robust interaction with 5caC. This finding is consistent with a proteomics survey identifying CTCF as a 5caC-specific reader. While not definitive proof of CTCF binding preferences in the context of chromosomal DNA, this finding reveals an intriguing link between TET function and CTCF binding and provides a rational basis for our observations in the CD45 model system. Please see the response to Referee 1 and new Discussion for an expansion on these concepts.

The conclusion based on Figure 7F that “In sum, through high-stringency iterative filtering in naïve and activated CD4+ T cells genome-wide data, we provide evidence that oxidation of methylcytosine at CTCF binding sites globally regulates upstream exon inclusion” is not justified as it applies to only 41 (25+16) splicing events where a large difference in exon inclusion correlated with reciprocal changes in 5meC/5hmC.

We understand the reviewers’ concern but stand by our assessment that by focusing on sites that both lose 5hmC and gain 5mC, or vice versa, we have certainly filtered away a large number of meaningful events. As we clearly demonstrate in the CD45 model system, changes in 5hmC do not show an equal magnitude change in 5mC, likely due to the presence of 5mC elsewhere in the IP fragment. I would also like to remind the reviewer that our filtering strategy omitted CD45 from our genome-wide analysis at the first step through removal of complex splicing patterns. With that said, even if the final determinations were accurate reflections of 5hmC/CTCF regulated exons, we would not have been surprised. Alternative pre-mRNA splicing is regulated at many levels and we would not have expected widespread changes at the nuanced developmental transition from a naïve to activated T cell type, especially in the absence of polarizing cytokines. Importantly, despite these
limitations, the genome-wide analysis succeeded in generalizing the observed changes at the CD45 model gene and revealed an intriguing link to upstream CTCF sites.

A secondary point is that as I stated in my original review, the relation between 5hmC and pol II pausing in exon 5 remains a bit flimsy as it still only supported by data for a single PCR amplicon rather than multiple amplicons or higher resolution pol II ChIP-seq analysis.

As mentioned in the previous rebuttal, we sincerely apologize, but we do not have any material remaining from the previous pol II ChIP. This is not a simple repeat experiment as the unmethylated clone ultimately showed methyl-drift in culture. As a result, we would have to return to the initial clone generation stage to produce material for additional amplicon testing. As the goal of the clones was to establish that pol II pausing and exon inclusion are dependent on an intact CTCF binding site, rather than indirectly through 5oxiC, we would hope that the combined information in the figure would reduce the Reviewer’s concerns.

Response to Referee #3:

We thank the Referee for his/her continuing support of this manuscript. We are glad to find that the Reviewer found the revised manuscript containing a reanalysis of the genome-wide data to now be suitable for publication in EMBOJ. We also appreciate the Referee’s acknowledgement that our presentation of the complicated CTCF/5mC EMSA was done in good faith. We were aware that the demonstration that CTCF binds to CpG methylated (5mC) CD45 DNA was at odds with the observed in vivo effects. However, we had hoped that considering that others in the CTCF/DNA methylation field have reported similar observations with low CpG content substrates in vitro, the Referees would instead focus on the cellular observations wherein we show a clear relationship between TET catalyzed 5mC oxidation and exon inclusion in a number of distinct cellular settings. However, as the EMSA did emerge as a contentious issue during review, we have revisited these assays in the current submission with a focus on first establishing conditions wherein CTCF fails to bind 5mC-containing CD45 DNA. In so doing, we find that CTCF preferentially interacts with the terminal TET-catalyzed oxidation product, 5caC. These data reveal a direct link between CTCF and oxidation derivatives, but are not definitive evidence of binding preferences in vivo where nucleosome assembly and other determinants have the capacity to influence CTCF binding. The expanded discussion focuses on these and related possibilities. In a continued spirit of transparency, the previous 5mC CpG methylated EMSA is referenced in the text and shown in the Supplementary Figures. We hope that the Referee will continue to support this manuscript with the newly presented data.

Referee’s Comments

This revised manuscript by Bailly et al. shows substantial improvements. Authors appropriately addressed most of my points. While their amendment of the manuscript has led to a clearer version overall, the changes brought to the genome-wide part are satisfactory. First, they have discarded, as suggested, “unrelated” observations which help the flow of the manuscript. Second, they have pushed their analysis in the right direction and managed to observe patterns allowing them to support their model and better understand CTCF-mediated alternative splicing and its modulation by DNA epigenetic modifications. Furthermore, the additional work they have executed and their transparent reporting of additional experiments, which sometimes, do not bring additional support to their model, shows the authors’ good faith. These results, while slightly disappointing, are well discussed by the authors and do not render the proposed model obsolete. In conclusion, in view of the modifications brought to the earlier version I would recommend the manuscript for publication.

3rd Editorial Decision 30 October 2015

Thank you for submitting a new and revised version of your manuscript. 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 major criticisms have been sufficiently addressed and recommend the manuscript for publication, pending minor revision. I would therefore encourage you to submit a final version of the manuscript, including the following:

-> Please provide data on the 5caC contribution to CTCF binding on a second target site (ref#1), if
this is possible within a short and reasonable time frame. If not, then I'd be happy to discuss this further.

-> Please also comment on the possible contribution from residue composition relative to probe length in the EMSA experiments (ref #1). Just for your clarification, the referee numbering is the same as in the previous round (ie ref #2 was not consulted on the new version)

-> As a final point, and this is really a suggestion rather than a request, I noticed that the final manuscript text has become very long. We do not have an official limitation on manuscript length, but for the sake of accessibility to the reader you may want to think about shortening a few passages (the standard manuscript length is around 60,000 characters).

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:

The new Fig6C and 6D showing that CTCF binding is recovered not at the 5hmC stage but only at the 5caC oxidation stage are convincing and constitute an important addition to the manuscript. However, I would recommend verifying the importance of 5caC for the binding of CTCF on at least one other well-characterized CTCF binding site. Also, the 72bp probes was generated by PCR in the presence of 5mC, 5hmC, 5fC, or 5caC. This means that all the Cs get modified, while in earlier probes only some Cs were modified. Possibly, this may be a difference more important than the length of the probe.

Referee #3:

For the newest version of their manuscript, authors have produced new EMSA results showing a binding preference of CTCF for 5CaC-containing DNA over unmethylated DNA. With this, they add strength to their hypothesized model in which TET-mediated oxidation of 5mC (5CaC being an oxidized derivative of 5hmC) facilitates upstream exon inclusion by antagonising 5mC-mediated CTCF eviction. Hence, I continue to support these results.

3rd Revision - authors' response 12 November 2015

Response to Referee #1:

We are happy that the Referee found our demonstration that CTCF preferentially interacts with 5caC-containing DNA in the context of CD45 sequence to be a “convincing” and “important addition to the manuscript.” We agree and thank the Referee for pointing us in this direction through previous assessments that the failure of CTCF to distinguish between 5mC and 5hmC obviated the mechanistic basis for TET1/2 involvement in CTCF dependent splicing. The Referee raised two additional concerns related to the new 5caC EMSAs. We have addressed both points as outlined below and hope that the Referee will now find the manuscript acceptable for publication in EMBOJ. We thank the Referee for helpful comments throughout the review process.

Response to Specific Comments

1) I would recommend verifying the importance of 5caC for the binding of CTCF on at least one other well-characterized CTCF binding site.

We now provide evidence that 5caC promotes CTCF interaction at a second CTCF binding site. For this purpose, we chose the KCNA2B gene based on a strong ChIP-seq peak with evidence of overlapping 5hmC in our genome-wide analysis. Importantly, the KCNA2B gene was amenable to assay design due to clear overlap between the ChIP-Seq peak and a defined CTCF motif, as well as favorable proximal sequence for primer design. Notably, as APOPT1 and HAPLN3 were used for
validation of the genome-wide analysis, we had hoped to use either of these genes for the supporting EMSA. Unfortunately, the sequences flanking the CTCF binging site of both genes were repetitive and GC-rich, thereby obviating PCR based assay design.

As shown in the new FigS6, generation of KCNA2B probes through PCR in the presence of modified dCTP recapitulated the CD45 results: CTCF failed to interact with 5mC or 5hmC, but formed complexes in the presence of unmethylated and 5caC-containing DNA. Notably, in this case binding was substantially increased in the presence of 5caC as compared to unmethylated DNA and a minor complex was formed in the presence of 5fC, suggesting that the KCNA2B binding site is particularly sensitive to increasing oxidation state. While these results rely on PCR amplification, creating an artificially high level of modified cytosines, they establish the clear preference of CTCF for 5caC containing sequences. This is further indicated in the context of CpG methylation, as described below.

2) Also, the 72bp probes was generated by PCR in the presence of 5mC, 5hmC, 5fC, or 5caC. This means that all the Cs get modified, while in earlier probes only some Cs were modified. Possibly, this may be a difference more important than the length of the probe.

We fully agree with this point and only relied on the PCR assays as they allowed us to establish conditions wherein CTCF did not interact with 5mC-containing DNA. Nonetheless, as we were concerned about the contribution of non-CpG methylation, we initiated commercial synthesis of 41 base pair CD45 probes that only contain 5caC at the 3 CpG that directly overlap the CTCF binding site. We previously established that CTCF binding to the 41 bp CD45 probes was very weak. We now show that symmetric addition of just the 3 5caC residues results in a 1.6 fold increase in CTCF binding as compared to unmethylated probe. These findings clearly establish that CTCF binding is enhanced in the presence of 5caC-containing DNA and provide a rationale basis for TET protein involvement in CTCF-dependent alternative pre-mRNA splicing.

Response to Referee #3:

We would again like to thank the Referee for his/her continuing support of this manuscript. Indeed, the new data showing preferential interaction of CTCF with 5caC-containing DNA bolster our model that the TET proteins promote CTCF-dependent splicing events. In this resubmission, we provide evidence that CTCF binding at a site in the KCNA2B gene is also enhanced in the presence of 5caC, suggesting some generality. In addition, we show that introduction of 5caC at just the CpGs directly overlapping the CD45 CTCF binding site is sufficient to enhance interaction. We thank the Referee for many valuable comments and input throughout this review process.

Referee’s Comments

For the newest version of their manuscript, authors have produced new EMSA results showing a binding preference of CTCF for 5CaC-containing DNA over unmethylated DNA. With this, they add strength to their hypothesized model in which TET-mediated oxidation of 5mC (5CaC being an oxidized derivative of 5hmC) facilitates upstream exon inclusion by antagonising 5mC-mediated CTCF eviction. Hence, I continue to support these results.