THE
EMBO
JOURNAL

Manuscript EMBO-2015-93561

Transcription rate and transcript length drive formation of chromosomal interaction domain boundaries

Tung B. K. Le and Michael T. Laub

Corresponding author: Michael T. Laub, MIT

Review timeline:

Submission date:	25 November 2015
Editorial Decision:	08 January 2016
Revision received:	07 April 2016
Editorial Decision:	06 May 2016
Revision received:	16 May 2016
Accepted:	20 May 2016

Editor: Hartmut Vodermaier

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 08 January 2016

Thank you for submitting your manuscript on chromosomal boundary formation in *Caulobacter* for our consideration. I would like to apologize for the delay in getting back to you with a decision, related to the fact that the reviewing process unfortunately ran into the end-of-the year holiday break.

We have now received all comments from three expert referees who agreed to review your study, copied below for your information. As you will see, the referees' opinions are somewhat equivocal; while the referees appreciate the current extension and broadening of your earlier Hi-C analyses, they raise a number of substantial concerns with the presentation, interpretation and conclusiveness of some of the data. In our view, these concerns currently preclude EMBO Journal publication of the study in its present form.

In light of the overall interest of the work, I would nevertheless be inclined to give you an opportunity to address the referees' criticisms by way of a major revision of the study. For such a revision to be successful, it will however be essential to not only improve the presentation and interpretation of the results, but also to add further experimental analyses (such as the Hi-C replicates demanded by referee 2) to strengthen the conclusions, increase the decisiveness of the data, and resolve certain ambiguities in the current results (such as the potential confounding effect of RNA polymerase presence on crosslinking efficiency). Moreover, since it is our policy to to allow only a single round of major revision, please note that it will be important to carefully and comprehensively respond to all points raised during this round. I would therefore be open to discussing an extension of our standard three-months revision time here if this might be necessary. As always at The EMBO Journal, competing manuscripts published elsewhere during such an official revision period would have no negative impact on our final assessment of your revised study. Additional information on preparing and uploading a revision can be found below.

Thank you again for the opportunity to consider this work for The EMBO Journal, and please do not hesitate to contact me should you have any comments or questions regarding the referee reports or this decision. I look forward to your revision.

REFEREE COMMENTS

Referee #1:

The manuscript by Le and Laub, "Transcription rate and transcript length drive chromosomal interaction domain boundary formation" extends the analysis of the Caulobacter crescentus chromosome using the chromosome conformation captures technique coupled with Hi-C deep sequencing. Here, they compare chromosome structure in cells growing in rich medium with cells starved for nitrogen and carbon, probing the local interactions along each replichore. Based on selected results from fluorescence microscopy and Hi-C chromosome crosslinking analysis they conclude that "transcription rate and transcription length, independent of concurrent translation, drive formation of domain boundaries"

This Hi-C study follows a ground breaking 2013 paper in which the Laub lab showed the first examples of local long range interactions along a prokaryotic chromosome plus cross interactions between the two replichores. However, the writing style clarity needs to ne improved on several critical points. For one, topological domain structure has been studied in both E. coli and Salmonella for many years. Results from γδ resolution analysis using time restricted resolvases and microarray analysis of supercoil responsive genes in E. coli agree. Supercoil domains exist, they form at stochastic sites along the chromosome that change frequently, and that the average domain size is 10 - 20 KB.

Specific points in order:

1. The authors redefine chromosome domains to be something that can only be reliably detected with Hi-C technology. They do experiments to prove that "transcription lengthening of DNA" is the one true solution to domain structure. This is sort of a branding strategy. Actually, two of the domains studied here (ribosomal RNA genes and the Atp operon) were initially discovered and analyzed in Salmonella 1,2 Hi-C appears to miss many dynamic structures that change with time, and by using the generic term domains for the Hi-C clusters they ignore or assume that all the others are insignificant.

2. One may ask why are only the most actively transcribed genes efficiently captured as crosslinked species in the genome using Hi-C. The authors addressed one issue, which is a restriction enzyme skew. However, they neglect a more obvious explanation, which has nothing to do with domain boundaries. The bacterial chromosome is hard to crosslink compared to eukaryotic DNA because it is more naked than histone-DNA. Crosslinking is more efficient on protein-coated DNA, so RNA polymerases, which are locked on the template as hyper stable ensembles are able to crosslink to each other more efficiently than to regions with very few RNAPs. The crosslinking efficiency argument explains why Rifampicin treatment causes a disappearance of Hi-C regions in bacteria but has almost no effect for Hi-C in eukaryotic chromosomes.

3) p.6 "Although high gene expression and domain boundaries are clearly correlated, not all highly transcribed genes were associated with domain boundaries, in both growing and starved cells. For example, genes encoding tRNAs are very highly-expressed in fast-growing cells, but were rarely associated with domain boundaries. Notably, tRNAs (~76-90 bp) are much shorter than an average bacterial gene (~1000 bp)."

This fits the RNAP crosslinking efficiency hypothesis above.

4) p.6 "To test this hypothesis, we first calculated 'rpkpm * transcript length' for each gene or operon in the Caulobacter genome and plotted these values as a function of genome position (Fig 1B-C, Appendix Fig S2, Appendix Table S2)."

Transcription has two relevant numbers, the RNAP initiation frequency and the elongation rate (NTP/sec) With knowledge of where promoters are, I can see how it is possible to determine regions with long transcripts. However, obtaining the initiation efficiency is not so obvious.
Just knowing the abundance of the RNA for a specific gene is a start but there is also the question of half-life. I have no idea of what was done, and anybody who printed out Table 2 in the supplementary data is a sucker. It is incomprehensible and seems to be 60 pages long from just trying to scroll to the bottom.

5). p7. "(Fig 2). The Hi-C maps indicated that a sharp domain boundary was formed when the terminator cassette was inserted at least 2060 bp from the transcriptional start site with no or very weak domain boundaries formed when the transcript length was engineered to be less than 2060 bp. Taken together, our results indicate that the length of a highly expressed gene plays a critical role in determining whether it creates a chromosomal domain boundary."

They state throughout the manuscript that transcript length is an issue. However, it doesn't seem to be so important . If a 2 kb gene makes a domain boundary that is as clear cut as a 10 kb ATP operon, where's the beef?. Again, the only thing that seems to be important for Hi-C is the high density of RNAP occupying the transcribed track, which is consistent with a crosslinking efficiency effect that involves RNAP.

Supercoil diffusion during transcription.

6) p.9
"To quantify supercoil diffusion, we developed a recombination assay based on the γδ resolvase TnpR (Deng et al, 2005; Higgins et al, 1996) (Fig 5A).....
"Recombination was highly efficient, with TnpR induction, driving a 4-log drop in plating viability in the presence of tetracycline (Fig 5B)....."
then
"A promoterless rsaA variant failed to block supercoil diffusion yielding only ~4% tetR colonies....

What caused a 4 orders of magnitude resolution efficiency to drop to under 2 orders of magnitude?

7) p.9. "Transcription based inhibition of supercoil diffusion is likely not simply a consequence of DNA-binding by RNA polymerase between two res sites. Cells harboring 10 consecutive lacO sites in between the res sites and expressing LacI-YFP did not affect recombination (Fig 5D)."

??? Binding of a few lac repressors would not be expected to inhibit supercoil diffusion because each operator/repressor module is flexible and will rotates with the DNA.

8) "We conclude that it is likely the DNA unwinding associated with active transcription, rather than protein binding, that drives supercoil diffusion inhibition."

??? This is the same conclusion reached by Booker 1

9) P. 11 "Taken together, our microscopy and Hi-C analyses indicate that nucleoid expansion does not occur uniformly across the chromosome. The DNA at chromosomal domain boundaries is likely less compacted and therefore stretches more than the DNA inside a domain. These results further support our conclusion that domain boundaries produced by highly expressed and long genes produce a spatial barrier in vivo that diminishes contacts between loci in adjacent domains."

Translation can exerts a large effect on plasmid topology in E. coli The effect involves co-transcriptal translation and insertion of proteins that are membrane associated or transported through the membrane to external locations on the cell surface or periplasm 3. Anchoring the whole transcription-translation to a stable surface exerts forces on the E. coli nucleoid that cause expansion. When cells are treated with chloramphenicol, nucleoids contract 4. This could be relevant for both the rsaA operon (an extruded cell surface component) and the ATP operon, which encodes multiple integral membrane proteins. A quick look at other domain genes in the Caulobacter suggest that membrane proteins may be prominent players. While it would not obliterate the RNAP binding to DNA it may well have an impact on distant markers in each replichore.

It might also be an interesting experiment to analyze marked positions flanking the ectopic rsaA operon and see what happens when expression changes from on to off and also from off to on.

10 p.15
"Thus, the high rates of transcription at domain boundaries may enable yet more transcription, forming a positive feedback loop that could potentially contribute to the burstiness of gene expression."

"burstiness" ??? I doubt that many readers of EMBO J will understand this term. If the reference is to transcription bursts, an important paper was published last year from Sunney Xie's group and the explanation involves positive supercoiling that builds up to shut gene expression off. They analyzed this phenomenon in vitro and in vivo. 5

1 Booker, B. M., et al. Mol Microbiol 78, 1348-1364, (2010).
2 Rovinskiy, N., PLoS Genet 8, e1002845 (2012).
3 Lynch, A. S. & Wang, J. C. J. Bacteriol. 175, 1645-1655 (1993).
4 Van Helvoort, J. M. L. M., Kool, J. & Woldringh, C. L. J. Bacteriol. 178, 4289-4293 (1996).
5 Chong, S., Chen, C., Ge, H. & Xie, X. S. Cell 158, 314-326, doi:10.1016/j.cell.2014.05.038 (2014).

Referee #2:

Previously the authors have used Hi-C methodologies to show that topological domain boundaries in C. crescentus appear to be defined by the presence of highly expressed genes. In this study they again use Hi-C experiments to show that high levels of transcription alone are not sufficient for boundary formation. Rather it appears that a combination of long transcript length and high levels of transcription define chromosomal domain boundary.

I think the paper is interesting and potentially important for the field. However, at present I do not think that the manscript is ready to be published in EMBO journal

My principal concerns are:

1) From Table S5 there do not appear to be any replicates of the Hi-C datasets. Personally I think that it is essential for Hi-C datasets to be replicated. For some of the experiments, I think an argument can be made for only analyzing a single dataset. For example, it could be argued that in figures 2 and 3 that the two datasets with transcripts greater than 2kb and those less than 2kb are basically equivalent. In these instances analysis of the global correlations between the datasets may be acceptable. However, I think that the datasets of starved cells should be replicated - because the changes relative to exponentially growing cells are the core result of the paper. Also, since the differences for interdomain interactions shown in figure 6 appear, to me, to be quite subtle these should also be replicated.

2) In all the experiments using the ectopically positioned rsaA gene the authors definition of a domain boundary appears to be subjective and imprecise. This is puzzling since in fig S2 and fig. 4 the authors do use directional preference of contacts to define where a domain boundary is. Directional preference should be used for all the HiC boundary analysis throughout the paper. This will allow the reader some scale of how big the changes in ectopic boundary formation are between the different conditions used.

Other concerns;

3) I understand that the proposition that high levels of transcription inhibit supercoil diffusion in bacteria has been made previously (B. M. Booker et al. Molecular Microbiology. 78, 1348-1364 (2010)). The authors should make clear that their results confirm these observations.

4) In fig.S6C the authors use changes in global contact probability as cells elongate to argue in favour of their hypothesis. I do not think that the global contact probabilities provide relevant information for their hypothesis since boundaries primarily inhibit inter-domain contacts < 100kb. To make their point they need to subset the the intra and interdomain contacts < 100kb. Their hypothesis predicts that <100kb inter-domain contacts should be relatively decreased relative to <100kb intra-domain contacts as the cells elongate.

Referee #3:

The article by Le and Laub "Transcription rate and transcript length drive chromosomal interaction domain boundary formation" proposes a mechanism (the self-explanatory title) for the formation of boundaries preventing adjacent DNA regions to collide with each other's along the Caulobacter crescentus genome. The data rely primarily on capture of chromosome conformation experiments performed in a variety of growth and mutant conditions. The mechanisms proposed here could presumably apply to other species as well.

The authors start by describing the concomitant changes that occur in the transcriptional pattern of C. crescentus and genomic domain organization in different growth conditions. They analyzed the boundaries between chromosomal domains presenting enrichment in DNA-DNA contacts (CIDs). In an initial report, these borders were found to correlate with the presence of highly expressed genes, suggesting that supercoiling was generating structures preventing contacts between two CIDs. As noted by the authors and others, not all highly expressed genes resulted in a border, nor did all borders consist in such a gene. Here they identify another parameter, gene length, that when combined with a high level of gene expression positively correlate with the presence of a boundary. They propose that this combination explains to a large extent the discrepancies pointed out previously. The paper is clear and reads easily. The experimental setups allowed the authors to test for interesting hypothesis and some of the results are really clear. I nevertheless think that some of the data are a bit interpreted to match the model, and that some of the conclusions are not fully supported by the experiments. These concerns consist essentially in two points. The first one is the match between the observations made in WT cultures presented in Figure 1 and the experimental setup in Figure 2, while the other one relates to the results in Figure 6.

Comparison between high expression levels of long genes and the presence of domain boundaries in different growth conditions are presented in Fig. 1B and C, with three regions emphasized in panels 1D, E, F that present convincing examples of the claim. The observations are interesting, but the authors could improve the quantification of these changes in the text and on the figures. Indeed, when I compare panel 1B and 1C, it is not that obvious that boundaries change along with variation in gene expression in the two experiments (given the gene length does not vary when one compare two experiments). Either it is due to misalignment of the different panels, but often a change in gene expression does not affect the topological boundaries and reciprocally, new boundaries do not necessary appear along with a higher gene expression level (1B: barrier 1-2, 2-3, 6-7, 10-11, etc. 1C: 3-4, 5-6, 14-15, etc.) Actually, it seems that boundaries that do not correlate with a high expression level in one condition will correlate in the other condition (e.g. barrier 1-2 in rich medium, barrier 5-6 in starvation, 6-7 rich medium, 10-11 rich medium, 18-19 rich medium, etc.) It suggests that these boundaries are positioned independently from the expression level of the (long) gene. Here again it would be informative to indicate on figure 1 the long genes and provide more statistics. How often a boundary does appear concomitantly with the increased expression of a gene? The claim that length of a highly expressed transcript determines the strength of a domain boundary is therefore only partially supported by the data from Figure 1. The authors should describe more precisely the various changes observed in the different growth conditions. Then, they should integrate these observations in light of the nice experiment presented in Figure 2, and maybe ponder the claim they draw from the later experiment. To my eyes, the title is therefore a bit of an overstatement.

The contact maps of Figure 2 should be more focused on the region of interest. Here they span 1Mb, for no obvious reason (?). It would actually be nice to see a resolution slightly lower than 10kb but given the author used BglII, a six-cutter, I assume this would be difficult. The diagonal in the last panel of Figure 2 is narrower than the other panels. It is a detail but this is unfortunate given this is the most important panel. More details about the generation of the contact maps in the MM or SM should be provided here, such as number of contacts for each map, etc. Also, the color scale is missing, so it is not clear whether they are exactly the same for all of these panels. Not that this will change the interpretation, but that will help the reader to focus on the region and results.

The authors demonstrate elegantly that the boundaries do not result from enrichment in proteins, such as ribosomes or DNA polymerase. The controls they made are convincing and provide interesting new insights that will be helpful to decipher the actual mechanisms behind the generation of borders.

I have another concern regarding the "elongation" model proposed by the authors to explain how long genes generate boundaries, relating principally to Figure 6 and S6. First, I could not find in the supplementary figures representative microscopy images of the intra- and inter-domain pairs of loci for the data presented in Figure 6 and S6. Can examples be provided and will these data be made available somewhere? Also, in the synchronization experiments to recover elongated cells or the CtrA(D51E)Δ3Ω overexpression cells, the first time point always appears as quite fuzzy, as pointed out by the authors. Then the post-release time points appear quickly more contrasted and quite similar. These experiments could indeed provide information about the nature of the domains by providing clues about what is happening at their boundaries but as they are presented this is not the case to my eyes. First, I cannot see exactly where in the WT contact map the synchronized pattern matches. The maps are so distorted to improve the visual appearance of the contact signal that it makes comparison tricky. It would be important to include for of each of these time courses a panel with the WT data. Second, could an alternative explanation for the increase in sharpness of the border compare to time 0 be that the synchronization procedure is not that harmless? If these Hi-C experiments were performed on a sample containing a few dead/sick cells, maybe the pattern would also look like this? After release the cells would then gradually recover and the entire population converge towards a "sharper" signal. This is where having the WT signal (sync and/or async) would be valuable. Then, for the region presented in figure 6B, the intra/inter ratio of the interfoci distance measured in elongating cells also appear quite conserved: 0.65 (time 0) vs. 0.60 (3 hours). I agree that the dispersion of the distances is increased in the inter-domain positions, but the conserved ratio does not point at a more pronounced difference. On the other hand, the other example (presumably involving a border made of a long highly expressed gene? Annotation would be helpful) in Figure 3C shows the opposite: inter-domain distances are initially smaller than intra-domain distances, then become clearly larger. But the dispersion of distances is quite conserved between the two datasets. Therefore, these experiments suggest that the organization of the domains, as defined by Hi-C, is more complicated and that all visualized domains may not correspond to the same structural properties. Maybe this result from the definition of domain boundaries by using directionality index analysis, which is quite sensitive to local signal on the diagonal? More generally, more quantification of the cell morphology and aspect of these synchronized cultures should be made. Overall, Figure 6 does not present very conclusive data to my eyes. Another example are the two P(s) curves on Figure S6F, which do not look that different to me: I am not convince that the small discrepancy in short vs. long range contact observed here would withhold a biological duplicate. To support their claim, the authors should provide a clear experiment showing the synchronization does not induce "noise" in the contact data. They should also track additional pairs of loci to reach at a more robust conclusion.

Other comments: citation of existing literature should be improved. There is to my eyes a number of omissions or approximations that have to be fixed. In the introduction: that the authors were the first to perform "Hi-C" on bacteria is clearly an overstatement. Although this was not Hi-C strictly speaking, a very similar analysis was performed by the groups of Shapiro, Dekker and Church (Umbarger et al., 2011), and this article should be cited instead of Le et al 2013 as the first instance of genomic contact analysis in a bacteria. Several interesting findings were present in this manuscript, even though the presence of CID/TADs were not identified at the time. In addition, Marbouty et al. 2015 also noticed that long genes are found at domain boundaries as reported in the Figure 1 of their paper: this should be explicitly stated in the text. Marbouty et al should also be cited along Wang et al. as both papers report domains with highly expressed genes at their boundaries in B. subtilis (including the long gene borders) and were published concomitantly.

Minor comment: nowhere in the text the name of the restriction enzyme used is mentioned (only on Figure 1, apparently). Even if referring to Le et al 2013 is ok, this is important information that should be indicated in the MM of the present manuscript as well.

1st Revision - authors' response 07 April 2016

Referee #1:

The manuscript by Le and Laub, "Transcription rate and transcript length drive chromosomal interaction domain boundary formation" extends the analysis of the Caulobacter crescentus chromosome using the chromosome conformation captures technique coupled with Hi-C deep sequencing. Here, they compare chromosome structure in cells growing in rich medium with cells starved for nitrogen and carbon, probing the local interactions along each replichore. Based on selected results from fluorescence microscopy and Hi-C chromosome crosslinking analysis they conclude that "transcription rate and transcription length, independent of concurrent translation, drive formation of domain boundaries."

This Hi-C study follows a ground breaking 2013 paper in which the Laub lab showed the first examples of local long range interactions along a prokaryotic chromosome plus cross interactions between the two replichores. However, the writing style clarity needs to ne improved on several critical points. For one, topological domain structure has been studied in both E. coli and Salmonella for many years. Results from γδ resolution analysis using time restricted resolvases and microarray analysis of supercoil responsive genes in E. coli agree. Supercoil domains exist, they form at stochastic sites along the chromosome that change frequently, and that the average domain size is 10 - 20 KB.

Specific points in order:
1. The authors redefine chromosome domains to be something that can only be reliably detected with Hi-C technology. They do experiments to prove that "transcription lengthening of DNA" is the one true solution to domain structure. This is sort of a branding strategy. Actually, two of the domains studied here (ribosomal RNA genes and the Atp operon) were initially discovered and analyzed in Salmonella 1,2 Hi-C appears to miss many dynamic structures that change with time, and by using the generic term domains for the Hi-C clusters they ignore or assume that all the others are insignificant.

We do not redefine or 'brand' chromosome domains, and are puzzled by this comment. In our prior work (Le et al. 2013) we used Hi-C analyses to define 'chromosomal interaction domains' or 'CIDs' as regions of the chromosome in which loci interact preferentially with other loci in the same region, *i.e.* regions of relatively high pairwise interaction frequencies as measured by Hi-C. However, we were careful in that paper, and this current paper, to avoid equating these CIDs with other types of domains, including the supercoil domains referred to by the reviewer. Indeed, CIDs are likely different entities: supercoil domains are probably smaller (up to 10-20 kb) than CIDs (on the order of 100 kb) and possibly more dynamic. However, there could be a hierarchical organization with multiple supercoil domains comprising a CID and multiple CIDs comprising what the *E. coli* community calls a 'macrodomain'. We have not made any assumption or statements in the paper that other domains are insignificant - rather, because our paper is focused on CIDs and the mechanisms that drive CID formation, these domains naturally garner most of our attention. Additionally, we make no claim that transcription-induced lengthening of DNA is the "one true solution" to all domains. We argue, based on several independent and mutually reinforcing pieces of data, that transcription likely plays a critical role in forming many CIDs. Moreover, we note when discussing data in Fig. 5 that short and long transcripts are both effective at blocking supercoil diffusion even though only long transcripts yield clear boundaries by Hi-C. These results underscore the notion that chromosomal interaction domains (CIDs) and supercoil domains are not equivalent. Hence, we wrote (p. 10 of the manusucript) that "our results suggest that transcript length affects whether a gene or operon produces a chromosomal domain boundary, but that longer transcripts are not substantially better at blocking supercoil diffusion, suggesting that chromosomal domains are not identical to supercoil diffusion barriers." Similarly, in the Discussion we state that "both short and long transcripts can block supercoil diffusion (Fig 5), but only the latter create clear domain boundaries by Hi-C, indicating that Hi-C and supercoiling domains are not equivalent. Moreover, most plectonemes are ~10-20 kb (Higgins *et al*, 1996), so the ability of highly expressed genes to diminish interactions at longer length scales, as is seen in Hi-C, presumably arises primarily from a different mechanism." Through such statements we have made it clear that there are different types or layers of chromosomal organization. So the assertion that we assume other domains to be insignificant is incorrect and not a fair or accurate reflection of our manuscript. Nevertheless, to ensure that the distinction in our terminology regarding different types of 'domains' is clear to readers of this paper, we have now added a paragraph to the Introduction clarifying the definition of a CID and noting its differences and relationship to supercoil domains and macrodomains. We would also add that in a recent review - Badrinarayanan, Le, Laub 2015 Annual Review of Cell & Dev Bio - we discuss in even more depth the different types of domains present in bacterial chromosomes as a long review allows the space and opportunity for such discussions.

2. One may ask why are only the most actively transcribed genes efficiently captured as crosslinked species in the genome using Hi-C. The authors addressed one issue, which is a restriction enzyme skew. However, they neglect a more obvious explanation, which has nothing to do with domain boundaries. The bacterial chromosome is hard to crosslink compared to eukaryotic DNA because it is more naked than histone-DNA. Crosslinking is more efficient on protein-coated DNA, so RNA polymerases, which are locked on the template as hyper stable ensembles are able to crosslink to each other more efficiently than to regions with very few RNAPs. The crosslinking efficiency argument explains why Rifampicin treatment causes a disappearance of Hi-C regions in bacteria but has almost no effect for Hi-C in eukaryotic chromosomes.

The reviewer may have missed important data in our paper that speak to this issue. In addition to addressing restriction enzyme skew, we show that domains do not arise from higher rates of crosslinking for the DNA at domain boundaries that could result either because RNA polymerase itself promotes higher rates of crosslinking or because the unwound, ssDNA in these regions have higher rates of crosslinking. If either of these crosslinking issues were at play one would expect to see higher numbers of reads involving these regions of the chromosome, which we clearly demonstrate in Fig. S1D is not the case. Additionally, if RNAP were simply increasing the crosslinking efficiency of domain boundaries, as the reviewer suggests, one would expect to see a completely different pattern in the Hi-C heatmaps (demonstrated in Fig. S1E), *i.e.* one would see a dark strip of high interaction values extending from a boundary in both directions. This is not what is observed. Moreover, it should be emphasized that a domain boundary manifests as *diminished* interactions between loci on either side of the boundary, not increased interactions of these flanking loci with boundary-associated loci. In other words, boundaries arise from changes in the interaction frequencies of loci not at the boundary, measurements that do not involve or rely on the crosslinking properties of the boundary loci. Finally, we note that our microscopy studies and recombination-assays, neither of which involve crosslinking, support and substantiate the existence of the domains documented by Hi-C. In sum: a careful and complete consideration of the data and analyses presented in our paper, along with an understanding of how Hi-C works, does not support the alternative hypothesis suggested.

With regard to Hi-C studies in eukaryotes: rifampicin likely has no effect in eukaryotes simply because this antibiotic binds to bacterial RNAP but not to eukaryotic RNA pol II, which is structurally quite different. However, we note that regions of high gene expression likely play a similar role in eukaryotes as we have shown in bacteria; for example, see the recent report Hsieh et al. 2015 Cell.

3) p.6 "Although high gene expression and domain boundaries are clearly correlated, not all highly transcribed genes were associated with domain boundaries, in both growing and starved cells. For example, genes encoding tRNAs are very highlyexpressed in fast-growing cells, but were rarely associated with domain boundaries.
Notably, tRNAs (~76-90 bp) are much shorter than an average bacterial gene (~1000 bp)."
This fits the RNAP crosslinking efficiency hypothesis above.

Please see the response above detailing why RNAP crosslinking efficiency is not consistent with multiple pieces of data and analysis.

*4) p.6 "To test this hypothesis, we first calculated 'rpkpm * transcript length' for each gene or operon in the Caulobacter genome and plotted these values as a function of genome position (Fig 1B-C, Appendix Fig S2, Appendix Table S2)."*

Transcription has two relevant numbers, the RNAP initiation frequency and the elongation rate (NTP/sec) With knowledge of where promoters are, I can see how it is possible to determine regions with long transcripts. However, obtaining the initiation efficiency is not so obvious. Just knowing the abundance of the RNA for a specific gene is a start but there is also the question of half-life. I have no idea of what was done, and anybody who printed out Table 2 in the supplementary data is a sucker. It is incomprehensible and seems to be 60 pages long from just trying to scroll to the bottom.

We agree that rpkpm values can only estimate transcription initiation frequency given that mRNA half-lives have not been globally measured in *Caulobacter*. Nevertheless, given that bacterial mRNAs are generally relatively short-lived, transcript abundance is a reasonable, albeit imperfect, proxy for the frequency of transcription. The correlation observed between 'rpkpm*transcript length' and domain boundaries is, thus, quite striking. We think the text of the paper already makes clear what was done in calculating the 'rpkpm*transcript length' values used in Table S2 (now Dataset EV1), but we have added some additional explanation in the text and Methods section.

With regard to Table S2 (now Dataset EV1): this table is indeed long, but is clearly labeled and comprehensible. And importantly, it includes critical data that allows interested readers to further explore, on their own, the correlations between gene expression and domain boundaries. The table is necessarily long as it includes data for the entire genome, but that is simply the nature of genome-scale studies. And it has been standard practice over the past 10-15 years to include long tables associated with genome-scale studies as supplemental material. Such tables are, of course, often not suitable for printing, but are routinely posted online by journals and they are an important component of many papers. Omitting this table would do a disservice to readers by preventing easy access to a data set that is central to this paper and could be useful for future studies.

5). p7. "(Fig 2). The Hi-C maps indicated that a sharp domain boundary was formed when the terminator cassette was inserted at least 2060 bp from the transcriptional start site with no or very weak domain boundaries formed when the transcript length was engineered to be less than 2060 bp. Taken together, our results indicate that the length of a highly expressed gene plays a critical role in determining whether it creates a chromosomal domain boundary."

They state throughout the manuscript that transcript length is an issue. However, it doesn't seem to be so important. If a 2 kb gene makes a domain boundary that is as clear cut as a 10 kb ATP operon, where's the beef?. Again, the only thing that seems to be important for Hi-C is the high density of RNAP occupying the transcribed track, which is consistent with a crosslinking efficiency effect that involves RNAP.

We are confused by this criticism, which is clearly not shared by the other reviewers and which several pieces of data and analyses in the paper directly address. Figure 2 demonstrates that as we engineer the *rsaA* transcript to be progressively longer, a sharp Hi-C domain boundary emerges only after the transcript is > 2 kb, demonstrating that transcript length is important. Additionally, Figure 1 shows that domain boundaries often coincide with long, highly expressed genes, but not short, highly expressed genes, in two different growth conditions. Collectively, our results clearly demonstrate that transcript length is a key parameter in determining whether a Hi-C domain boundary forms. Perhaps the reviewer's concern is why a 10 kb operon doesn't make a boundary 5X stronger than a 2 kb operon. However, we never make any claims that the strength of boundaries scales linearly with length and it seems plausible that CID boundary strength, as measured by Hi-C, saturates above a certain length. Whether the two parameters, transcription rate and transcript length, contribute equally to boundary formation is unclear, and understanding exactly how they combine to influence domain boundary formation remains a challenge for the future. But the notion that each contributes is clearly substantiated by the data in our paper. Finally, with regard to the point that crosslinking efficiency of RNAP is somehow responsible, we refer the reviewer again to our in-depth response to this issue above.

Supercoil diffusion during transcription.

6) p.9 "To quantify supercoil diffusion, we developed a recombination assay based on the γδ resolvase TnpR (Deng et al, 2005; Higgins et al, 1996) (Fig 5A).....
"Recombination was highly efficient, with TnpR induction, driving a 4-log drop in plating viability in the presence of tetracycline (Fig 5B)....."
then "A promoterless rsaA variant failed to block supercoil diffusion yielding only ~4% tetR colonies....
What caused a 4 orders of magnitude resolution efficiency to drop to under 2 orders of magnitude?

In Figure 5A-B, there is only a *tetAR* cassette (~3 kb) between the two res sites, but in the *rsaA* experiment in Figure 5C, there is a *tetAR* cassette and the *rsaA* gene (combined ~6 kb). Recombination rate is thought to fall off exponentially with distance between res sites (Higgins *et al,* 1996 *J Bact*), likely accounting for the difference noted.

7) p.9. "Transcription based inhibition of supercoil diffusion is likely not simply a consequence of DNA-binding by RNA polymerase between two res sites. Cells harboring 10 consecutive lacO sites in between the res sites and expressing LacI-YFP did not affect recombination (Fig 5D)."

??? Binding of a few lac repressors would not be expected to inhibit supercoil diffusion because each operator/repressor module is flexible and will rotates with the DNA.

This experiment sought to address whether RNAP binding alone may be responsible for, or contribute to, domain boundary formation. LacI-YFP molecules bind to *lacO* sites very strongly (nanomolar affinity; Riggs et al, 1970 *J Mol Bio*) and likely much stronger than RNAP. Additionally, the reviewer's comment is puzzling as our goal was to show that the inhibition of supercoiling that results from high gene expression is not simply due to RNAP binding, but likely due to its ability to unwind the DNA. So we agree that it might be expected that DNA binding proteins alone cannot block supercoil diffusion, but we thought it was important to experimentally verify such an expectation and not assume it.

8) "We conclude that it is likely the DNA unwinding associated with active transcription, rather than protein binding, that drives supercoil diffusion inhibition."

??? This is the same conclusion reached by Booker 1.

We agree and have added a sentence to highlight that this conclusion agrees with that reached in Booker et al. To the best of our knowledge, this is the first time this sort of study has been done in *Caulobacter* supporting the notion that this phenomenon is broadly conserved in bacteria.

9) P. 11 "Taken together, our microscopy and Hi-C analyses indicate that nucleoid expansion does not occur uniformly across the chromosome. The DNA at chromosomal domain boundaries is likely less compacted and therefore stretches more than the DNA inside a domain. These results further support our conclusion that domain boundaries produced by highly expressed and long genes produce a spatial barrier in vivo that diminishes contacts between loci in adjacent domains."

Translation can exerts a large effect on plasmid topology in E. coli The effect involves co-transcriptal translation and insertion of proteins that are membrane associated or transported through the membrane to external locations on the cell surface or periplasm 3. Anchoring the whole transcription-translation to a stable surface exerts forces on the E. coli nucleoid that cause expansion. When cells are treated with chloramphenicol, nucleoids contract 4. This could be relevant for both the rsaA operon (an extruded cell surface component) and the ATP operon, which encodes multiple integral membrane proteins. A quick look at other domain genes in the Caulobacter suggest that membrane proteins may be prominent players. While it would not obliterate the RNAP binding to DNA it may well have an impact on distant markers in each replichore. It might also be an interesting experiment to analyze marked positions flanking the ectopic rsaA operon and see what happens when expression changes from on to off and also from off to on.

We agree that translation and insertion of proteins into the membrane (transertion) can affect plasmid topology and likely chromosome topology. However, transertion remains a very poorly understood phenomenon (see Roggiani and Goulian, 2015 for an in-depth review). More to the point though, transertion simply cannot explain all domain boundaries. Some boundaries are clearly associated with rRNA loci and ribosomal proteins, which are not membrane-bound. And although *rsaA*, the subject of many of our experiments, is a secreted protein, it does not rely on a traditional N-terminal secretion signal and instead relies on a C-terminal signal that likely directs export of the fully translated protein (e.g. see Bingle et al JBact 2000) such that transertion is probably not relevant. In short, we agree that a cursory analysis may suggest that transertion could be relevant, but a more careful, in-depth analysis demonstrates that transertion does not explain domain boundaries.

10) p.15 "Thus, the high rates of transcription at domain boundaries may enable yet more transcription, forming a positive feedback loop that could potentially contribute to the burstiness of gene expression."

"burstiness" ??? I doubt that many readers of EMBO J will understand this term. If the reference is to transcription bursts, an important paper was published last year from Sunney Xie's group and the explanation involves positive supercoiling that builds up to shut gene expression off. They analyzed this phenomenon in vitro and in vivo. 5.

We have modified the wording of this sentence to eliminate the word burstiness.

1 Booker, B. M., et al. Mol Microbiol 78, 1348-1364, (2010).
2 Rovinskiy, N., PLoS Genet 8, e1002845 (2012).
3 Lynch, A. S. & Wang, J. C. J. Bacteriol. 175, 1645-1655 (1993).
4 Van Helvoort, J. M. L. M., Kool, J. & Woldringh, C. L. J. Bacteriol. 178, 4289-4293 (1996).
5 Chong, S., Chen, C., Ge, H. & Xie, X. S. Cell 158, 314-326, doi:10.1016/j.cell.2014.05.038 (2014).
6 Riggs AD, Suzuki H, Bourgeois S. J. Mol. Biol. 1970;48:67–83.

Referee #2:

Previously the authors have used Hi-C methodologies to show that topological domain boundaries in C. crescentus appear to be defined by the presence of highly expressed genes. In this study they again use Hi-C experiments to show that high levels of transcription alone are not sufficient for boundary formation. Rather it appears that a combination of long transcript length and high levels of transcription define chromosomal domain boundary.

I think the paper is interesting and potentially important for the field. However, at present I do not think that the manscript is ready to be published in EMBO journal

My principal concerns are:

1) From Table S5 there do not appear to be any replicates of the Hi-C datasets. Personally I think that it is essential for Hi-C datasets to be replicated. For some of the experiments, I think an argument can be made for only analyzing a single dataset. For example, it could be argued that in figures 2 and 3 that the two datasets with transcripts greater than 2kb and those less than 2kb are basically equivalent. In these instances analysis of the global correlations between the datasets may be acceptable. However, I think that the datasets of starved cells should be replicated - because the changes relative to exponentially growing cells are the core result of the paper. Also, since the differences for interdomain interactions shown in figure 6 appear, to me, to be quite subtle these should also be replicated.

We agree and have now repeated the Hi-C for starved cells. The two repeats are shown in Fig. S1B-C, along with the corresponding directional preference plots, and discussed on p. 5-6 of the revised text. The correlation coefficient for the two independent repeats was 0.98 and the correlation between the directional preference values was 0.82. For Figures 2 and 3: we have now included (also see our response to the next point below) the directional preference plots, which indicate that the individual matrices are extremely reproducible except at the site where the various *rsaA* constructs have been inserted; in the cases where a long, highly expressed transcript is made the directional preference plots indicate a new sharp CID boundary.

For Figure 6, we note that we performed Hi-C (and microscopy analysis) on elongated cells generated by two methods, depletion of *dnaA* and overexpression of CtrAD51EΔ3Ω. These two cases are, of course, not exact replicates, but they effectively represent two independent assessments of how elongation affects Hi-C patterns. We had previously only shown the contact probability plots for the *dnaA* depletion. We now have added the contact probability plots for the CtrAD51EΔ3Ω overexpression strain. The comparison nicely demonstrates that, although the differences in Figure 6 are somewhat subtle, the two independent cases are similar, with short-range interactions becoming increasingly higher in frequency as cells elongate with longer-range interactions decreasing in frequency as cell elongate. We have also included contact probability plots that explicitly separate intra- and inter-domain interactions to further demonstrate that as cells elongate intra-domain interactions increase preferentially relative to inter-domain interactions.

2) In all the experiments using the ectopically positioned rsaA gene the authors definition of a domain boundary appears to be subjective and imprecise. This is puzzling since in fig S2 and fig. 4 the authors do use directional preference of contacts to define where a domain boundary is. Directional preference should be used for all the HiC boundary analysis throughout the paper. This will allow the reader some scale of how big the changes in ectopic boundary formation are between the different conditions used.

We agree and have now modified the figures throughout the paper, particularly those involving the ectopically positioned *rsaA* and its variants, to include directional preference information, enabling a more quantitative and objective assessment of domain boundary formation and comparison of Hi-C experiments. The quantitative values of the directional preference change at the site of *rsaA* insertion are also now referenced directly in the revised text (p. 7 and p. 8).

Other concerns:

3) I understand that the proposition that high levels of transcription inhibit supercoil diffusion in bacteria has been made previously (B. M. Booker et al. Molecular Microbiology. 78, 1348-1364 (2010)). The authors should make clear that their results confirm these observations.

We agree and have added a sentence indicating that our results are in accord with and confirm those presented in Booker et al.

4) In fig.S6C the authors use changes in global contact probability as cells elongate to argue in favour of their hypothesis. I do not think that the global contact probabilities provide relevant information for their hypothesis since boundaries primarily inhibit interdomain contacts < 100kb. To make their point they need to subset the the intra and interdomain contacts < 100kb. Their hypothesis predicts that <100kb inter-domain contacts should be relatively decreased relative to <100kb intra-domain contacts as the cells elongate.

This is an excellent point and we have now prepared contact probability graphs by separating out the intra- and inter-domain interactions. The new graphs (see Appendix Fig. S3E-F) show that the intra-domain interaction frequencies increase relative to the inter-domain interaction frequencies as cells elongate.

Referee #3:

The article by Le and Laub "Transcription rate and transcript length drive chromosomal interaction domain boundary formation" proposes a mechanism (the self-explanatory title) for the formation of boundaries preventing adjacent DNA regions to collide with each other's along the Caulobacter crescentus genome. The data rely primarily on capture of chromosome conformation experiments performed in a variety of growth and mutant conditions. The mechanisms proposed here could presumably apply to other species as well.

The authors start by describing the concomitant changes that occur in the transcriptional pattern of C. crescentus and genomic domain organization in different growth conditions. They analyzed the boundaries between chromosomal domains presenting enrichment in DNA-DNA contacts (CIDs). In an initial report, these borders were found to correlate with the presence of highly expressed genes, suggesting that supercoiling was generating structures preventing contacts between two CIDs. As noted by the authors and others, not all highly expressed genes resulted in a border, nor did all borders consist in such a gene. Here they identify another parameter, gene length, that when combined with a high level of gene expression positively correlate with the presence of a boundary. They propose that this combination explains to a large extent the discrepancies pointed out previously. The paper is clear and reads easily. The experimental setups allowed the authors to test for interesting hypothesis and some of the results are really clear. I nevertheless think that some of the data are a bit interpreted to match the model, and that some of the conclusions are not fully supported by the experiments. These concerns consist essentially in two points. The first one is the match between the observations made in WT cultures presented in Figure 1 and the experimental setup in Figure 2, while the other one relates to the results in Figure 6.

Comparison between high expression levels of long genes and the presence of domain boundaries in different growth conditions are presented in Fig. 1B and C, with three regions emphasized in panels 1D, E, F that present convincing examples of the claim. The observations are interesting, but the authors could improve the quantification of these changes in the text and on the figures. Indeed, when I compare panel 1B and 1C, it is not that obvious that boundaries change along with variation in gene expression in the two experiments (given the gene length does not vary when one compare two *experiments). Either it is due to misalignment of the different panels, but often a change in gene expression does not affect the topological boundaries and reciprocally, new boundaries do not necessary appear along with a higher gene expression level (1B: barrier 1-2, 2-3, 6-7, 10-11, etc. 1C: 3-4, 5-6, 14-15, etc.) Actually, it seems that boundaries that do not correlate with a high expression level in one condition will correlate in the other condition (e.g. barrier 1-2 in rich medium, barrier 5-6 in starvation, 6-7 rich medium, 10-11 rich medium, 18-19 rich medium, etc.) It suggests that these boundaries are positioned independently from the expression level of the (long) gene. Here again it would be informative to indicate on figure 1 the long genes and provide more statistics. How often a boundary does appear concomitantly with the increased expression of a gene? The claim that length of a highly expressed transcript determines the strength of a domain boundary is therefore only partially supported by the data from Figure 1. The authors should describe more precisely the various changes observed in the different growth conditions. Then, they should integrate these observations in light of the nice experiment presented in Figure 2, and maybe ponder the claim they draw from the later experiment. To my eyes, the title is therefore a bit of an overstatement.*

We thank the reviewer for this comment. The panels in Fig. 1 and Fig. S2 were not misaligned, but there was an error on the lines indicating where domain boundaries are. We have now fixed that. Additionally, as suggested by the reviewer, we now calculate and report in the text more statistics to bolster the conclusion that changes in the Hi-C boundaries in different growth conditions are mirrored by changes in gene expression. In particular, we note that there are 12 domain boundaries specific to cells grown in a rich medium - 11 of the 12 coincide with long (> 1 kb) genes that are highly expressed only in rich media and not in starved cells. Conversely, there are 18 boundaries specific to starved cells, with 14 of these 18 coinciding with long genes that are highly expressed only in starved cells. We include on Fig. 1 a notation (orange bars) of where long, highly expressed genes are so that one can compare these to the positions of domain boundaries. Additionally, we have revised former Fig. S2 (now Fig. EV1) to integrate and align the Hi-C data, the plots of 'rpkm*transcript length', and the directional preference plots. This figure helps to capture the notion that condition-specific CID boundaries typically coincide with condition-specific expression of long genes. Dataset EV1 contains the complete data, including exact domain boundary locations and the identities of all genes, transcript lengths, and expression levels in each condition.

The contact maps of Figure 2 should be more focused on the region of interest. Here they span 1Mb, for no obvious reason (?). It would actually be nice to see a resolution slightly lower than 10kb but given the author used BglII, a six-cutter, I assume this would be difficult. The diagonal in the last panel of Figure 2 is narrower than the other panels. It is a detail but this is unfortunate given this is the most important panel. More details about the generation of the contact maps in the MM or SM should be provided here, such as number of contacts for each map, etc. Also, the color scale is missing, so it is not clear whether they are exactly the same for all of these panels. Not that this will change the interpretation, but that will help the reader to focus on the region and results.

We have now added panels to Figures EV2 and EV3 that show magnified and more focused regions of the Hi-C maps shown in Fig. 2 and 3. We agree that it helps the reader to focus on the region of interest and the changes observed there, with the larger regions also capturing the notion that the rest of the Hi-C maps show very little change.

We have also now added more details about the number of valid reads used to create each map and the how each map is normalized such that differences in the maps cannot arise trivially from differences in the number of contacts. Finally, we have added scale bars to all Hi-C data throughout the paper - note that the identical scaling is used throughout the paper to facilitate comparisons between the maps.

The authors demonstrate elegantly that the boundaries do not result from enrichment in proteins, such as ribosomes or DNA polymerase. The controls they made are convincing and provide interesting new insights that will be helpful to decipher the actual mechanisms behind the generation of borders.

I have another concern regarding the "elongation" model proposed by the authors to explain how long genes generate boundaries, relating principally to Figure 6 and S6. First, I could not find in the supplementary figures representative microscopy images of the intra- and inter-domain pairs of loci for the data presented in Figure 6 and S6. Can examples be provided and will these data be made available somewhere?

We have now added example images corresponding to Figure 6 in Figure S3, although given the close proximity of all loci being examined, it should be emphasized that the difference between intra- and inter-domain locus pairs is only clear after quantifying a large number of cells, as was done for the box-plots shown in Fig. 6.

Also, in the synchronization experiments to recover elongated cells or the CtrA(D51E)Δ3Ω overexpression cells, the first time point always appears as quite fuzzy, as pointed out by the authors. Then the post-release time points appear quickly more contrasted and quite similar. These experiments could indeed provide information about the nature of the domains by providing clues about what is happening at their boundaries but as they are presented this is not the case to my eyes. First, I cannot see exactly where in the WT contact map the synchronized pattern matches. The maps are so distorted to improve the visual appearance of the contact signal that it makes comparison tricky. It would be important to include for of each of these time courses a panel with the WT data.

We have now added panels to Figure 6 showing the same portion of the Hi-C maps from WT swarmer cells so that the reader can compare to the Hi-C maps for the *dnaA* depletion strain. Also, we note here, and in the text of the paper, that the maps have not been distorted in any way to affect the visual appearance of the contact signal. The Hi-C maps for each time point taken as cells elongate were identically prepared, normalized, and visualized - the apparent distortion is, in fact, the increasing sharpness of the boundaries in these regions. The inclusion of directional preference plots (Fig. EV4) also now helps emphasize this point.

Second, could an alternative explanation for the increase in sharpness of the border compare to time 0 be that the synchronization procedure is not that harmless? If these Hi-C experiments were performed on a sample containing a few dead/sick cells, maybe the pattern would also look like this? After release the cells would then gradually recover and the entire population converge towards a "sharper" signal. This is where having the WT signal (sync and/or async) would be valuable.

We thank the reviewer for raising this point, which we had not previously considered. If the synchronization process led to a small number of dead/sick cells, it could potentially lead to an increase in non-specific interactions, but intra- and inter-domain interactions would presumably be similarly affected. Alternatively, it could be that synchronization somehow affects global compaction in most cells, leading, for example, to a slightly more compact overall state. Then, as cells elongate, the chromosome may expand, partly because of the decreased confinement of the chromosome, and partly as a response to this synchronization-induced compaction. Although the latter possibility is quite feasible, the key point is that the intra- and inter-domain interactions respond *differently* post-synchronization. As noted above (response to Reviewer 2, point #4), we now include contact probability graphs demonstrating that intra-domain interactions increase more than inter-domain interactions in these time courses. However, we have now, in response to this comment, adjusted the text to indicate that the effect observed could stem from cell elongation *and* a recovery from the synchronization procedure, which may lead to chromosome compaction. We think the recovery from synchronization is less likely to be a factor simply because the changes we observed continue over the course of 3 hours and any effects of synchronization are unlikely to persist that long (*e.g.* gene expression effects associated with synchronization are known - see Laub et al. 2000 - to last less than one generation).

Then, for the region presented in figure 6B, the intra/inter ratio of the interfoci distance *measured in elongating cells also appear quite conserved: 0.65 (time 0) vs. 0.60 (3 hours). I agree that the dispersion of the distances is increased in the inter-domain positions, but the conserved ratio does not point at a more pronounced difference. On the other hand, the other example (presumably involving a border made of a long highly expressed gene? Annotation would be helpful) in Figure 3C* [[We presume the reviewer means 6C here]] *shows the opposite: inter-domain distances are initially smaller than intra-domain distances, then become clearly larger. But the dispersion of distances is quite conserved between the two datasets. Therefore, these experiments suggest that the organization of the domains, as defined by Hi-C, is more complicated and that all visualized domains may not correspond to the same structural properties. Maybe this result from the definition of domain boundaries by using directionality index analysis, which is quite sensitive to local signal on the diagonal? More generally, more quantification of the cell morphology and aspect of these synchronized cultures should be made.*

We have tested two pairs of loci in two different strains (depleting DnaA and overproducing CtrA) - all four experiments are now combined in Fig. 6. In three of the four cases the intra/inter ratio clearly increases over time, with the one exception noted by the reviewer where the ratio remains approximately constant. Additionally in three of the four cases the dispersion is also significantly different for the inter-domain locus pair compared to the intra-domain locus pair. We think these data, taken collectively, support the notion that inter- and intra-domain locus pairs behave differently with interdomain pairs generally showing greater separation and greater variability as cells elongate.

Overall, Figure 6 does not present very conclusive data to my eyes. Another example are the two P(s) curves on Figure S6F, which do not look that different to me: I am not convince that the small discrepancy in short vs. long range contact observed here would withhold a biological duplicate. To support their claim, the authors should provide a clear experiment showing the synchronization does not induce "noise" in the contact data. They should also track additional pairs of loci to reach at a more robust conclusion.

We have now repeated the experiment previously shown in Figure S6F and again see that the slightly smaller cells have slightly increased short-range interaction frequencies. However, this experiment is not critical to the paper and we think it would be prudent to omit it. With regard to tracking additional loci in Figure 6: We have already examined two different pairs of loci in two different strains and see a consistent pattern across these four experiments. Also, the microscopy analysis would be extremely difficult for domains < ~100 kb given given the extremely small size of *Caulobacter* cells and the limited resolution of fluorescence microscopy. The locus pairs used in Figure 6 involve the largest domains that we can also perform microscopy on and reliably measure interfocus differences.

Other comments: citation of existing literature should be improved. There is to my eyes a number of omissions or approximations that have to be fixed. In the introduction: that the authors were the first to perform "Hi-C" on bacteria is clearly an overstatement. Although this was not Hi-C strictly speaking, a very similar analysis was performed by the groups of Shapiro, Dekker and Church (Umbarger et al., 2011), and this article should be cited instead of Le et al 2013 as the first instance of genomic contact analysis in a bacteria. Several interesting findings were present in this manuscript, even though the presence of CID/TADs were not identified at the time.

We have now added a reference to the Umbarger et al. paper although as noted, that work used 5C, not Hi-C, and the resolution was such that domains, the focus of our work, were not seen or documented.

In addition, Marbouty et al. 2015 also noticed that long genes are found at domain boundaries as reported in the Figure 1 of their paper: this should be explicitly stated in the text. Marbouty et al should also be cited along Wang et al. as both papers report *domains with highly expressed genes at their boundaries in B. subtilis (including the long gene borders) and were published concomitantly.*

We now cite Marbouty et al. along with Wang et al. when noting the fact that long genes are also often associated with domain boundaries in *B. subtilis*.

Minor comment: nowhere in the text the name of the restriction enzyme used is mentioned (only on Figure 1, apparently). Even if referring to Le et al 2013 is ok, this is an important information that should be indicated in the MM of the present manuscript as well.

We agree and have added this information to the Hi-C section of the Materials and Methods in the main text and supplement.

2nd Editorial Decision 06 May 2016

Thank you for submitting your revised manuscript for our consideration. It has now been assessed once more by two of the original referees, whose comments are copied below. As you will see, both of them consider the study significantly improved and now in principle suitable for publication in The EMBO Journal. Nevertheless, referee 3 is at this stage still not fully satisfied with all responses to the original reviews, and retains a number of reservations regarding interpretations, which I would kindly ask you to address through an additional round of minor revision. These additional modifications should not require further experiments, but it will be important to carefully respond to the reviewer's points, to more carefully state/qualify various conclusions and assertions throughout the manuscript, as well as to conduct and include the comparison of borders requested by referee 3 (second/main paragraph in their report).

I hope you will be able to add the requested comparisons and modifications as early as possible, and that we should then be able to swiftly proceed with acceptance and publication of the study. Should you have any further questions in this regard, please do not hesitate to get back to me.

REFEREE COMMENTS

Referee #2:

The authors of the manuscript have now dealt satisfactorily with all my previous concerns. I am now happy to recommend that the manuscript be publication in EMBO journal.

One minor point. In the introduction the manuscript states that "TAD boundaries have also been correlated with several other factors, including CTCF and condensin". I assume that the authors actually mean to say "CTCF and cohesin".

Referee #3:

The revised version of Le and Laub article "Transcription rate and transcript length drive formation of chromosomal interaction domain boundaries" addresses several of my original concerns. They have modified the figures accordingly, and overall I am satisfied with the answers. But the revised manuscript still contains statements I disagree with and, also, raises a couple of issues. Notably, a tendency to overstate the importance and/or to oversimplify the nature of the boundaries associated with the CID domains observed in chromosome contact maps of Caulobacter. I advise the authors to be more careful with their assertions. For instance the last sentence of the introduction is clearly an overstatement. The work presented in this paper and the former does not reveal the mechanisms driving the formation of chromosomal domains in general. It does support a mechanism for the formation of boundaries in a bacterial chromosome contact map. There are others ways to generate such signal, including the possibility that some of the domains may be generated by the Hi-C protocol itself (protein occupancy, etc.) This is important, as some microbiologists now appear to take for granted the existence and mechanisms of formation of CIDs describe by the authors.

Regarding boundaries (Figure 1; one boundaries in panel 1C between domains 10 and 11 was modified): overall, although they now include a couple of statistics, my comment and interrogations remain valid. It may relates to the fact that the authors rely extensively on the directionality index analysis to define their domains. Pretty much all the variations between domains boundaries present between the 2 maps shown in Figure 1 seem really close to the statistical threshold chosen. Most of the extra boundaries found in the chromosomes of the starved cells can still be observed in the rich medium maps and DI green/red stripes, but did not pass the statistical test. I would like to be sure that the loss in statistical relevance is not related to the higher background clearly visible in the rich medium conditions. This is a bit puzzling. Is there an explanation for the sharper signal in the starved conditions?
Interestingly, the starved condition duplicates shown in FigS1C, even though globally conserved, also seem to exhibit differences in the DI at several of the positions that also change between rich and starved cells. Those with p values really close to the threshold chosen. The authors should take this into account when they make broad statements. In addition, they compare duplicate 1 and duplicate 2 only through correlation analysis. For the global maps, of course those are very high (0.98) because of the strong signal along the diagonal: it does not mean anything. But the correlation of the two DI signal is "only" of 0.82, which may not be very high for such data. Therefore the authors must compare finely the borders of the starved replicates if they want to be fair with their comparison between rich and starved cells. And the must also do a fine comparison of borders for each replicate of the 2 conditions with those of the other condition. They should also compute the DI correlation between all datasets. I am not sure they will be that different, and this is an essential analysis to be done. Otherwise what is the point to do a replicate?

In the same vein, I am also puzzled by the panels 1D, 1E. During starvation the rpkpm drops sharply but nevertheless the boundaries appear much sharper. In panel 1D the authors point at a difference in the contact map but when one refers to the DI above, the boundary is significant in both cases. Whether more transcripts are found, or not. Same for panel 1F (which is actually the same region than panel 1E; it would be good to indicate where these panels are coming from with their genomic coordinates), boundary using DI is the same in both maps.

These few comments illustrate why they should ponder some of their assertions, and really be more precise and quantitative about what they are really looking at. I think that at the moment they include within the analysis valid boundaries but also a lot of others that I cannot differentiate confidently from noise... And I would really appreciate to see clear comparisons between duplicates and between each starved replicate and the rich medium experiment. That will reinforce their point by discriminating between noise and real changes, alleviating the readers concerns.

Also, regarding their answer to the concern of referee 1 about crosslinking efficiency with respect to protein occupancy, I have several comments. Although I am ready to consider their transcription mediated hypothesis for some boundaries, I have strong objections to the arguments raised in the response. First of all, a DNA fragment has only two extremities. There is no particular reasons to assess (and I am not aware of published data demonstrating it convincingly or even addressing the question seriously) that there would be an increased coverage if this fragment was more frequently crosslinked. This would depend on its size, of its neighbor's sizes, of their protein occupancy, etc. Second, Figure S1E is pure fantasy to me and should be removed. The authors kind of ignore all the factors that are at play during a HiC experiment. For instance, how do they know that more crosslinked region is actually not more constrained and less likely to collide with distant regions? Also, a heavily crosslinked region made of very long (or very short) restriction fragments may be almost invisible in the experiment (due to restriction pattern biases, especially using a 6 cutter). HiC experiments are sometimes tricky to interpret and it is essential to remain open to any possibility. These drawings are not only oversimplifying but wrong, and are likely to be very misleading.

FInally, the image processing point has to be clarified. Maybe we do not look at the same thing when I mentioned "distorsion" of the contact maps in panels Figure 1B and Figure 4C. But to me, the fact that the contact maps do not present 45 degrees angles at the extremities (and between domains) fits my definition of distorsion. I agree this may be a bit too general and may suggest a more complex treatment, so maybe vertical rescaling would be a better term. There is (as in the 2013 publication) important vertical rescaling of the maps. I understand the will to make the domains boundaries look clearer, but want to be sure this is an ok practice, because it does not help the reader to identify the region on the genomewide map.

Regarding citations, the author could cite Val et al., for the domains in Vibrio cholera in Science Advance. Hsieh et al is cited twice. And finally I did not see reference to long gene in the intro (on the opposite to what is stated).

2nd Revision - authors' response 16 May 2016

We thank Referees 2 and 3 for their thoughtful comments on the initial manuscript and the revised version. Below we respond to their latest comments and note additional changes made to the manuscript.

Referee #2:

The authors of the manuscript have now dealt satisfactorily with all my previous concerns. I am now happy to recommend that the manuscript be publication in EMBO journal.

One minor point. In the introduction the manuscript states that "TAD boundaries have also been correlated with several other factors, including CTCF and condensin". I assume that the authors actually mean to say "CTCF and cohesin".

We thank the reviewer for catching this typo, which has now been corrected.

Referee #3:

The revised version of Le and Laub article "Transcription rate and transcript length drive formation of chromosomal interaction domain boundaries" addresses several of my original concerns. They have modified the figures accordingly, and overall I am satisfied with the answers. But the revised manuscript still contains statements I disagree with and, also, raises a couple of issues. Notably, a tendency to overstate the importance and/or to oversimplify the nature of the boundaries associated with the CID domains observed in chromosome contact maps of Caulobacter. I advise the authors to be more careful with their assertions. For instance the last sentence of the introduction is clearly an overstatement. The work presented in this paper and the former does not reveal the mechanisms driving the formation of chromosomal domains in general. It does support a mechanism for the formation of boundaries in a bacterial chromosome contact map. There are others ways to generate such signal, including the possibility that some of the domains may be generated by the Hi-C protocol itself (protein occupancy, etc.) This is important, as some microbiologists now appear to take for granted the existence and mechanisms of formation of CIDs describe by the authors.

We agree and have ensured that the text now indicates throughout that long, highly expressed genes represent a *primary* mechanism for chromosomal interaction domain boundary formation in bacteria, but may not be the only mechanism. We would also underscore that the Discussion already included a section noting that there could be other mechanisms for creating boundaries, although no strong evidence in favor of such mechanisms exists yet. We also note here that if the Hi-C protocol itself were creating boundaries, those boundaries should exist in all Hi-C experiments, but that is not the case. For example, rifampicin and novobiocin treated cells show virtually no domains by Hi-C. So, while we are certainly open to the reviewer's suggestion that CID boundaries could arise from sources other than highly expressed genes, there is simply no evidence that the Hi-C procedure itself introduces domains. Stably bound proteins seem like a plausible explanation for some boundaries in bacteria, but strong evidence in favor of this mechanism has not been reported yet.

Regarding boundaries (Figure 1; one boundaries in panel 1C between domains 10 and 11 was modified): overall, although they now include a couple of statistics, my comment and interrogations remain valid. It may relates to the fact that the authors rely extensively on the directionality index analysis to define their domains. Pretty much all the variations between domains boundaries present between the 2 maps shown in Figure 1 seem really close to the statistical threshold chosen. Most of *the extra boundaries found in the chromosomes of the starved cells can still be observed in the rich medium maps and DI green/red stripes, but did not pass the statistical test. I would like to be sure that the loss in statistical relevance is not related to the higher background clearly visible in the rich medium conditions. This is a bit puzzling. Is there an explanation for the sharper signal in the starved conditions?*

We are confused by this comment. The domain boundaries specific to each growth condition are, in fact, quite striking in most cases and not close to the threshold. And in most cases, these boundaries coincide with changes in gene expression, as is well documented in the manuscript. Moreover, the new repeat of the starvation experiment supports the notion that the changes in boundaries are reproducible (see our response to the next comment as well). It's also important to remember that genes do not completely turn ON or OFF and there may be some vestiges of expression that lead to the persistence of a weaker boundary at a given location even if the gene expression level drops. For instance, Fig. 1D shows a domain that is clearly stronger in rich media than in starvation conditions. This region coincides with one of the rRNA loci, which is more highly expressed in rich media but presumably still expressed at some level in starvation conditions. Consistently, the directional preference values show an abrupt change at this boundary location from -17 to +20 (a net change of 37) in rich media but a change only from -2 to 6 (a net change of 8) in starvation experiment #1 and from -1 to 5 (net change of 6) in starvation experiment #2. Similar trends hold at the other condition-specific boundaries (in Fig. 1E-F and Dataset EV1 and EV2), supporting our conclusion that long, highly expressed genes play a crucial role in determining the positions of most chromosomal interaction domain boundaries. Finally, we have now added the directional preference values for the rich media and starvation experiments as Dataset EV1.

Interestingly, the starved condition duplicates shown in FigS1C, even though globally conserved, also seem to exhibit differences in the DI at several of the positions that also change between rich and starved cells. Those with p values really close to the threshold chosen. The authors should take this into account when they make broad statements. In addition, they compare duplicate 1 and duplicate 2 only through correlation analysis. For the global maps, of course those are very high (0.98) because of the strong signal along the diagonal: it does not mean anything. But the correlation of the two DI signal is "only" of 0.82, which may not be very high for such data. Therefore the authors must compare finely the borders of the starved replicates if they want to be fair with their comparison between rich and starved cells. And the must also do a fine comparison of borders for each replicate of the 2 conditions with those of the other condition. They should also compute the DI correlation between all datasets. I am not sure they will be that different, and this is an essential analysis to be done. Otherwise what is the point to do a replicate?

We now include the complete set of directional preference values for growing and starved cells along with the correlations between directional preference values for all data sets - see new Dataset EV1 - with the associated text on p. 5-6 of the manuscript. The correlation for directional preference values is 0.82 for the starvation repeats with the correlation between the rich media and starvation condition being only 0.34 and 0.4 for the two repeats. So, although the reviewer was not expecting a difference for reasons that are unclear to us, there is a substantial difference. Moreover, the two independent repeats produced 29 and 26 boundaries, respectively, with 25 in common. These comparisons of the two repeats are now discussed on p. 6 of the text. Additionally, to enable readers to perform their own "fine comparisons" we now include the directional preference values for the rich media and two starvation repeats in Dataset EV1 - inspection of this table demonstrates that all starvation-specific boundaries are reproducibly seen in each repeat and not in rich media conditions. We would also remind the reviewer, as noted in the response above, that some genes responsible for condition-specific boundaries are not fully "OFF" in the other condition and so can, in a few cases, still yield weak signatures of a boundary below the thresholds used to call boundaries. This is not noise in the data but likely another reflection of how long, highly expressed transcript impact Hi-C maps and chromosome organization.

In the same vein, I am also puzzled by the panels 1D, 1E. During starvation the rpkpm drops sharply but nevertheless the boundaries appear much sharper. In panel 1D the authors point at a difference in the contact map but when one refers to the DI above, the boundary is significant in both cases. Whether more transcripts are found, or not. Same for panel 1F (which is actually the same region than panel 1E; it would be good to indicate where these panels are coming from with their genomic coordinates), boundary using DI is the same in both maps.

We are confused by this comment. The examples in panels 1D-1G show domain boundaries, marked with a dashed line, that are clearly condition specific and associated with changes in the expression of long genes/operons. For instance, panel 1D shows a rRNA locus, whose expression drops dramatically in starvation. In accord with this change, one sees an unambiguous change in the Hi-C map at this locus with the dashed line coinciding with a clear boundary between two domains/triangles in rich media and then coinciding with the middle of a domain/triangle in starvation conditions. This change is also evident in the directional preference plots, both in Fig. 1B-C and in the revised Fig. 1 where we include the relevant region of the directional preference plot in Fig. 1D-1G. Note, that in the latest version of Fig. 1 we fixed an error in which panel 1D showed an rRNA locus but was labeled as corresponding to a ribosomal protein locus. We also then added the ribosomal protein locus as Fig. 1E and shifted the former panels 1E-F to become 1F-G.

These few comments illustrate why they should ponder some of their assertions, and really be more precise and quantitative about what they are really looking at. I think that at the moment they include within the analysis valid boundaries but also a lot of others that I cannot differentiate confidently from noise... And I would really appreciate to see clear comparisons between duplicates and between each starved replicate and the rich medium experiment. That will reinforce their point by discriminating between noise and real changes, alleviating the readers concerns.

As noted above, we now include Dataset EV1 which includes all of the directional preference values and correlation analysis - together these analyses confirm that the changes documented when comparing rich media to starvation are real changes, not noise.

Also, regarding their answer to the concern of referee 1 about crosslinking efficiency with respect to protein occupancy, I have several comments. Although I am ready to consider their transcription mediated hypothesis for some boundaries, I have strong objections to the arguments raised in the response. First of all, a DNA fragment has only two extremities. There is no particular reasons to assess (and I am not aware of published data demonstrating it convincingly or even addressing the question seriously) that there would be an increased coverage if this fragment was more frequently crosslinked. This would depend on its size, of its neighbor's sizes, of their protein occupancy, etc. Second, Figure S1E is pure fantasy to me and should be removed. The authors kind of ignore all the factors that are at play during a HiC experiment. For instance, how do they know that more crosslinked region is actually not more constrained and less likely to collide with distant regions? Also, a heavily crosslinked region made of very long (or very short) restriction fragments may be almost invisible in the experiment (due to restriction pattern biases, especially using a 6 cutter). HiC experiments are sometimes tricky to interpret and it is essential to remain open to any possibility. These drawings are not only oversimplifying but wrong, and are likely to be very misleading.

We are happy to remove Figure S1E - it was included primarily to address reviewer 1 and the issue of whether cross-linking efficiencies are higher in certain regions of the genome. That reviewer argued that highly expressed genes would be more highly crosslinked. We agree that it's difficult to know with certainty what the consequent Hi-C pattern would be, but, as in ChIP experiments, higher crosslinking would make certain fragments appear more frequently in the sequencing, and Figure S1D shows that simply isn't the case.

Finally, the image processing point has to be clarified. Maybe we do not look at the same thing when I mentioned "distorsion" of the contact maps in panels Figure 1B and Figure 4C. But to me, the fact that the contact maps do not present 45 degrees angles at the extremities (and between domains) fits my definition of distorsion. I agree this may be a bit too general and may suggest a more complex treatment, so maybe vertical rescaling would be a better term. There is (as in the 2013 publication) important vertical rescaling of the maps. I understand the will to make the domains boundaries look clearer, but want to be sure this is an ok practice, because it does not help the reader to identify the region on the genomewide map.

Ah, now we understand and apologize for not understanding this point in the initial review. The contact maps in 1B-1C and 4B are not "processed" or "distorted" in any way. But when the diagonal region of the contact map was rotated 45 degrees, we also stretched the images to span an entire page and simply stretched the images differently in the x and y dimensions. In particular we stretched the images more in the y dimension to facilitate visualization of domain boundaries. This doesn't affect the data in any way, but we now note this adjustment in the figure legends to avoid any additional confusion.

Regarding citations, the author could cite Val et al., for the domains in Vibrio cholera in Science Advance. Hsieh et al is cited twice. And finally I did not see reference to long gene in the intro (on the opposite to what is stated).

Val et al. 2016, which was just published, is now cited and we removed one Hsieh citation.

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: Michael Laub
Journal Submitted to: EMBO Journal
Manuscript Number: EMBOJ-2015-93561

Reporting Checklist For Life Sciences Articles (Rev. July 2015)

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.

Please ensure that the answers to the following questions are reported in the manuscript itself. We encourage you to include a specific subsection in the methods section for statistics, reagents, animal models and human subjects.

In the pink boxes below, provide the page number(s) of the manuscript draft or figure legend(s) where the information can be located. Every question should be answered. If the question is not relevant to your research, please write NA (non applicable).

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-tum

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/

B- Statistics and general methods

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

1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size?	p. 18 and Appendix p. 3
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?	NA
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?	yes, Fig. 6 legend and p. 18, Fig. 5 and its legend
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	yes
Is there an estimate of variation within each group of data?	yes, Fig. 6
Is the variance similar between the groups that are being statistically compared?	yes

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 accession codes for deposited data. See author guidelines, under '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	GEO, GSE74364
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).	Table S1
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. As far as possible, primary and referenced data should be formally cited in a Data Availability section. Please state whether you have included this section. Examples: **Primary Data** Wetmore KM, Deutschbauer AM, Price MN, Arkin AP (2012). Comparison of gene expression and mutant fitness in Shewanella oneidensis MR-1. Gene Expression Omnibus GSE39462 **Referenced Data** Huang J, Brown AF, Lei M (2012). Crystal structure of the TRBD domain of TERT and the CR4/5 of TR. Protein Data Bank 4O26 AP-MS analysis of human histone deacetylase interactions in CEM-T cells (2013). PRIDE PXD000208	p. 16
22. 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

23. 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