THE
EMBO
JOURNAL

Manuscript EMBO-2011-78248

Dynamic long-range chromatin interactions control *Myb* proto-oncogene transcription during erythroid development

Ralph Stadhouders, Supat Thongjuea, Charlotte Andrieu-Soler, Robert-Jan Palstra , Jan Christian Bryne, Anita van den Heuvel, Mary Stevens, Ernie de Boer, Christel Kockx, Antoine van der Sloot, Mirjam van den Hout, Wilfred van IJcken, Dirk Eick, Boris Lenhard, Frank Grosveld and Eric Soler

Corresponding author: Eric Soler, Erasmus MC

Review timeline:

Submission date:	23 May 2011
Editorial Decision:	05 July 2011
Revision received:	02 November 2011
Editorial Decision:	16 November 2011
Accepted:	16 November 2011

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 05 July 2011

Thank you for submitting your manuscript for consideration by The EMBO Journal. It has been now been evaluated by three referees and I enclose their reports below.
As you will see from their comments they provide mixed recommendations with referees #2 and #3 being more positive than referee #1. While both referee #2 and #3 find that describing the control of the Myb locus in interesting referee #1 raises significant concerns regarding the conclusions of the study, especially the role for CTCF mediated looping in regulating polymerase elongation, and lack of colocalisation with CTCF (this is also an issue raised by referee #3) and actively elongating polymerase. After discussing these issues with the Chief Editor we require that these points are specifically experimentally addressed and made more compelling. It is important that the issues raised by referee #1 and #2 are addressed, together with the SNP discussion requested by referee #3. It is also important to note that if these issues cannot be resolved to the satisfaction of the referees and the journal's editorial board, we will not be able to proceed with publication. Nevertheless, given the interest in the study, should you be able to address the concerns we would be happy to consider a revised version of the manuscript for publication in The EMBO Journal.

I should remind you that it is EMBO Journal policy to allow a single round of revision only and that, therefore, acceptance or rejection of the manuscript will depend on the completeness of your responses included in the next, final version of the manuscript. When you submit a revised version to the EMBO Journal, please make sure you upload a letter of response to the referees' comments. Please note that when preparing your letter of response to the referees' comments that this will form part of the Review Process File, and will therefore be available online to the community. For more details on our Transparent Editorial Process initiative, please visit our website:

http://www.nature.com/emboj/about/process.html

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

Yours sincerely,

Editor
The EMBO Journal

REFEREE COMMENTS

Referee #1:

Myb is an important regulator of hematopoietic transcription. This study examines the regulation of the Myb locus by analyzing prior ChIP seq data for the nuclear factors Ldb1, GATA1 and Tal1. In combination with transient transfection assays this led to the characterization of Myb distal regulatory elements with enhancer function. Chromatin conformation assays were carried out to demonstrate that some of these elements interact to form a chromatin hub that also involves the Myb promoter and a region in the first intron. When erythroid cells differentiate, Myb is downregulated which is associated with a loss of hub formation. Finally, reduction of Ldb1 impairs long range interactions consistent with previous studies.
The understanding of how Myb expression is controlled is important and the work presented here adds to some new details to this. However, I do find that this paper is really quite descriptive. The data linking transcription factors to looping and gene control are at best correlative which is disappointing in light if the promising title of the paper. Thus, this report lacks the kind of conceptual innovation that is required for publication in a highly ranked journal such as EMBO.

Specific points

1) The authors play up a potentially interesting finding that an intronic CTCF binding site that is involved in looping marks a transition from transcription initiation to elongation. However, it seems that the CTCF site does not co-localize with the changes in H3K36me3 and Ser5 polII. It seems ~ 2kb away, which argues against the model. The bold claim by the authors that transcription elongation factors are brought to the elongation initiation site via looping remains entirely unproven, even doubtful in the absence of any functional studies. A lot more work would be required to demonstrate a role for CTCF mediated looping and control of transcriptional elongation.

2) The role of Ldb in chromatin looping needs to be presented with a more critical view on the actual data. Loss of Ldb1 impacts on all interactions not just those bound by Ldb1. This suggests that there are indirect effects perhaps resulting from a more general impact of Ldb1 loss on cell differentiation and gene expression.
Work by others has already shown that Ldb1 loss affects the expression and chromatin binding of other factors including GATA1.

3) The authors claim that 3C and ChIP data correlate but ignore important data where these correlations break down. There are several examples of this, most notably at -61kb where transcription factor binding is lost upon Myb repression but the 3C interaction is maintained. In contrast, at -36kb the 3C interaction is lost but Tal1, GATA1, and Ldb1 remain. The authors should be more critical in the interpretation of their findings.

Referee #2:

This is an extremely interesting mannuscript. It involves the use of very difficult state-of-the-art genomics techniques including 3C-seq and ChIP-seq to prove the existence of another interesting

active chromatin hub (ACH) in erythroid progenitor cells. In this case (in comparison to the b-globin locus), the myb hub, which consists of the key erythroid transcription factors - the LDB1 complext and KLF1/EKLF, is formed in progenitor cells to drive c-myb expression and is then dismantled upon differentiation. The authors also show a very nice interaction between enhancrs and a CTCF-bound site in the first intron of myb is likely to be involved in overcoming a transcriptional elongation block. The findings in this paper provide new insights into gene regulation by distal regulatory elements...providing a novel way to think about enhancer function.
I have a few comments which should be addressed before publication. Most importantly, I do not have a clear understanding from the text about how 'true' interactions were defined in the 3C-seq assays. For example, in Figure 2 it is not clear to me how a decision was made to colour some of the Hind3 fragments as grey interactors (i.e. 'real') and some as black (i.e. not real). There is obviously not a lot of difference in signal intensity is some cases to my eyes. Was this decided on a fold change relative to fetal brain, or other statistical methods? The methodology needs more explanation as it is fundamental to the interpretation of the 3C-seq data.
Also, how was the mapping to the genome done? Does the data in Figure 2, Figure 4, etc only show sequence tags in which some component of sequence was derived from the viewpoint and some from the test H3 fragment? I presume to but this would not be easy to map using off the shelf mapping programs.
Why does there not appear to be any specific interaction between the -81 Hind3 fragment and the myb promoter fragment when the -81 region was used as the viewpoint - bottom section of Figure 2b?

Minor suggestions include:
page 11, lines 4-5. The reference to Tallack et al 2010 should be included here also.
page 14, line 16. E2f4 is also misregulated in Klf1-/- fetal liver cells and is likely to contribute to the cell cycle defect. Both e2f2 and e2f4 have intronic enhancers which are very strongly bound by KLF1...and which may function like the myb enhancer described herein. Although Pilon was the first to show e2f2 is KLf1-dependent, Tallack et al found the key KLF1-dependent enhancers for these two genes. Tallack et al, JBC 2009 should be refenced here.
In short I think this paper will make a very important contribution to the field of erythropoiesis and also to the general mechanisms of gene regulation by distant enhancers.

Referee #3:

This manuscript examines in detail how transcription of the Myb proto-oncogene is regulated via the interaction between long-range upstream enhancer-like elements and local cis-acting sequences at the promoter and first intron of the gene. The authors demonstrate that the upstream sequences are active elements using a full range of experimental criteria. They demonstrate that these elements bind the relevant transcription factors, bridging factors and co-factors in vivo. Using chromosome conformation capture, they show that these elements interact with the promoter of the Myb gene and a CTCF binding site in its first intron. They make a convincing case that the upstream elements act in early erythroid precursors to activate Myb and control proliferation. These interactions are abrogated in later cells as differentiation proceeds, in keeping with previous work on Myb expression during erythropoiesis. They provide strong circumstantial evidence that the upstream elements control Myb expression via an effect on transcriptional elongation.

Overall, I think this manuscript provides and excellent, and very well presented, example of how to rigorously dissect the role of long-range elements in controlling gene expression. While much of what they find is not unexpected, relatively few genes have been carefully examined in this way. Furthermore the evidence that the upstream elements might control transcriptional elongation provides to my knowledge only the second convincing example of this. The role of CTCF in this process is new but not addressed in detail. They also add to the idea that Ldb1 may act as a bridging protein.

Clearly all of these points are of great general interest. Of more specific interest is the role of polymorphisms in the Myb-Hbs1l locus on erythropoiesis and the expression of gamma globin. This is of considerable current interest and was mentioned in the introduction but they do not return to it in the discussion. It would be important to include a brief discussion of where and how these

polymorphisms might effect expression of the Myb gene. This is clearly a key issue to discuss since presumably this was a major impetus for understanding how the Myb gene is regulated.

1st Revision - authors' response 02 November 2011

Answers to referee#1's comments:

"1) The authors play up a potentially interesting finding that an intronic CTCF binding site that is involved in looping marks a transition from transcription initiation to elongation. However, it seems that the CTCF site does not co-localize with the changes in H3K36me3 and Ser5 polII. It seems ~ 2kb away, which argues against the model."

We agree with referee#1 that the ChIP-qPCR data do not have sufficient resolution to very precisely localize the site of transition from initiation to elongation. To address referee#1's comment and more accurately localize the transition site from initiating polII (characterized by Ser5 phosphorylation) to productively elongating polII (characterized by Ser2 phosphorylation and the H3K36me3 mark), we performed Ser5-P polII and H3K36me3 ChIP-Seq experiments in MEL cells. As shown in Figure 3B, Ser5-P polII is detected within the *Myb* first intron starting from the transcription start site up to the intronic CTCF site, suggesting that stalling occurs at or just 5' of the CTCF site. In addition, histone H3K36me3 starts increasing after the intronic CTCF site (Figure 3D), suggesting that elongation starts around the CTCF site. While performing these experiments, a study showing histone H3K36me3 profiles in mouse primary erythroid cells was published (Wong et al, Blood 2011; 118(16):e128-e138), which shows a similar pattern of H3K36me3 starting to peak after the CTCF site (Supplementary Figure 5). In addition, data obtained from the ENCODE Project database exhibits a similar H3K36Me3 pattern in human K562 erythroid cells (Supplementary Figure 5). Combined, these data provide a refined analysis of the initiation to elongation transition, localizing it in the vicinity of the CTCF site. To avoid any confusion in the text, we now refer to the elongation transition site as occurring around the intronic CTCF site in the manuscript. Altogether, these new data show that the CTCF and transition sites are not *"~2kb away"* but close to each other, a finding that therefore does not *"argue against the model"*.

"The bold claim by the authors that transcription elongation factors are brought to the elongation initiation site via looping remains entirely unproven, even doubtful in the absence of any functional studies. A lot more work would be required to demonstrate a role for CTCF mediated looping and control of transcriptional elongation."

We do not agree with referee#1 using the term "*bold claim*" when referring to our model of long-range stimulation of transcription elongation. In our manuscript, we propose a model where the upstream enhancers provide a positive elongation environment at the *Myb* gene first intron based on the following findings:

1- We showed that productive elongation takes place around the CTCF site (see above and Figure 3).

2- When analyzing the binding of the positive elongation factors TIF1γ and CDK9, we saw no evidence of binding at the intronic CTCF site, and low signal at the promoter. In contrast, these factors show strong enrichments at all the upstream enhancer elements (Figure 3E and F), suggesting they are brought in by looping.

3- The upstream enhancers are physically contacting the intron, positioning the bound transcription and elongation factors in direct proximity to the initiation-elongation transition site (Figure 2B).

Taken together, the obvious interpretation of our data is that local RNA polII Ser2 phopshorylation is stimulated by the upstream enhancers via chromatin looping. This is not what we would consider a *"totally unproven bold statement"*, but rather a plausible explanation based on our experimental findings.

We agree with referee#1 that the ultimate demonstration of *"a role for CTCF mediated looping and control of transcriptional elongation"* requires a lot more work. However, the only way to definitely address this point would be to conditionally knock-out the CTCF site. We are working on the generation of such mouse models, but these experiments require much more time and represent a full research project on their own.

We did perform a CTCF knock-down in primary erythroid progenitors (from E13.5 fetal livers) to show that reducing CTCF levels indeed results in a reduction of *Myb* transcription, without any effect on the two neighboring genes *Ahi1* and *Hbs1l* (Supplementary Figure 8). This effect is not caused by a differentiation of the cells, since differentiation markers tested are not significantly affected. This experiment suggests that CTCF is required for efficient *Myb* transcription.

To further support our model that the enhancers stimulate transcription elongation via chromatin looping, we carried out CDK9 inhibition experiments (Figure 4). CDK9 is primarily bound to the upstream regulatory elements (within the LDB1/GATA1 complex). Its inhibition results in a severe drop of elongating (Ser2-P) polymerase and 3' *Myb* transcription, while the initiating (Ser5-P) polymerase and 5' transcription are retained, and without affecting looping. The simplest explanation would therefore be that CDK9 is brought in by looping, as represented in our model (see above and Figure 8).

"2) The role of Ldb in chromatin looping needs to be presented with a more critical view on the actual data. Loss of Ldb1 impacts on all interactions not just those bound by Ldb1. This suggests that there are indirect effects perhaps resulting from a more general impact of Ldb1 loss on cell differentiation and gene expression. Work by others has already shown that Ldb1 loss affects the expression and chromatin binding of other factors including GATA1."

We show in Supplementary Figure 8 that reduced levels of LDB1 do not affect the differentiation markers *Hbb-b1* and *Gypa*, and that the expression levels of master erythroid transcription factors *Gata1* and *Tal1* remain unchanged. This argues against an indirect effect due to cellular differentiation or altered transcription factor expression.

Regarding the role of LDB1 in looping, loss of LDB1 indeed impacts on all interactions,

© European Molecular Biology Organization

not just those bound by LDB1. A likely explanation is that in order to be maintained and stabilized, the chromatin hub requires several if not all the interactions (i.e. the enhancer sites and the CTCF sites).

“3) The authors claim that 3C and ChIP data correlate but ignore important data where these correlations break down. There are several examples of this, most notably at -61kb where transcription factor binding is lost upon Myb repression but the 3C interaction is maintained. In contrast, at -36kb the 3C interaction is lost but Tal1, GATA1, and Ldb1 remain. The authors should be more critical in the interpretation of their findings.”

Interactions within the locus during terminal differentiation were first analyzed using 3C-Seq, which we always back-up with the much more quantitative 3C-qPCR method (Figures 2 and 5). 3C-qPCR shows that the interaction with the -61kb element also decreases (Figure 5C). Concerning the -36kb element, where binding of the complex is largely maintained upon differentiation (although with lower levels of LDB1 and ETO2), the loss of looping may be due to a number of reasons: it may be a general destabilization of the chromatin hub since 4 out of 5 sites show a strong reduction in LDB1-complex components binding. It is also possible that other as yet unknown factors are involved. We apologize for not having made this point more clearly in the original manuscript, and we now included it in the discussion of the revised version.

Answers to Referee#2’s comments:

“I do not have a clear understanding from the text about how 'true' interactions were defined in the 3C-seq assays. For example, in Figure 2 it is not clear to me how a decision was made to colour some of the Hind3 fragments as grey interactors (i.e. 'real') and some as black (i.e. not real). There is obviously not a lot of difference in signal intensity is some cases to my eyes. Was this decided on a fold change relative to fetal brain, or other statistical methods? The methodology needs more explanation as it is fundamental to the interpretation of the 3C-seq data.”

We apologize for the confusing representation of the 3C-Seq data. We did not choose to highlight ‘true interactions’ but rather highlighted in grey the fetal liver-specific interactions that overlap with transcription factor and/or CTCF binding. This is now made clear in the figure legends. The statistical method used to define fetal liver-specific interactions has been added to the supplementary methods section and is described below:

Interactions were defined as “true” when they satisfied the following two criteria:

(1) An adjusted p-value ≤ 0.001

(2) A fold change (interaction signal of fetal liver/interaction signal of fetal brain) ≥ 2.

P-values were calculated based on a combination of the Poisson distribution and running mean. In order to obtain p-values for the interactions per restriction fragment from the Poisson distribution, we calculated λ (the mean interaction signal per restriction fragment) for the Poisson probability function. We generated λ as follows: (1) In the same restriction fragment from both fetal liver and

brain, we defined λ(local) as the mean interaction signal from fetal brain and (2) we defined λ(global) by calculating the mean interaction signal for every 100 KB overlapping window chromosome-wide (running mean) from fetal liver and taking the mean of these. The final λ was defined as the maximum value of λ(local) and λ(global). This λ was used in the Poisson probability function described below:

$$p(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

Where x is the interaction signal per restriction fragment, $p(x)$ is the probability of each interaction signal per restriction fragment and λ is λ(final). Finally, p-values were corrected using the Bonferroni correction method.

"Also, how was the mapping to the genome done? Does the data in Figure 2, Figure 4, etc only show sequence tags in which some component of sequence was derived from the viewpoint and some from the test H3 fragment?"

Yes, this is correct.

"I presume to but this would not be easy to map using off the shelf mapping programs. Why does there not appear to be any specific interaction between the -81 Hind3 fragment and the myb promoter fragment when the -81 region was used as the viewpoint - bottom section of Figure 2b?"

The 3C-Seq procedure implies sequencing viewpoint-derived PCR fragments. This means that the resulting sequence tags are a combination of viewpoint-specific primers and sequences derived from the captured interacting fragments. As a whole, these composite sequences are unmappable. In order to achieve efficient mapping to the reference genome, the sequence tags were trimmed to remove the viewpoint sequence, thus leaving only the captured sequence fragments for mapping.

The -81kb 3C-Seq result is an apparent contradiction with the promoter viewpoint result. However it should be noted that a weaker interaction, e.g. the -81kb element with the promoter in this case, may look as a relative strong interaction compared to others when selected as a viewpoint. The tracks are not comparable in quantitative terms and within a track comparisons should be considered semi-quantitative.

"Minor suggestions include:

page 11, lines 4-5. The reference to Tallack et al 2010 should be included here also.
page 14, line 16. E2f4 is also misregulated in Klf1-/- fetal liver cells and is likely to contribute to the cell cycle defect. Both e2f2 and e2f4 have intronic enhancers which are very strongly bound by KLF1...and which may function like the myb enhancer described herein. Although Pilon was the first to show e2f2 is KLf1-dependent, Tallack et al found the key KLF1-dependent enhancers for these two genes. Tallack et al, JBC 2009 should be refenced here."

We agree with Referee#2 that some references needed to be adjusted. These have now been added.

Answers to Referee#3's comments:

"Clearly all of these points are of great general interest. Of more specific interest is the role of polymorphisms in the Myb-Hbs1l locus on erythropoiesis and the expression of gamma globin. This is of considerable current interest and was mentioned in the introduction but they do not return to it in the discussion. It would be important to include a brief discussion of where and how these polymorphisms might effect expression of the Myb gene. This is clearly a key issue to discuss since presumably this was a major impetus for understanding how the Myb gene is regulated."

We thank Referee#3 for his/her positive comments and the interest in our work. Indeed, our study provides a framework for investigating the impact of intergenic SNPs on Myb expression in patients with a high F phenotype (expressing high levels of fetal gamma globin in their adult blood). A correlation between our study in the mouse and the human situation can be drawn, and we have now included this aspect in the discussion although we cannot provide any functional data on this particular matter.

2nd Editorial Decision 16 November 2011

Thank you for submitting your revised manuscript for consideration by The EMBO Journal. It has been seen by two of the original referees who find that you have satisfactorily addressed the issues raised in the initial round of review and both recommend publication of the study. I therefore accept the manuscript for publication in The EMBO Journal. You will receive the official acceptance letter in the next day or so.

Yours sincerely,

Editor
The EMBO Journal

Referee #2

I think this is an outstanding manuscript. I thought the original version was very good but, to their credit, the authors have addressed most of the criticisms of the original reviewers. The reveised manuscript is even stronger for the additional data relating to mechanim of transcriptional elongation through the intronic CTCF site. This paper is certainly at the very cutting edge of technical quality.
I think this paper makes outstanding contributions in a number of areas relating to general mechanims of transcriptional regulation of gene expression as well as in the specific filed of erythropoiesis. Perhaps the most interesting of finding is the relationship between distant enhancers, intronic CTCF sites and transcriptional elongation mechanisms. The approach taken in this paper paper will certainly serve as a paradigm for future experiments addressing gene regulation of any complex locus.

The discussion is very long but it is well written and intersting.

© European Molecular Biology Organization

Referee #3

I have reviewed the revised manuscript. The authors have addressed my point concerning the role of human SNPs playing a role in regulating expression of Myb.

In addition I looked at the responses to Referee #1. I think they have adequately addressed the points made by this referee. In particular they have strengthened the relationship between the transition in transcription and the CTCF binding site. They have also made a reasonably strong and plausible case for the involvement of the upstream elements playing a role in controlling transcriptional elongation.

I have no further comments on the revised manuscript.

© European Molecular Biology Organization