nature research

Peer Review Information

Journal: Nature Genetics
Manuscript Title: Transcription factor competition at the γ-globin promoters controls hemoglobin switching
Corresponding author name(s): Dr. Stuart Orkin

Editorial Notes:

N/A

Reviewer Comments & Decisions:

Decision Letter, initial version:

28th Jul 2020

Dear Dr Orkin,

Your Article, "Transcription factor competition at the γ-globin promoters controls hemoglobin switching" has now been seen by 3 referees. You will see from their comments copied below that while they find your work of considerable potential interest, they have raised substantial concerns that must be addressed. In light of these comments, we cannot accept the manuscript for publication in its current form, but would be very interested in considering a revised version that addresses these concerns.

Overall, referees #1 and #2 are highly supportive of your work, finding it to be both interesting and a productive advance for this field. Their comments are mostly focused on technical/methodological questions which, in our interpretation, are constructive and would improve the manuscript if addressed. We note they both also have more biological questions that would, again, add to the work if they could be answered in a revision.

Conversely, reviewer #3 - while not criticising the technical details - finds this work to be lacking in novelty and mechanistic insight, given what has been previously published. As the other reviewers were supportive, we are prepared to disregard this criticism. Nevertheless, we believe that their detailed comments are pertinent to the manuscript (and, for some points, overlap with the other reviewers' comments e.g. referee #1's questions about other activating factors), and would improve the paper if they could be addressed. Therefore, we would invite you to address those comments as well.

We hope you will find the referees' comments useful as you decide how to proceed. If you wish to submit a substantially revised manuscript, please bear in mind that we will be reluctant to approach the referees again in the absence of major revisions.

To guide the scope of the revisions, the editors discuss the referee reports in detail within the team, including with the chief editor, with a view to identifying key priorities that should be addressed in revision and sometimes overruling referee requests that are deemed beyond the scope of the current study. We hope that you will find the prioritised set of referee points to be useful when revising your study. Please do not hesitate to get in touch if you would like to discuss these issues further.

If you choose to revise your manuscript taking into account all reviewer and editor comments, please highlight all changes in the manuscript text file. At this stage we will need you to upload a copy of the manuscript in MS Word .docx or similar editable format.

We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

If revising your manuscript:

*1) Include a "Response to referees" document detailing, point-by-point, how you addressed each referee comment. If no action was taken to address a point, you must provide a compelling argument. This response will be sent back to the referees along with the revised manuscript.

*2) If you have not done so already please begin to revise your manuscript so that it conforms to our Article format instructions, available here.
Refer also to any guidelines provided in this letter.

*3) Include a revised version of any required Reporting Summary:
https://www.nature.com/documents/nr-reporting-summary.pdf
It will be available to referees (and, potentially, statisticians) to aid in their evaluation if the manuscript goes back for peer review.
A revised checklist is essential for re-review of the paper.

Please be aware of our guidelines on digital image standards.

You may use the link below to submit your revised manuscript and related files:

[REDACTED]

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

If you wish to submit a suitably revised manuscript we would hope to receive it within 6 months. If you cannot send it within this time, please let us know. We will be happy to consider your revision so

long as nothing similar has been accepted for publication at Nature Genetics or published elsewhere. Should your manuscript be substantially delayed without notifying us in advance and your article is eventually published, the received date would be that of the revised, not the original, version.

Please do not hesitate to contact me if you have any questions or would like to discuss the required revisions further.

Nature Genetics is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit please visit www.springernature.com/orcid.

Thank you for the opportunity to review your work, and I hope to hear from you and your co-authors soon.

Sincerely,

Michael Fletcher, PhD
Associate Editor, Nature Genetics

ORCID: 0000-0003-1589-7087

Referee expertise: all three reviewers are experts in haemoglobin gene regulation and/or haematological development.

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:
In this paper Liu and co-authors use Cas9 and dCas9 to screen the beta-globin locus to identify sequences involved in the activation and repression of the fetally expressed gamma-globin genes. Unexpectedly, the dCas9 screen highlights sequences required for the gamma-globin expression close to the binding site of the known repressor BCL11A. The authors hypothesise Cas9 binding at this site prevents the binding of an activating factor due to steric inhibition and motif analysis identifies a CCAAT site and NF-Y as a candidate activator. The authors test this using base editing of the NF-Y binding motif in the HUDEP-2 erythroid cell line and CD34+ve cells and siRNA knockdown of NF-Y. The authors show the CCAAT box is occupied by the activator NF-Y and that this factor appears to compete with the repressor BCL11A, which binds 35 bp away, to maintain silencing or expression of the gamma-globin genes. NF-Y reoccupies the site after BCL11A depletion, neatly explaining why these sites are capable of being bound by NF-Y in primitive but not definitive erythropoiesis, despite the activator being present in both cell types.

I have a few reservations about the way some of the data are presented and I also think the authors' data may point to the presence of more potent activators than NF-Y that are being displaced by the dCas9, which adds further interest. However, this paper over all is extremely well written and together this work represents a substantial advance in a really interesting and important area of research.

Major Comments

The ChIP-seq data in Supplementary Figure 2b appears to be very low depth, are the scales correct or could the authors mention in the text whether this is a product of normalisation?

Is 3C-qPCR the most sensitive method to determine chromatin interactions? In addition the effect of NFYA depletion may be direct or indirect and authors could comment on this.

Could the authors comment on why the CUT and RUN protocol identifies twice as many NF-Y binding sites as the equivalent ChIP-seq experiment and what the degree of overlap is between the techniques.

For Figure 2c and Supplementary Figure 2b there is no scale, making it difficult to properly assess these data.

Figure legends need to be clearer.

Proposed reasons for the discrepancy between the effects dCas9 and Cas9 should be more clearly explained (Figure 1b).

The statistical method used in Figure 4a should be better explained. Ideally a value for BCL11A protein level before KO should be added and used as a comparison to show a fold change over time. The authors could then use the data for the AAVS1 control sgRNA to prove the specificity of the targeting.

The dynamic assessment of BCL11A binding in CD34+ve cells is a really excellent aspect of the work, we wonder whether this could be further improved by using more quantitative technique?

In Figure 5b why does gamma expression reduce when dCas9 is moved further leftwards of the NF-Y binding site? Is it displacing other positive regulators (does the underlying sequence give any clues to this?).

Could the authors explain in more detail the advantages of using the Target-AID-NG method? It is not clear to this reviewer why the combined plasmid produced such low levels of protein. The plasmid does not appear to be over 10kb so I would not have anticipated any problems with transfection efficiency. However, a split Cas9 could be beneficial for others who work with difficult to transfect cell lines.

In Supplementary Figure 5e, we see that the HBG expression level at 72h only gets partially restored when both BCL11A and NFYA are targeted. This interestingly suggests NF-Y is a weak activator and/or works in concert with other factors. Can the authors comment on this? Could NF-Y work as a pioneer

factor (explaining why it is recruited to the site at the early time point/32h) which then recruits other proteins?

Minor Comments

Legend for Fig 4b should be corrected.

Figure 5B would be clearer with a title (perhaps NF-Y CUT and RUN).

The legends for Figure 1a and b should describe the tracks more precisely.

On page 5: "CCAAT box often co-occurs with other TF motifs with precise spatial positioning, suggesting an architectural role in promoter activation". This is very interesting. Is there any example in haematopoietic lineages the authors could use to illustrate this point?

Use of both HUDEP-1 and HUDEP-2 cell lines should be acknowledged.

A clearer annotation of Figure 2e would be appreciated.

In Figure 3b: it would be useful to add "NF-Y motif base edited" on the X axis (as it has been done on Figure 3c).

Can the authors highlight in Figure 3C which clones have been chosen for subsequent work? The authors could use a different shape on the dataset to highlight the HBG expression level of the clones selected.

Page 16: E. coli should be written in italic. Can the name of the bacterial strain be added? The word "transduce" should be changed to "Transduction". The unit "g" (centrifugal units) should be written in italic.

Page 19, paragraph "shRNA knockdown": the phrase "medium were exchanged" should be altered.

Page 20: the word "by" in the sentence "brought up to 1.5 mL by distilled water" should be changed to "with".

Reviewer #2:
Remarks to the Author:
'Transcription factor competition at the gamma-globin promoters controls hemoglobin switching' by Liu et al.

Hemoglobin switching is clinically important as increased levels of fetal hemoglobin (alpha2gamma2) ameliorate the symptoms of beta-hemoglobinopathy patients. Hemoglobin switching also provides a model system for the study of developmentally regulated gene expression. Work of the last ten years has established that expression of the fetal gamma-globin genes (HBG1 and HBG2) is repressed in

adult erythroid cells through the action of transcriptional repressors BCL11A and ZBTB7A. It has recently been shown that these factors bind directly to the HBG gene promoters. It is not known how repression is relieved upon depletion of the repressors. In an elegant series of experiments the authors show that depletion of BCL11A or mutation of the BCL11A binding site in the HBG gene promoters allows binding of the canonical CCAAT-box binding transcriptional activator NF-Y. The authors start by performing Cas9-based screens for regulatory elements in the HBB locus, testing almost 10,000 gRNAs. Their read-out is increased expression of gamma-globin. The screen based on classical Cas9 appears less informative than the screen based on dCas9. The latter reveals that binding of dCas9 to the HBG promoters prevents NF-Y binding and hence blocks activation of the genes. The authors perform a large number of experiments including base editing, CUT&RUN, chromosome conformation capture, and even luciferase assays to substantiate the model that BCL11A (or dCas9) binding prevents interaction of NF-Y with its binding sites in the HBG promoters. This is demonstrated convincingly.

Questions.
1) In the Cas9 screen (Fig 1a top) the region covering HBG2-HBG1 is almost entirely positive for HbF. Given that these genes arose by a recent duplication event, most gRNAs in this region will cut twice which may result in deletions leaving a single HBG2-HBG1 fusion gene. As a fallout of this cut&paste mechanism the fusion gene may become (temporarily) activated. Do the authors have any evidence for this and if so it would be worth mentioning since this is valuable information for scientists performing similar screens.

2) In contrast, the dCas9 screen (Fig 1a bottom) gives much cleaner results. The hypersensitive sites of the locus control region (LCR) super-enhancer and the HBG promoters stand out; dCas9 binding to these areas reduces HBG expression. How do the authors explain that, despite the fact that the screen was based on activation of the HBG genes (by selecting high HbF cells), it is repression that shows superior performance in detecting regulatory elements? This is particularly puzzling since HUDEP2 cells express extremely low levels of gamma-globin.

3) While the results are consistent with a large body of literature, it is regretful that the authors pay no attention to the LCR in the dCas9 screen. The contrast with the Cas9 screen is stunning; the Cas9 screen doesn't reveal any of the LCR hypersensitive sites. This indicates that destroying a single binding site has little impact on LCR function, but binding of dCas9 may -similar to NF-Y binding in the HBG promoters- knock off several neighboring transcription factors. It would be worthwhile to point out critical transcription factor binding sites blocked by dCas9 in HS2 and HS3 in particular. These binding sites are well known (GATA1, KLF1, TAL1, NF-E2 and also BCL11A).

4) Fig 1b, top, Cas9 screen. gRNAs targeting the proximal CCAAT box would be expected to reduce HBG expression. This is not the case. Would deletions leaving a single HBG2-HBG1 fusion gene (see 1)) provide an explanation?

5) Time course experiment displayed in Fig 4. I would conclude that the ATAC peaks mirror NF-Y binding, rather than NF-Y binding preceding ATAC peak formation. Knowing how NF-Y binds to DNA (Fig 2e), this is to be expected.

6) The experiments with altered spacing of the two CCAAT boxes are based on luciferase assays in wildtype and BCL11A knockout mouse erythroleukemia cells (MEL). Although the authors discuss the limitations of reporter assays based on transient transfections, the results are not at all convincing.

For instance, expression of the wildtype promoter construct increases barely 2-fold in BCL11A knockout MEL cells, raising doubts about the validity of this experimental system. These doubts are not taken away by the results obtained with the mutant promoter constructs. The large error bars suggest that the experiments lack statistical power. The authors should either improve the data or take them out altogether.

7) While steric hindrance of NF-Y binding by BCL11A remains to be demonstrated directly, this doesn't affect the main conclusions of the authors or the impact of the current work. Steric hindrance can be put forward as a model, the authors provide reasonable evidence to support this model. However, since BCL11A requires NuRD for repression, it is also possible that NF-Y is evicted by NuRD.

8) The authors are unclear about NF-Y binding. Results section: 'Binding of two NF-Y molecules was impeded as the distance between motifs was reduced to 14 bp'. Discussion section: 'In vitro studies showed that two NF-Y molecules cannot synergistically occupy a promoter if the distance between two CCAAT boxes is less than 27 bp'. Please clarify.

9) Related to this, a major point is whether dCas9 is a neutral molecule. Given the dCas9 activities described by the authors, it would be important to establish whether dCas9 can interact with NuRD or other eukaryotic repressors.

10) Finally, the model (Fig 6) should include NF-Y binding to the HBB promoter. In BCL11A loss of function, the LCR remains free to interact with the HBB promoter. This ties in nicely with classical experiments on the importance of gene order and distance to the LCR for activation of the individual globin genes in the HBB locus.

Reviewer #3:
Remarks to the Author:
Liu and colleagues analyzed the mechanism of how BCL11A represses gamma globin transcription and developed evidence for a model in which BCL11A evicts an activator, NFY, which was implicated previously in activating beta globin genes. Applying state of the art technical approaches added more detail into the existing repression and activation mechanisms. Since the resulting model maintains an essential role of BCL11A as a critical repressor, which has been published in foundational papers, and NFY as an activator, which has been described to be a globin activator, the contribution of the work can be considered to relate to generating additional knowledge on mechanistic steps, rather than developing a new paradigm or surprising mechanistic findings. The work raises additional mechanistic questions, including those stated below.

Specific Comments:

1. In an NFY depletion context, in which gamma expression declines, does this abrogate the cCas9 activity to increase gamma expression? Multiple CCAAT box factors exist and have been implicated in globin transcription. Is NFY really the predominant or sole player in the mechanism?

2. What precludes NFY binding and function through the distal CCAAT box?

3. Are there novel mechanistic aspects of how NFY functions in this globin context, relative to its

established mechanisms in any system or principles of transcription factor function?

4. Can NFY activating function in this system be mimicked with a Cas9 fusion to a generic activating module, or does NFY bring something unique to the locus – something that might introduce novelty with regard to the transcriptional mechanism?

5. The transient transfection/reporter assay in MEL cells contributes nothing to the analysis and may be misleading for the reasons noted in the manuscript. Based on the BCL11A mutation in the transient reporter context, which activated expression, what does this say about the authors' model that BCL11A functions by redirecting chromatin loops? There are likely many mechanistic insights that remain to be determined.

6. Is the distance between BCL11A and NFY sites important at the endogenous locus, and if so, what is the mechanism? Extrapolating from plasmid to chromosome is fraught with problems.

Author Rebuttal to Initial comments

Responses to reviewers' comments:

Reviewer #1:

Remarks to the Author:

In this paper Liu and co-authors use Cas9 and dCas9 to screen the beta-globin locus to identify sequences involved in the activation and repression of the fetally expressed gamma-globin genes. Unexpectedly, the dCas9 screen highlights sequences required for the gamma-globin expression close to the binding site of the known repressor BCL11A. The authors hypothesise Cas9 binding at this site prevents the binding of an activating factor due to steric inhibition and motif analysis identifies a CCAAT site and NF-Y as a candidate activator. The authors test this using base editing of the NF-Y binding motif in the HUDEP-2 erythroid cell line and CD34+ve cells and siRNA knockdown of NF-Y. The authors show the CCAAT box is occupied by the activator NF-Y and that this factor appears to compete with the repressor BCL11A, which binds 35 bp away, to maintain silencing or expression of the gamma-globin genes. NF-Y reoccupies the site after BCL11A depletion, neatly explaining why these sites are capable of being bound by NF-Y in primitive but not definitive erythropoiesis, despite the activator being present in both cell types.

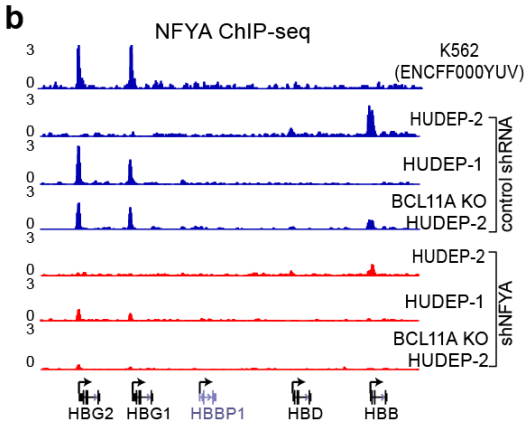
I have a few reservations about the way some of the data are presented and I also think the authors' data may point to the presence of more potent activators than NF-Y that are being displaced by the dCas9, which adds further interest. However, this paper over all is extremely well written and together this work represents a substantial advance in a really interesting and important area of research.

Major Comments

1) The ChIP-seq data in Supplementary Figure 2b appears to be very low depth, are the scales correct or could the authors mention in the text whether this is a product of normalisation?

The signals were normalized based on sequencing depth, using the SPMR (signal per million reads) option of MACS2. Processing ENCODE data (ENCFF000YUV, NFYA in K562) with the same option resulted in similar data scaling. We have added the following details to the methods section, and updated Extended Data Fig. 2b.

"Peaks were called using MACS2 with the following parameters: -g hs -f BAMPE -q 0.01 -B –SPMR."



2) Is 3C-qPCR the most sensitive method to determine chromatin interactions? In addition the effect of NFYA depletion may be direct or indirect and authors could comment on this.

Common techniques to view chromatin interactions for one or a few viewpoints include 3C-qPCR, 4C-seq and Capture-C, among others. The relationship between 3C-qPCR and the other two is similar to that of ChIP-qPCR and ChIP-seq. Whereas 3C-qPCR quantifies one-to-few interactions, 4C-seq and Capture-C identify many more interactions to one or a number of viewpoints in a less biased way. The downside of 4C-seq and Capture-C seq is that the 20-25 cycles of PCR amplification prior to sequencing inevitably introduces bias. Therefore, they are not substantially better in quantifying one-to-one interactions than 3C-qPCR.

We agree that the role of NFYA in chromatin interaction might be indirect. It may act through other looping factors. We added the following sentence to the main text to clarify.

"Whether NF-Y mediates looping directly or through other factors is as yet unknown."

3) Could the authors comment on why the CUT and RUN protocol identifies twice as many NF-Y binding sites as the equivalent ChIP-seq experiment and what the degree of overlap is between the techniques.

CUT&RUN identifies both direct binding sites and genomic contact sites, since a pA-MNase tethered to a TF not only cleaves the TF bound sequences but also cleaves DNA in close proximity. Therefore CUT&RUN also maps chromatin interactions. This has been demonstrated by comparing CTCF CUT&RUN and CTCF ChIA-PET (Skene, eLife 2017, Figure 8). We show in Extended Data Figure 3b that NFYA CUT&RUN identified 59% of NFYA ChIP-seq peaks, similar to what we previously reported for

CTCF and GATA1 CUT&RUN (46% and 51%, Liu, Cell 2018).

4) For Figure 2c and Supplementary Figure 2b there is no scale, making it difficult to properly assess these data.

We have updated the figures to include scales.

5) Figure legends need to be clearer.

We have carefully reviewed and edited the Figure legends.

6) Proposed reasons for the discrepancy between the effects dCas9 and Cas9 should be more clearly explained (Figure 1b).

We have added the following text to explain why the Cas9 screen identified widespread HbF "repressor sequences":

"We suspect that the induction of HbF by targeting these sequences reflects secondary effects of Cas9 editing and subsequent DNA repair. In addition, gRNAs targeting the duplicated HBG1 and HBG2 genes frequently result in 4.9 kb deletions (Traxler, Nat Med 2016), which may lead to remodeling of local chromatin or removal of repressive elements. Authentic cis-acting elements may therefore be obscured."

In contrast, dCas9 acts through interfering with binding of endogenous regulators in a narrow region. We have added the following sentence:

"Therefore, dCas9 screens allow for finer dissection of cis-acting elements in a target genomic region."

7) The statistical method used in Figure 4a should be better explained. Ideally a value for BCL11A protein level before KO should be added and used as a comparison to show a fold change over time. The authors could then use the data for the AAVS1 control sgRNA to prove the specificity of the targeting.

We have updated Figure 4a to include the time point before KO. We also updated the methods section to include the following sentence describing the band quantification method:

"Western blot was used to assess BCL11A levels before and after nucleofection, and the band intensities were quantified using ImageJ. Briefly, the pixel densities of target bands and blank regions were measured. The pixel numbers of blank regions represented background and were subtracted from the band. Three independent measurements were performed for each band, and BCL11A signals were normalized using Histone H3."

8) The dynamic assessment of BCL11A binding in CD34+ve cells is a really excellent aspect of the work, we wonder whether this could be further improved by using more quantitative technique?

We add an Extended Data Figure 5d to quantify PRO-seq counts on each globin gene for Figure 4d, which aids the visualization of the results.

9) In Figure 5b why does gamma expression reduce when dCas9 is moved further leftwards of the NF-Y binding site? Is it displacing other positive regulators (does the underlying sequence give any clues to this?).

We believe the high level of γ-globin expression in -62 and -102 is caused by incomplete eviction of NF-Y by dCas9. When dCas9 is moved further leftwards to -115 or -124, dCas9 can evict NF-Y completely through steric hindrance. This is supported by NF-Y CUT&RUN: weak binding of NF-Y at the γ-globin promoters was detected in cells with dCas9-gRNA(-62/-102), whereas no binding was detected in dCas9-gRNA(-115/-124) cells. Difference in NF-Y displacement activities may be attributed to different binding efficiencies of the gRNAs.

Could the authors explain in more detail the advantages of using the Target-AID-NG method? It is not clear to this reviewer why the combined plasmid produced such low levels of protein. The plasmid does not appear to be over 10kb so I would not have anticipated any problems with transfection efficiency. However, a split Cas9 could be beneficial for others who work with difficult to transfect cell lines.

We have included the sentence below in the methods section to clarify why we engineered the split Target-AID-NG.

"Introduction of the original Target-AID-NG produced protein at very low level (Extended Data Figure 4b) which did not lead to detectable base editing in HUDEP-2 cells. We reasoned that a higher level of base editor protein was needed for successful base editing. To this end, we engineered a split Target-AID-NG."

We do not know the reason why Target-AID-NG produced very little protein. We infer that this was due to poor virus production, transcription, or translation.

In Supplementary Figure 5e, we see that the HBG expression level at 72h only gets partially restored when both BCL11A and NFYA are targeted. This interestingly suggests NF-Y is a weak activator and/or works in concert with other factors. Can the authors comment on this? Could NF-Y work as a pioneer factor (explaining why it is recruited to the site at the early time point/32h) which then recruits other proteins?

It is highly likely that NF-Y works in concert with other transcription factors to activate γ-globin. We added a comment on this in the result section:

"Similar results were obtained in primary human CD34+ cell derived erythroid precursors, though with a lesser degree of γ-globin reduction, possibly due to the action of other transcription activators"

With regard to the pioneer activity of NF-Y, it was shown in embryonic stem cells that NF-Y promotes chromatin accessibility and enhances the binding of other TFs, such as Oct4, Sox2 and Nanog, to their targets (Oldfield, Mol Cell 2014). It is possible that NF-Y activates γ-globin through a similar mechanism. We added a comment on this in the results section.

"These data reveal that NF-Y binds rapidly to the γ-globin promoters upon BCL11A depletion to open up the local chromatin, which precedes formation of enhancer-promoter contacts and transcriptional activation. These findings are consistent with the reported pioneer transcription factor activity of NF-Y"

Minor Comments
Legend for Fig 4b should be corrected.
Addressed.
Figure 5B would be clearer with a title (perhaps NF-Y CUT and RUN).
Updated.
The legends for Figure 1a and b should describe the tracks more precisely.
We have updated the figure legends.
On page 5: "CCAAT box often co-occurs with other TF motifs with precise spatial positioning, suggesting an architectural role in promoter activation". This is very interesting. Is there any example in haematopoietic lineages the authors could use to illustrate this point?
The statement is based on findings reported in reference 23-25. Two of the studies were conducted by analyzing ENCODE ChIP-seq data in K562 and GM12878 cells. In these studies, the binding motifs of several TFs, including cFOS, MAX, etc, were found to be enriched at a precise distance to CCAAT, strongly suggesting cooperative binding in a stereo-specific manner. The exact nature of such cooperativity and why a precise distance is required is interesting and worthy of additional study.

Use of both HUDEP-1 and HUDEP-2 cell lines should be acknowledged.
We have added the following statement in the acknowledgment section:
"We thank Yukio Nakamura for HUDEP-1 and HUDEP-2 cell lines."
A clearer annotation of Figure 2e would be appreciated.
We have updated Figure 2e by labeling each subunit of NF-Y. The legend of Figure 2e is also update with the following information:
"flanking sequences are wrapped around NF-Y through histone-fold domains of NFYB and NFYC. NFYA is responsible for motif recognition."
In Figure 3b: it would be useful to add "NF-Y motif base edited" on the X axis (as *it has been done on Figure 3c).*
Updated.
Can the authors highlight in Figure 3C which clones have been chosen for subsequent work? The authors could use a different shape on the dataset to highlight the HBG expression level of the clones selected.
We have circled the clones in Figure 3c and 3g that were used in Figure 3d and added a description in legends.
Page 16: E. coli should be written in italic. Can the name of the bacterial strain be added? The word "transduce" should be changed to "Transduction". The unit "g" (centrifugal units) should be written in italic.
Corrected. E. coli strain has been added to the method section.
Page 19, paragraph "shRNA knockdown": the phrase "medium were exchanged" should be altered.

Rephrased to "fresh media supplemented with puromycin was added to select transduced cells"
Page 20: the word "by" in the sentence "brought up to 1.5 mL by distilled water" should be changed to "with".
Corrected.

Reviewer #2:
Remarks to the Author:
'Transcription factor competition at the gamma-globin promoters controls hemoglobin switching' by Liu et al.
Hemoglobin switching is clinically important as increased levels of fetal hemoglobin (alpha2gamma2) ameliorate the symptoms of beta-hemoglobinopathy patients. Hemoglobin switching also provides a model system for the study of developmentally regulated gene expression. Work of the last ten years has established that expression of the fetal gamma-globin genes (HBG1 and HBG2) is repressed in adult erythroid cells through the action of transcriptional repressors BCL11A and ZBTB7A. It has recently been shown that these factors bind directly to the HBG gene promoters. It is not known how repression is relieved upon depletion of the repressors. In an elegant series of experiments the authors show that depletion of BCL11A or mutation of the BCL11A binding site in the HBG gene promoters allows binding of the canonical CCAAT-box binding transcriptional activator NF-Y. The authors start by performing Cas9-based screens for regulatory elements in the HBB locus, testing almost 10,000 gRNAs. Their read-out is increased expression of gamma-globin. The screen based on classical Cas9 appears less informative than the screen based on dCas9. The latter reveals that binding of dCas9 to the HBG promoters prevents NF-Y binding and hence blocks activation of the genes. The authors perform a large number of experiments including base editing, CUT&RUN, chromosome conformation capture, and even luciferase assays to substantiate the model that BCL11A (or dCas9) binding prevents interaction of NF-Y with its binding sites in the HBG promoters. This is demonstrated convincingly.

Questions.
1) In the Cas9 screen (Fig 1a top) the region covering HBG2-HBG1 is almost entirely positive for HbF. Given that these genes arose by a recent duplication event, most gRNAs in this region will cut twice which may result in deletions leaving a single HBG2-HBG1 fusion gene. As a fallout of this cut&paste mechanism the fusion gene may become (temporarily) activated. Do the authors have any evidence for this and if so it would be worth mentioning since this is valuable information for scientists performing similar screens.
It is correct that the sequences of HBG1 and HBG2 and their promoters are very similar. Cas9 editing with a gRNA that binds to both genes often leads to large deletions, leaving only one copy of the γ-globin. We observed frequent large deletions upon Cas9

editing. We also suspect that such events lead to activation of γ-globin through unknown secondary effects.
We have added the following text to the result section:
"We suspect that the induction of HbF by targeting these sequences reflects secondary effects of Cas9 editing and subsequent DNA repair. In addition, gRNAs targeting the duplicated HBG1 and HBG2 genes frequently result in 4.9 kb deletions (Traxler, Nat Med 2016), which may lead to remodeling of local chromatin or removal of repressive elements. Authentic cis-acting elements may therefore be obscured."

2) In contrast, the dCas9 screen (Fig 1a bottom) gives much cleaner results. The hypersensitive sites of the locus control region (LCR) super-enhancer and the HBG promoters stand out; dCas9 binding to these areas reduces HBG expression. How do the authors explain that, despite the fact that the screen was based on activation of the HBG genes (by selecting high HbF cells), it is repression that shows superior performance in detecting regulatory elements? This is particularly puzzling since HUDEP2 cells express extremely low levels of gamma-globin.

In the screen a comparison is made between HbF-high and unsorted cells. Expression of HbF is low but non-zero in HUDEP-2 cells, reminiscent of low-level expression of HbF in adult stage human erythroid cells. We speculate that, by blocking activators at the LCR, especially at HS2 and HS3 sites, dCas9 reduces expression of both HBB and HBG1/2 and thus reduces expression of both HbA and HbF. Therefore, cells carrying these gRNAs are relatively depleted from the HbF+ pool as compared to the unsorted population which includes some poorly hemoglobinized cells.

3) While the results are consistent with a large body of literature, it is regretful that the authors pay no attention to the LCR in the dCas9 screen. The contrast with the Cas9 screen is stunning; the Cas9 screen doesn't reveal any of the LCR hypersensitive sites. This indicates that destroying a single binding site has little impact on LCR function, but binding of dCas9 may -similar to NF-Y binding in the HBG promoters- knock off several neighboring transcription factors. It would be worthwhile to point out critical transcription factor binding sites blocked by dCas9 in HS2 and HS3 in particular. These binding sites are well known (GATA1, KLF1, TAL1, NF-E2 and also BCL11A).

The dramatic effect of dCas9 binding at HS2 and HS3 is indeed interesting. The simplest explanation, as the reviewer pointed out, is that dCas9 blocks the binding of multiple key transcription factors at the core enhancer, whereas indels generated by Cas9 may not be perfectly positioned at the core motifs. We have added the following text and Extended data fig 1d to point this out:
"Perturbations of HbF were observed when dCas9 was targeted to HS2, HS3 and the γ-globin promoters. Within HS2 and HS3, the most depleted gRNAs mapped to the core GATA1-TAL1 composite motifs (Extended Data Fig. 1d) , suggesting that eviction of GATA1 and TAL1 proteins at each of the GATA1/TAL1 motifs is sufficient to impair the LCR activity."

4) Fig 1b, top, Cas9 screen. gRNAs targeting the proximal CCAAT box would be

expected to reduce HBG expression. This is not the case. Would deletions leaving a single HBG2-HBG1 fusion gene (see 1)) provide an explanation?
Yes, we believe that the large deletion provides an explanation, as mentioned in the response to Question 1. The other technical reason is that there is no gRNA in the screen that directs SpCas9 cleavage exactly at the proximal CCAAT box (due to lack of NGG PAM availability). Both could confound the analysis.
5) Time course experiment displayed in Fig 4. I would conclude that the ATAC peaks mirror NF-Y binding, rather than NF-Y binding preceding ATAC peak formation. Knowing how NF-Y binds to DNA (Fig 2e), this is to be expected.
We agree that the ATAC peaks showed a slight increase at 32 hr after BCL11A KO, similar to the increase of NF-Y binding. We quantified the changes with MAnorm, which measures fold changes and associated P-values. The increase of ATAC-seq signal was not statistically significant, though we acknowledge the limitation of quantitative methods when dealing with very low signal/noise data. We rephrased the result to reflect the subtle increase in chromatin accessibility.
6) The experiments with altered spacing of the two CCAAT boxes are based on luciferase assays in wildtype and BCL11A knockout mouse erythroleukemia cells (MEL). Although the authors discuss the limitations of reporter assays based on transient transfections, the results are not at all convincing. For instance, expression of the wildtype promoter construct increases barely 2-fold in BCL11A knockout MEL cells, raising doubts about the validity of this experimental system. These doubts are not taken away by the results obtained with the mutant promoter constructs. The large error bars suggest that the experiments lack statistical power. The authors should either improve the data or take them out altogether.
As another reviewer raised similar concerns about luciferase assay, we chose to remove these experiments from the manuscript as they do not significantly add to the existing data.
7) While steric hindrance of NF-Y binding by BCL11A remains to be demonstrated directly, this doesn't affect the main conclusions of the authors or the impact of the current work. Steric hindrance can be put forward as a model, the authors provide reasonable evidence to support this model. However, since BCL11A requires NuRD for repression, it is also possible that NF-Y is evicted by NuRD.
We agree with the comments. NuRD is huge protein complex at a size of 2 MDa. Considering that NF-Y recruits coactivator complexes, the steric effect between these bulky cofactors may also affect their occupancy.

8) The authors are unclear about NF-Y binding. Results section: 'Binding of two NF-Y molecules was impeded as the distance between motifs was reduced to 14 bp'. Discussion section: 'In vitro studies showed that two NF-Y molecules cannot synergistically occupy a promoter if the distance between two CCAAT boxes is less

than 27 bp'. Please clarify.
We have updated the text to make it clearer.
"In vitro studies revealed that two NF-Y molecules exhibit synergistic binding to double CCAAT motifs only when the distance between the two motifs is 27 bp (three helical turns). NF-Y may still occupy double CCAAT motifs without synergy when the distance is reduced by a few basepairs, but co-binding was not detected when the distance is reduced to 17 bp."
9) Related to this, a major point is whether dCas9 is a neutral molecule. Given the dCas9 activities described by the authors, it would be important to establish whether dCas9 can interact with NuRD or other eukaryotic repressors.
Based on our observations and the literature, it is unlikely that dCas9 interacts with repressors.
1) dCas9 exhibited gene activating function when targeted to ~-50 and -200 bp of γ-globin promoter (Fig .1a, b), suggesting that dCas9 function is context dependent, and faithfully reflects the disruption of transcription factor binding. This is further supported by a previous report that used dCas9 to activate Pax6 or Nanog expression by disrupting corresponding repressors (Shariati, Mol Cell 2019).
2) As shown by the newly incorporated data in Fig. 1a, when dCas9 is fused to a repressor (KRAB), it can exert repressive activity across the entire HBG1/2 gene body and even at a distance (Fig. 1a), in contrast to the discrete action range of dCas9.
10) Finally, the model (Fig 6) should include NF-Y binding to the HBB promoter. In BCL11A loss of function, the LCR remains free to interact with the HBB promoter. This ties in nicely with classical experiments on the importance of gene order and distance to the LCR for activation of the individual globin genes in the HBB locus.
We updated Fig. 6 to reflect NF-Y binding at the HBB gene.
We have also included analysis and discussion with regard to NF-Y binding at HBB. Extended Data Fig. 2e show reduced NF-Y binding at the HBB gene after NF-Y depletion. Extended Data Fig. 3g showed that NF-Y directly binds to the first intron of HBB. We have not investigated the role of NF-Y in HBB expression.

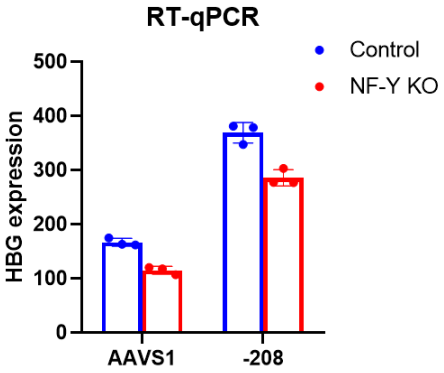
Reviewer #3:
Remarks to the Author:
Liu and colleagues analyzed the mechanism of how BCL11A represses gamma globin transcription and developed evidence for a model in which BCL11A evicts an activator, NFY, which was implicated previously in activating beta globin genes. Applying state of the art technical approaches added more detail into the existing repression and activation mechanisms. Since the resulting model maintains an essential role of BCL11A as a critical repressor, which has been published in foundational papers, and NFY as an activator, which has been described to be a globin activator, the contribution of the work can be considered to relate to generating additional knowledge on mechanistic steps, rather than developing a new paradigm or surprising mechanistic

findings. The work raises additional mechanistic questions, including those stated below.

Specific Comments:

1. In an NFY depletion context, in which gamma expression declines, does this abrogate the cCas9 activity to increase gamma expression? Multiple CCAAT box factors exist and have been implicated in globin transcription. Is NFY really the predominant or sole player in the mechanism?

Question 1 above: To address this, we acutely depleted NF-Y in dCas9-gRNA(-208) HUDEP-2 cells. We observed abrogated γ-globin derepression to a similar extent as that in BCL11A null HUDEP-2 cells, as shown below.



Question 2 above: Besides NF-Y, known factors that may associate with CCAAT box are CDP (CCAAT displacement protein)/CUX1, CTF (CCAAT box-binding transcription factor)/NF-I, and C/EBP (CCAAT enhancer binding protein). We present the following evidence indicating that these factors are not dominant γ-globin activators.

a) Although previous gel shift based in vitro experiments showed that these factors binds to CCAAT, later ChIP-seq have determined different and distinct binding motifs. CDP/CUX1 binds to ATCRAT, C/EBP binds to GCAAT, which is supported by crystal structure (PMID: 30566668, 12578822), and CTF/NF-I preferentially binds to a palindrome TGGCANNNTGCCA motif. None of these consensus sequences matches the γ-globin CCAAT boxes.

b). We knocked out each of these proteins (Acute depletion by CRISPR/Cas9 for 3 days) in BCL11A null HUDEP-2 cells. Proteins that are not successfully depleted with two different gRNAs, or not detected with two different antibodies are not shown. Only NFYA KO displayed significant reduction of γ-globin expression. The small change of γ-globin upon depletion of other factors may be attributed to secondary effects.

Based on the above data, we conclude that γ-globin activation may involve other players, but NF-Y is the predominant factor that acts at the CCAAT box of γ-globin promoter.

2. What precludes NFY binding and function through the distal CCAAT box?
We have two hypotheses but no direct data.

1) *Another transcription activator binds near the distal CCAAT and blocks NF-Y binding to the distal CCAAT.*
2) *Synergistic binding of NF-Y and another transcription factor. Previous research suggested an architectural role of NF-Y in organizing the promoters. CCAAT boxes are typically located around -80 bp of the promoters (the proximal CCAAT is at -88 to -84 bp of the γ-globin promoters), and are often associated with certain TF motifs with a precise distance (Mantovani 1999, Dolfini 2009, 2016, Fleming 2013). Such stereotypical binding may allow synergy for gene activation, and in the meantime, confine the binding of NF-Y to certain CCAAT sequences.*

3. Are there novel mechanistic aspects of how NFY functions in this globin context, relative to its established mechanisms in any system or principles of transcription factor function?
The mechanistic novelty is that NF-Y competes with transcription repressor BCL11A for binding to γ-globin promoters. BCL11A represses γ-globin through sterically excluding NF-Y in adult erythroid cells to initiate silencing of γ-globin transcription. The competitive binding in this 35 bp region encompassing two motifs determines the stage specificity of the entire β-globin gene cluster. We contend this is a novel and important finding as it brings mechanistic clarity to a gene switch that has been under study for more than four decades.

4. Can NFY activating function in this system be mimicked with a Cas9 fusion to a generic activating module, or does NFY bring something unique to the locus – something that might introduce novelty with regard to the transcriptional mechanism?
Yes, we have incorporated dCas9-VP64 and dCas9-KRAB screens in Fig.1 in parallel to Cas9 and dCas9 screens. These data indicate that targeting dCas9-VP64 to the γ-globin gene and even a few kilobases upstream or downstream could activate γ-globin, suggesting that γ-globin can be activated by generic activating factors.
We believe that NF-Y functions as a pioneer transcription factor, and recruits transcriptional coactivators to the promoter.

5. The transient transfection/reporter assay in MEL cells contributes nothing to the analysis and may be misleading for the reasons noted in the manuscript. Based on the BCL11A mutation in the transient reporter context, which activated expression, what does this say about the authors' model that BCL11A functions by redirecting chromatin loops? There are likely many mechanistic insights that remain to be determined.

As another reviewer raised similar concerns about the luciferase assay, we have removed these experiments from the manuscript.

6. Is the distance between BCL11A and NFY sites important at the endogenous locus, and if so, what is the mechanism? Extrapolating from plasmid to chromosome is fraught with problems.

We agree that investigating the effect of motif distance in the endogenous locus could provide more direct confirmation of our model. We have attempted to modify the BCL11A and NF-Y motif distance in vivo using HDR-mediated knock-in. We tested two gRNAs that cleave between the two motifs, and chose the more efficient one to isolate cell clones. We used ssDNA as HDR donor as this method achieves 5-10% HDR efficiency in HDUEP-2 cells. However, after screening >300 clones, we were not able to identify a single clone that meet the following criteria for subsequent analysis.

1) *All four γ-globin alleles are intact. The duplicated HBG1/HBG2 genes have identical core promoters and are separated by 4.9 kb. Any editing at the promoters results in two double strand breaks and frequently leads to genomic inversion or large deletions that remove HBG2. In Fig. 1 and associated text, we discussed that such events may contribute to undesired derepression of γ-globin transcription due to secondary effects.*
2) *Random indels are unfavorable compared to HDR alleles, as they may complicate the interpretation of results. Therefore, we aimed to obtain clones with all HDR alleles, or combinations of HDR and wild-type alleles.*

Given a 5-10% KI efficiency in bulk HUDEP-2 cells, identification of a qualified clone based on the above criteria was not technically feasible. HDR efficiencies need to be increased significantly to generate genotypes for analysis.

Decision Letter, first revision:

14th Dec 2020

Dear Stu,

Firstly, my sincere apologies for the long delay in our decision; the final reviewer submitted their report only in the past few days.

Your Article, "Transcription factor competition at the γ-globin promoters controls hemoglobin switching" has now been seen by the 3 original referees.

You will see from their comments below that while they agree that this study has improved in revision

and appear supportive of publication, there are still a number of outstanding concerns that we would like addressed. We remain interested in the possibility of publishing your study in Nature Genetics, but would like to consider your response to these concerns in the form of a revised manuscript before we make a final decision on publication.

Briefly, referees #2 and #3 approve of your responses to their previous reviews and have no remaining comments. However, referee #1 - while signing off on your responses to most of their comments from the previous round - does make some suggestions for additional details, explanation and discussion in the text.

In our reading of these reviews, we believe that referee #1's remaining points are likely addressable with only textual changes, and we believe that - depending on your response to the points - a further round of review may not be required.

To guide the scope of the revisions, the editors discuss the referee reports in detail within the team, including with the chief editor, with a view to identifying key priorities that should be addressed in revision and sometimes overruling referee requests that are deemed beyond the scope of the current study. We hope that you will find the prioritized set of referee points to be useful when revising your study. Please do not hesitate to get in touch if you would like to discuss these issues further.

We therefore invite you to revise your manuscript taking into account all reviewer and editor comments. Please highlight all changes in the manuscript text file. At this stage we will need you to upload a copy of the manuscript in MS Word .docx or similar editable format.

We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

When revising your manuscript:

*1) Include a "Response to referees" document detailing, point-by-point, how you addressed each referee comment. If no action was taken to address a point, you must provide a compelling argument. This response will be sent back to the referees along with the revised manuscript.

*2) If you have not done so already please begin to revise your manuscript so that it conforms to our Article format instructions, available
here.
Refer also to any guidelines provided in this letter.

*3) Include a revised version of any required Reporting Summary:
https://www.nature.com/documents/nr-reporting-summary.pdf
It will be available to referees (and, potentially, statisticians) to aid in their evaluation if the manuscript goes back for peer review.
A revised checklist is essential for re-review of the paper.

Please be aware of our guidelines on digital image standards.

Please use the link below to submit your revised manuscript and related files:

[REDACTED]

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

We hope to receive your revised manuscript within four to eight weeks. If you cannot send it within this time, please let us know.

Please do not hesitate to contact me if you have any questions or would like to discuss these revisions further.

Nature Genetics is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit please visit www.springernature.com/orcid.

We look forward to seeing the revised manuscript and thank you for the opportunity to review your work.

Sincerely,

Michael Fletcher, PhD
Associate Editor, Nature Genetics

ORCID: 0000-0003-1589-7087

Referee expertise: all three referees are experts in the field of gene regulation in the haematopoetic system.

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:
Remarks to the Author:

In this paper Liu and co-authors use Cas9 and dCas9 to screen the beta-globin locus to identify sequences involved in the activation and repression of the fetally expressed gamma-globin genes. Unexpectedly, the dCas9 screen highlights sequences required for the gamma-globin expression close

to the binding site of the known repressor BCL11A. The authors hypothesise Cas9 binding at this site prevents the binding of an activating factor due to steric inhibition and motif analysis identifies a CCAAT site and NF-Y as a candidate activator. The authors test this using base editing of the NF-Y binding motif in the HUDEP-2 erythroid cell line and CD34+ve cells and siRNA knockdown of NF-Y. The authors show the CCAAT box is occupied by the activator NF-Y and that this factor appears to compete with the repressor BCL11A, which binds 35 bp away, to maintain silencing or expression of the gamma-globin genes. NF-Y reoccupies the site after BCL11A depletion, neatly explaining why these sites are capable of being bound by NF-Y in primitive but not definitive erythropoiesis, despite the activator being present in both cell types.

I have a few reservations about the way some of the data are presented and I also think the authors' data may point to the presence of more potent activators than NF-Y that are being displaced by the dCas9, which adds further interest. However, this paper over all is extremely well written and together this work represents a substantial advance in a really interesting and important area of research.

Major Comments
1) The ChIP-seq data in Supplementary Figure 2b appears to be very low depth, are the scales correct or could the authors mention in the text whether this is a product of normalisation?

The signals were normalized based on sequencing depth, using the SPMR (signal per million reads) option of MACS2. Processing ENCODE data (ENCFF000YUV, NFYA in K562) with the same option resulted in similar data scaling. We have added the following details to the methods section, and updated Extended Data Fig. 2b.
"Peaks were called using MACS2 with the following parameters: -g hs -f BAMPE -q 0.01 -B –SPMR."

The small peak heights we questioned are due to their usage of 'signal per million reads' which compresses the data, it loses some detail, but I think it's legitimate and compresses the data well for display in small figures.

2) Is 3C-qPCR the most sensitive method to determine chromatin interactions? In addition the effect of NFYA depletion may be direct or indirect and authors could comment on this.

Common techniques to view chromatin interactions for one or a few viewpoints include 3C-qPCR, 4C-seq and Capture-C, among others. The relationship between 3C-qPCR and the other two is similar to that of ChIP-qPCR and ChIP-seq. Whereas 3C-qPCR quantifies one-to-few interactions, 4C-seq and Capture-C identify many more interactions to one or a number of viewpoints in a less biased way. The downside of 4C-seq and Capture-C seq is that the 20-25 cycles of PCR amplification prior to sequencing inevitably introduces bias. Therefore, they are not substantially better in quantifying one-to-one interactions than 3C-qPCR.

CaptureC uses indexes to remove PCR duplicates. The analytical protocol also filters for sonication ends to minimise PCR bias so I think that despite using 20-25 cycles it does reflect quantitation. It is not clear how many cycles of amplification were used to detect the interactions described here. This should be included.

We agree that the role of NFYA in chromatin interaction might be indirect. It may act through other looping factors. We added the following sentence to the main text to clarify.
"Whether NF-Y mediates looping directly or through other factors is as yet unknown."

This is acceptable

3) Could the authors comment on why the CUT and RUN protocol identifies twice as many NF-Y binding sites as the equivalent ChIP-seq experiment and what the degree of overlap is between the techniques.

CUT&RUN identifies both direct binding sites and genomic contact sites, since a pA-MNase tethered to a TF not only cleaves the TF bound sequences but also cleaves DNA in close proximity. Therefore CUT&RUN also maps chromatin interactions. This has been demonstrated by comparing CTCF CUT&RUN and CTCF ChIA-PET (Skene, eLife 2017, Figure 8). We show in Extended Data Figure 3b that NFYA CUT&RUN identified 59% of NFYA ChIP-seq peaks, similar to what we previously reported for CTCF and GATA1 CUT&RUN (46% and 51%, Liu, Cell 2018).

This does raise problems in interpreting CUT&RUN data in that it is recording direct binding, indirect binding and chromatin interactions and these points should be included as caveats in the manuscript.

4) For Figure 2c and Supplementary Figure 2b there is no scale, making it difficult to properly assess these data.

We have updated the figures to include scales. OK

5) Figure legends need to be clearer.

We have carefully reviewed and edited the Figure legends. OK

6) Proposed reasons for the discrepancy between the effects dCas9 and Cas9 should be more clearly explained (Figure 1b).
We have added the following text to explain why the Cas9 screen identified widespread HbF "repressor sequences":
"We suspect that the induction of HbF by targeting these sequences reflects secondary effects of Cas9 editing and subsequent DNA repair. In addition, gRNAs targeting the duplicated HBG1 and HBG2 genes frequently result in 4.9 kb deletions (Traxler, Nat Med 2016), which may lead to remodeling of local chromatin or removal of repressive elements. Authentic cis-acting elements may therefore be obscured."
In contrast, dCas9 acts through interfering with binding of endogenous regulators in a narrow region. We have added the following sentence:
"Therefore, dCas9 screens allow for finer dissection of cis-acting elements in a target genomic region."

This is an interesting point

7) The statistical method used in Figure 4a should be better explained. Ideally a value for BCL11A protein level before KO should be added and used as a comparison to show a fold change over time. The authors could then use the data for the AAVS1 control sgRNA to prove the specificity of the targeting.
We have updated Figure 4a to include the time point before KO. We also updated the methods section to include the following sentence describing the band quantification method:
"Western blot was used to assess BCL11A levels before and after nucleofection, and the band intensities were quantified using ImageJ. Briefly, the pixel densities of target bands and blank regions were measured. The pixel numbers of blank regions represented background and were subtracted from the band. Three independent measurements were performed for each band, and BCL11A signals were normalized using Histone H3." OK

8) The dynamic assessment of BCL11A binding in CD34+ve cells is a really excellent aspect of the work, we wonder whether this could be further improved by using more quantitative technique?
We add an Extended Data Figure 5d to quantify PRO-seq counts on each globin gene for Figure 4d, which aids the visualization of the results. OK

9) In Figure 5b why does gamma expression reduce when dCas9 is moved further leftwards of the NF-Y binding site? Is it displacing other positive regulators (does the underlying sequence give any clues to this?).
We believe the high level of γ-globin expression in -62 and -102 is caused by incomplete eviction of NF-Y by dCas9. When dCas9 is moved further leftwards to -115 or -124, dCas9 can evict NF-Y completely through steric hindrance. This is supported by NF-Y CUT&RUN: weak binding of NF-Y at the γ-globin promoters was detected in cells with dCas9-gRNA(-62/-102), whereas no binding was detected in dCas9-gRNA(-115/-124) cells. Difference in NF-Y displacement activities may be attributed to different binding efficiencies of the gRNAs.

An explanation of this should be included in the manuscript

Could the authors explain in more detail the advantages of using the Target-AID-NG method? It is not clear to this reviewer why the combined plasmid produced such low levels of protein. The plasmid does not appear to be over 10kb so I would not have anticipated any problems with transfection efficiency. However, a split Cas9 could be beneficial for others who work with difficult to transfect cell lines.

We have included the sentence below in the methods section to clarify why we engineered the split Target-AID-NG.
"Introduction of the original Target-AID-NG produced protein at very low level (Extended Data Figure 4b) which did not lead to detectable base editing in HUDEP-2 cells. We reasoned that a higher level of base editor protein was needed for successful base editing. To this end, we engineered a split Target-AID-NG." OK

We do not know the reason why Target-AID-NG produced very little protein. We infer that this was due to poor virus production, transcription, or translation.

In Supplementary Figure 5e, we see that the HBG expression level at 72h only gets partially restored when both BCL11A and NFYA are targeted. This interestingly suggests NF-Y is a weak activator and/or works in concert with other factors. Can the authors comment on this? Could NF-Y work as a pioneer factor (explaining why it is recruited to the site at the early time point/32h) which then recruits other proteins?

It is highly likely that NF-Y works in concert with other transcription factors to activate γ-globin. We added a comment on this in the result section:
"Similar results were obtained in primary human CD34+ cell derived erythroid precursors, though with a lesser degree of γ-globin reduction, possibly due to the action of other transcription activators"
With regard to the pioneer activity of NF-Y, it was shown in embryonic stem cells that NF-Y promotes chromatin accessibility and enhances the binding of other TFs, such as Oct4, Sox2 and Nanog, to their targets (Oldfield, Mol Cell 2014). It is possible that NF-Y activates γ-globin through a similar mechanism. We added a comment on this in the results section.
"These data reveal that NF-Y binds rapidly to the γ-globin promoters upon BCL11A depletion to open up the local chromatin, which precedes formation of enhancer-promoter contacts and transcriptional activation. These findings are consistent with the reported pioneer transcription factor activity of NF-Y"
OK

Minor Comments
Legend for Fig 4b should be corrected.
Addressed. OK

Figure 5B would be clearer with a title (perhaps NF-Y CUT and RUN).
Updated. OK

The legends for Figure 1a and b should describe the tracks more precisely.
We have updated the figure legends. OK

On page 5: "CCAAT box often co-occurs with other TF motifs with precise spatial positioning, suggesting an architectural role in promoter activation". This is very interesting. Is there any example in haematopoietic lineages the authors could use to illustrate this point?

The statement is based on findings reported in reference 23-25. Two of the studies were conducted by analyzing ENCODE ChIP-seq data in K562 and GM12878 cells. In these studies, the binding motifs of several TFs, including cFOS, MAX, etc, were found to be enriched at a precise distance to CCAAT, strongly suggesting cooperative binding in a stereo-specific manner. The exact nature of such cooperativity and why a precise distance is required is interesting and worthy of additional study. OK

Use of both HUDEP-1 and HUDEP-2 cell lines should be acknowledged.

We have added the following statement in the acknowledgment section:
"We thank Yukio Nakamura for HUDEP-1 and HUDEP-2 cell lines." OK

A clearer annotation of Figure 2e would be appreciated.

We have updated Figure 2e by labeling each subunit of NF-Y. The legend of Figure 2e is also update with the following information:
"flanking sequences are wrapped around NF-Y through histone-fold domains of NFYB and NFYC. NFYA is responsible for motif recognition." OK

In Figure 3b: it would be useful to add "NF-Y motif base edited" on the X axis (as it has been done on Figure 3c).

Updated. OK

Can the authors highlight in Figure 3C which clones have been chosen for subsequent work? The authors could use a different shape on the dataset to highlight the HBG expression level of the clones selected.

We have circled the clones in Figure 3c and 3g that were used in Figure 3d and added a description in legends. OK

Page 16: E. coli should be written in italic. Can the name of the bacterial strain be added? The word "transduce" should be changed to "Transduction". The unit "g" (centrifugal units) should be written in italic.

Corrected. E. coli strain has been added to the method section. OK

Page 19, paragraph "shRNA knockdown": the phrase "medium were exchanged" should be altered.

Rephrased to "fresh media supplemented with puromycin was added to select transduced cells"

Page 20: the word "by" in the sentence "brought up to 1.5 mL by distilled water" should be changed to "with".

Corrected. OK

Reviewer #2:
Remarks to the Author:
The authors have addressed all my queries satisfactorily. The data are convincing ad reveal NFY as an

activator of the fetal gamma-globin genes.

Reviewer #3:
Remarks to the Author:
The revisions have appropriately addressed the recommendations made in my prior review. The molecular mechanism proposed is supported by the data presented.

Author Rebuttal, first revision:

Responses to reviewers' comments:

Reviewer #1:
Remarks to the Author:
In this paper Liu and co-authors use Cas9 and dCas9 to screen the beta-globin locus to identify sequences involved in the activation and repression of the fetally expressed gamma-globin genes. Unexpectedly, the dCas9 screen highlights sequences required for the gamma-globin expression close to the binding site of the known repressor BCL11A. The authors hypothesise Cas9 binding at this site prevents the binding of an activating factor due to steric inhibition and motif analysis identifies a CCAAT site and NF-Y as a candidate activator. The authors test this using base editing of the NF-Y binding motif in the HUDEP-2 erythroid cell line and CD34+ve cells and siRNA knockdown of NF-Y. The authors show the CCAAT box is occupied by the activator NF-Y and that this factor appears to compete with the repressor BCL11A, which binds 35 bp away, to maintain silencing or expression of the gamma-globin genes. NF-Y reoccupies the site after BCL11A depletion, neatly explaining why these sites are capable of being bound by NF-Y in primitive but not definitive erythropoiesis, despite the activator being present in both cell types.

I have a few reservations about the way some of the data are presented and I also think the authors' data may point to the presence of more potent activators than NF-Y that are being displaced by the dCas9, which adds further interest. However, this paper over all is extremely well written and together this work represents a substantial advance in a really interesting and important area of research.

Major Comments
1) The ChIP-seq data in Supplementary Figure 2b appears to be very low depth, are the scales correct or could the authors mention in the text whether this is a product of normalisation?

The signals were normalized based on sequencing depth, using the SPMR (signal per million reads) option of MACS2. Processing ENCODE data (ENCFF000YUV, NFYA in K562) with the same option resulted in similar data scaling. We have added the following details to the methods section, and updated Extended Data Fig. 2b.
"Peaks were called using MACS2 with the following parameters: -g hs -f BAMPE -q 0.01 -B –SPMR."

The small peak heights we questioned are due to their usage of 'signal per million reads' which compresses the data, it loses some detail, but I think it's legitimate and compresses the data well for display in small figures.

2) Is 3C-qPCR the most sensitive method to determine chromatin interactions? In addition the effect of NFYA depletion may be direct or indirect and authors could comment on this.

Common techniques to view chromatin interactions for one or a few viewpoints include 3C-qPCR, 4C-seq and Capture-C, among others. The relationship between 3C-qPCR and the other two is similar to that of ChIP-qPCR and ChIP-seq. Whereas 3C-qPCR quantifies one-to-few interactions, 4C-seq and Capture-C identify many more interactions to one or a number of viewpoints in a less biased way. The downside of 4C-seq and Capture-C seq is that the 20-25 cycles of PCR amplification prior to sequencing inevitably introduces bias. Therefore, they are not substantially better in quantifying one-to-one interactions than 3C-qPCR.

CaptureC uses indexes to remove PCR duplicates. The analytical protocol also filters for sonication ends to minimise PCR bias so I think that despite using 20-25 cycles it does reflect quantitation. It is not clear how many cycles of amplification were used to detect the interactions described here. This should be included.
Ct value for interactions in 3C-qPCR are usually between 27-30 in our experiments. We have included this information in the methods section.

We agree that the role of NFYA in chromatin interaction might be indirect. It may act through other looping factors. We added the following sentence to the main text to clarify.
"Whether NF-Y mediates looping directly or through other factors is as yet unknown."

This is acceptable

3) Could the authors comment on why the CUT and RUN protocol identifies twice as many NF-Y binding sites as the equivalent ChIP-seq experiment and what the degree of overlap is between the techniques.

CUT&RUN identifies both direct binding sites and genomic contact sites, since a pA-MNase tethered to a TF not only cleaves the TF bound sequences but also cleaves DNA in close proximity. Therefore CUT&RUN also maps chromatin interactions. This has been demonstrated by comparing CTCF CUT&RUN and CTCF ChIA-PET (Skene, eLife 2017, Figure 8). We show in Extended Data Figure 3b that NFYA CUT&RUN identified 59% of NFYA ChIP-seq peaks, similar to what we previously reported for CTCF and GATA1 CUT&RUN (46% and 51%, Liu, Cell 2018).

This does raise problems in interpreting CUT&RUN data in that it is recording direct binding, indirect binding and chromatin interactions and these points should be included as caveats in the manuscript.

We have included the following sentence in the manuscript.
"CUT&RUN does not include an immunoprecipitation step for chromatin enrichment and reports indirect peaks caused by pA-MNase cutting at proximal regions. However, peaks that reflect direct TF binding are identified by the presence of a TF footprint (see below)."

4) For Figure 2c and Supplementary Figure 2b there is no scale, making it difficult to properly assess these data.

We have updated the figures to include scales. OK

5) Figure legends need to be clearer.

We have carefully reviewed and edited the Figure legends. OK

6) Proposed reasons for the discrepancy between the effects dCas9 and Cas9 should be more clearly explained (Figure 1b).

We have added the following text to explain why the Cas9 screen identified widespread HbF "repressor sequences":
"We suspect that the induction of HbF by targeting these sequences reflects secondary effects of Cas9 editing and subsequent DNA repair. In addition, gRNAs targeting the duplicated HBG1 and HBG2 genes frequently result in 4.9 kb deletions (Traxler, Nat Med 2016), which may lead to remodeling of local chromatin or removal of repressive elements. Authentic cis-acting elements may therefore be obscured."
In contrast, dCas9 acts through interfering with binding of endogenous regulators in a narrow region. We have added the following sentence:
"Therefore, dCas9 screens allow for finer dissection of cis-acting elements in a target genomic region."

This is an interesting point

7) The statistical method used in Figure 4a should be better explained. Ideally a value for BCL11A protein level before KO should be added and used as a comparison to show a fold change over time. The authors could then use the data for the AAVS1 control sgRNA to prove the specificity of the targeting.
We have updated Figure 4a to include the time point before KO. We also updated the methods section to include the following sentence describing the band quantification method:
"Western blot was used to assess BCL11A levels before and after nucleofection, and the band intensities were quantified using ImageJ. Briefly, the pixel densities of target bands and blank regions were measured. The pixel numbers of blank regions represented background and were subtracted from the band. Three independent measurements were performed for each band, and BCL11A signals were normalized using Histone H3." OK

8) The dynamic assessment of BCL11A binding in CD34+ve cells is a really excellent aspect of the work, we wonder whether this could be further improved by using more quantitative technique?
We add an Extended Data Figure 5d to quantify PRO-seq counts on each globin gene for Figure 4d, which aids the visualization of the results. OK

9) In Figure 5b why does gamma expression reduce when dCas9 is moved further leftwards of the NF-Y binding site? Is it displacing other positive regulators (does the underlying sequence give any clues to this?).

We believe the high level of γ-globin expression in -62 and -102 is caused by incomplete eviction of NF-Y by dCas9. When dCas9 is moved further leftwards to -115 or -124, dCas9 can evict NF-Y completely through steric hindrance. This is supported by NF-Y CUT&RUN: weak binding of NF-Y at the γ-globin promoters was detected in cells with dCas9-gRNA(-62/-102), whereas no binding was detected in dCas9-gRNA(-115/-124) cells. Difference in NF-Y displacement activities may be attributed to different binding efficiencies of the gRNAs.

An explanation of this should be included in the manuscript

We have included the following explanation in the text.
"NF-Y occupancy was reduced, though partially, upon placement of dCas9 at -102 or -62, which overlaps the NF-Y motif or downstream flanking sequence, respectively. These findings are in line with the modest reduction of γ-globin expression."

Could the authors explain in more detail the advantages of using the Target-AID-NG method? It is not clear to this reviewer why the combined plasmid produced such low levels of protein. The plasmid does not appear to be over 10kb so I would not have anticipated any problems with transfection efficiency. However, a split Cas9 could be beneficial for others who work with difficult to transfect cell lines.

We have included the sentence below in the methods section to clarify why we engineered the split Target-AID-NG.
"Introduction of the original Target-AID-NG produced protein at very low level (Extended Data Figure 4b) which did not lead to detectable base editing in HUDEP-2 cells. We reasoned that a higher level of base editor protein was needed for successful base editing. To this end, we engineered a split Target-AID-NG." OK

We do not know the reason why Target-AID-NG produced very little protein. We infer that this was due to poor virus production, transcription, or translation.

In Supplementary Figure 5e, we see that the HBG expression level at 72h only gets partially restored when both BCL11A and NFYA are targeted. This interestingly suggests NF-Y is a weak activator and/or works in concert with other factors. Can the authors comment on this? Could NF-Y work as a pioneer factor (explaining why it is recruited to the site at the early time point/32h) which then recruits other proteins?

It is highly likely that NF-Y works in concert with other transcription factors to activate γ-globin. We added a comment on this in the result section:
"Similar results were obtained in primary human CD34+ cell derived erythroid precursors, though with a lesser degree of γ-globin reduction, possibly due to the action of other transcription activators"
With regard to the pioneer activity of NF-Y, it was shown in embryonic stem cells that NF-Y promotes chromatin accessibility and enhances the binding of other TFs, such as Oct4, Sox2 and Nanog, to their targets (Oldfield, Mol Cell 2014). It is possible that NF-Y activates γ-globin through a similar mechanism. We added a comment on this in the results section.
"These data reveal that NF-Y binds rapidly to the γ-globin promoters upon BCL11A depletion to open up the local chromatin, which precedes formation of enhancer-promoter contacts and transcriptional activation. These findings are consistent with the reported pioneer transcription factor activity of NF-Y" OK

Minor Comments
Legend for Fig 4b should be corrected.
Addressed. OK

Figure 5B would be clearer with a title (perhaps NF-Y CUT and RUN).
Updated. OK

The legends for Figure 1a and b should describe the tracks more precisely.
We have updated the figure legends. OK

On page 5: "CCAAT box often co-occurs with other TF motifs with precise spatial positioning, suggesting an architectural role in promoter activation". This is very interesting. Is there any example in haematopoietic lineages the authors could use to illustrate this point?

The statement is based on findings reported in reference 23-25. Two of the studies were conducted by analyzing ENCODE ChIP-seq data in K562 and GM12878 cells. In these studies, the binding motifs of several TFs, including cFOS, MAX, etc, were found to be enriched at a precise distance to CCAAT, strongly suggesting cooperative binding in a stereo-specific manner. The exact nature of such cooperativity and why a precise distance is required is interesting and worthy of additional study. OK

Use of both HUDEP-1 and HUDEP-2 cell lines should be acknowledged.

We have added the following statement in the acknowledgment section:

"We thank Yukio Nakamura for HUDEP-1 and HUDEP-2 cell lines." OK

A clearer annotation of Figure 2e would be appreciated.

We have updated Figure 2e by labeling each subunit of NF-Y. The legend of Figure 2e is also update with the following information:
"flanking sequences are wrapped around NF-Y through histone-fold domains of NFYB and NFYC. NFYA is responsible for motif recognition." OK

In Figure 3b: it would be useful to add "NF-Y motif base edited" on the X axis (as it has been done on Figure 3c).

Updated. OK

Can the authors highlight in Figure 3C which clones have been chosen for subsequent work? The authors could use a different shape on the dataset to highlight the HBG expression level of the clones selected.

We have circled the clones in Figure 3c and 3g that were used in Figure 3d and added a description in legends. OK

Page 16: E. coli should be written in italic. Can the name of the bacterial strain be added? The word "transduce" should be changed to "Transduction". The unit "g" (centrifugal units) should be written in italic.

Corrected. E. coli strain has been added to the method section. OK
Page 19, paragraph "shRNA knockdown": the phrase "medium were exchanged" should be altered.

Rephrased to "fresh media supplemented with puromycin was added to select transduced cells"

Page 20: the word "by" in the sentence "brought up to 1.5 mL by distilled water" should be changed to "with".

Corrected. OK

Reviewer #2:
Remarks to the Author:
The authors have addressed all my queries satisfactorily. The data are convincing ad reveal NFY as an activator of the fetal gamma-globin genes.
No responses required

Reviewer #3:
Remarks to the Author:
The revisions have appropriately addressed the recommendations made in my prior review. The molecular mechanism proposed is supported by the data presented.
No responses required

Decision Letter, second revision:

Our ref: NG-A54945R1

18th Dec 2020

Dear Stu,

Thank you for submitting your revised manuscript "Transcription factor competition at the γ-globin promoters controls hemoglobin switching" (NG-A54945R1).

We have made our editorial checks on your revision, and the changes have addressed the remaining comments to our satisfaction. Therefore, we will be happy in principle to publish it in Nature Genetics, pending minor revisions to satisfy the referees' final requests and to comply with our editorial and formatting guidelines.

** Note that we will send you a checklist detailing these editorial and formatting requirements in about a week. Please do not finalize your revisions or upload the final materials until you receive this additional information.**

In recognition of the time and expertise our reviewers provide to Nature Genetics's editorial process, we would like to formally acknowledge their contribution to the external peer review of your manuscript entitled "Transcription factor competition at the γ-globin promoters controls hemoglobin switching". For those reviewers who give their assent, we will be publishing their names alongside the published article.

While we prepare these instructions, we encourage the Corresponding Author to begin to review and collect the following:

-- Confirmation from all authors that the manuscript correctly states their names, institutional affiliations, funding IDs, consortium membership and roles, author or collaborator status, and author contributions.

-- Declarations of any financial and non-financial competing interests from any author. For the sake of transparency and to help readers form their own judgment of potential bias, the Nature Research Journals require authors to declare any financial and non-financial competing interests in relation to the work described in the submitted manuscript. This declaration must be complete, including author initials, in the final manuscript text.

If you have any questions as you begin to prepare your submission please feel free to contact our Editorial offices at genetics@us.nature.com. We are happy to assist you.

Thank you again for your interest in Nature Genetics.

Sincerely,

Michael Fletcher, PhD
Associate Editor, Nature Genetics

ORCID: 0000-0003-1589-7087

ORCID

Nature Genetics is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS) prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. For more information please visit http://www.springernature.com/orcid

For all corresponding authors listed on the manuscript, please follow the instructions in the link below to link your ORCID to your account on our MTS before submitting the final version of the manuscript. If you do not yet have an ORCID you will be able to create one in minutes. https://www.springernature.com/gp/researchers/orcid/orcid-for-nature-research

IMPORTANT: All authors identified as 'corresponding author' on the manuscript must follow these instructions. Non-corresponding authors do not have to link their ORCIDs but are encouraged to do so. Please note that it will not be possible to add/modify ORCIDs at proof. Thus, if they wish to have their ORCID added to the paper they must also follow the above procedure prior to acceptance.

To support ORCID's aims, we only allow a single ORCID identifier to be attached to one account. If you have any issues attaching an ORCID identifier to your MTS account, please contact the Platform Support Helpdesk.

Final Decision Letter:

In reply please quote: NG-A54945R2 Orkin

21st Jan 2021

Dear Stu,

I am very pleased to say that your manuscript "Transcription factor competition at the γ-globin promoters controls hemoglobin switching" has been accepted for publication in an upcoming issue of Nature Genetics.

Prior to setting your manuscript, we may make minor changes to enhance the lucidity of the text and with reference to our house style. We therefore ask that you examine the proofs most carefully to

ensure that we have not inadvertently altered the sense of your text in any way.

Once your manuscript is typeset and you have completed the appropriate grant of rights, you will receive a link to your electronic proof via email with a request to make any corrections within 48 hours. If, when you receive your proof, you cannot meet this deadline, please inform us at rjsproduction@springernature.com immediately.

Your paper will be published online after we receive your corrections and will appear in print in the next available issue. You can find out your date of online publication by contacting the Nature Press Office (press@nature.com) after sending your e-proof corrections. Now is the time to inform your Public Relations or Press Office about your paper, as they might be interested in promoting its publication. This will allow them time to prepare an accurate and satisfactory press release. Include your manuscript tracking number (NG-A54945R2) and the name of the journal, which they will need when they contact our Press Office.

Before your paper is published online, we shall be distributing a press release to news organizations worldwide, which may very well include details of your work. We are happy for your institution or funding agency to prepare its own press release, but it must mention the embargo date and Nature Genetics. Our Press Office may contact you closer to the time of publication, but if you or your Press Office have any enquiries in the meantime, please contact press@nature.com.

Acceptance is conditional on the data in the manuscript not being published elsewhere, or announced in the print or electronic media, until the embargo/publication date. These restrictions are not intended to deter you from presenting your data at academic meetings and conferences, but any enquiries from the media about papers not yet scheduled for publication should be referred to us.

Nature Genetics is a Transformative open access journal. In approximately 10 business days you will receive an email with a link to choose the appropriate publishing options for your paper and our Author Services team will be in touch regarding any additional information that may be required.

Please note that Nature Research offers an immediate open access option only for papers that were first submitted after 1 January, 2021.

You will not receive your proofs until the publishing agreement has been received through our system.

If you have any questions about our publishing options, costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com

In the event that you choose to publish under the subscription model, Nature Research allows authors to self-archive the accepted manuscript (the version post-peer review, but prior to copy-editing and typesetting) on their own personal website and/or in an institutional or funder repository where it can be made publicly accessible 6 months after first publication, in accordance with our self-archiving policy. Please review our self-archving policy for more information.

Several funders require deposition the accepted manuscript (AM) to PubMed Central or Europe PubMed Central. To enable compliance with these requirements, Nature Research therefore offers a free manuscript deposition service for original research papers supported by a number of PMC/EPMC

participating funders. If you do not choose to publish immediate open access, we can deposit the accepted manuscript in PMC/Europe PMC on your behalf, if you authorise us to do so.

If you have posted a preprint on any preprint server, please ensure that the preprint details are updated with a publication reference, including the DOI and a URL to the published version of the article on the journal website.

To assist our authors in disseminating their research to the broader community, our SharedIt initiative provides you with a unique shareable link that will allow anyone (with or without a subscription) to read the published article. Recipients of the link with a subscription will also be able to download and print the PDF.

As soon as your article is published, you will receive an automated email with your shareable link.

You can now use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

An online order form for reprints of your paper is available at https://www.nature.com/reprints/author-reprints.html. Please let your coauthors and your institutions' public affairs office know that they are also welcome to order reprints by this method.

If you have not already done so, we invite you to upload the step-by-step protocols used in this manuscript to the Protocols Exchange, part of our on-line web resource, natureprotocols.com. If you complete the upload by the time you receive your manuscript proofs, we can insert links in your article that lead directly to the protocol details. Your protocol will be made freely available upon publication of your paper. By participating in natureprotocols.com, you are enabling researchers to more readily reproduce or adapt the methodology you use. Natureprotocols.com is fully searchable, providing your protocols and paper with increased utility and visibility. Please submit your protocol to http://www.nature.com/protocolexchange/. After entering your nature.com username and password you will need to enter your manuscript number (NG-A54945R2). Further information can be found at http://www.natureprotocols.com.

Sincerely,

Michael Fletcher, PhD
Associate Editor, Nature Genetics

ORCID: 0000-0003-1589-7087