

Appendix for A functional overlap between actively transcribed genes and chromatin insulator elements

Table of Contents

<i>Appendix Figure S1: Mutated sequences of CTCF binding sites at the α globin locus</i>	<i>2</i>
<i>Appendix Figure S2: Editing strategy for introducing α-globin sequences into the ectopic site in the α-globin locus, to report effect on Hba-a1 expression</i>	<i>3</i>
<i>Appendix Figure S3: Extended characterisation of in vitro derived CD71+ cells with inserts.</i>	<i>4</i>
<i>Appendix Supplementary information: Limitations on interpretation of Capture-C from insertion models</i>	<i>5</i>
<i>Appendix Table S1: DNA sequences for sgRNAs, ssODNs and screening primers used for targeting CTCF binding sites at the mouse α-globin locus</i>	<i>6</i>
<i>Appendix Table S2: DNA sequences for sgRNAs and regions used for targeting the insertion site.</i>	<i>7</i>

Appendix Figure S1: Mutated sequences of CTCF binding sites at the α globin locus

A

WT HS+44	1	ATTAGAAAGCCAGTGGCGCCACCTTGGGGCCCACTCTCAGCTGTTTATGACTACCGGGCA
Δ 44-48	1	ATTAGAAAGCCAGTGGCGCCA-----
WT HS+44	61	GGAGGTCTCTGGGGCGCCCCCTGCAGGCCACTATAAGTAGGTGCAAGGCTTTTCCAGTAT
Δ 44-48	61	-----AGGTGCAAGGCTTTTCCAGTAT
WT HS+48	1	CCGGAGATATAAGAAGGCGACGAGCACCCCCGTGTGGCGGGACTCGGAACTTCGCAGCTC
Δ 44-48	1	CCGGAGATATAAGAAGGCGACGAGCACCCCCGTG-----GACTCGGAACTTCGCAGCTC

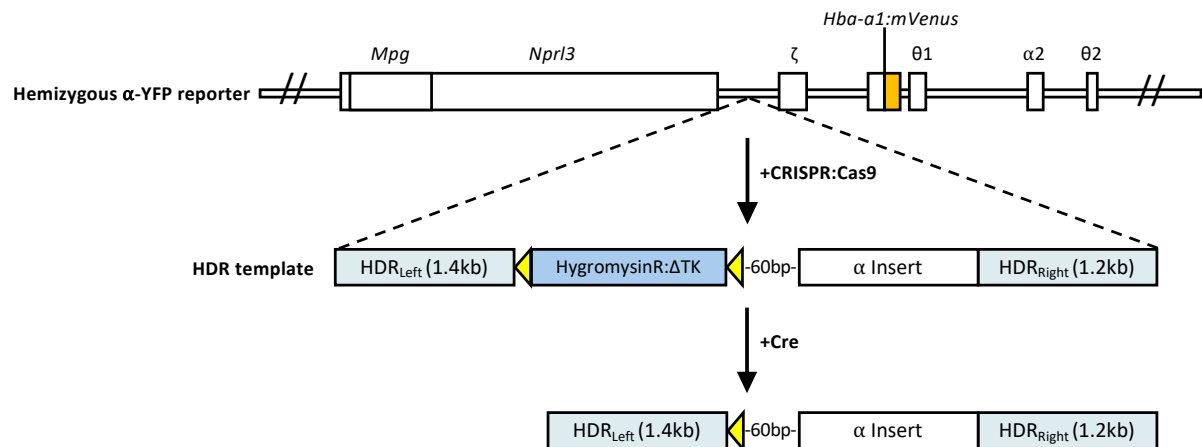
B

WT θ 1	1	CCTAGGCATCCTGGAACGATGTCAGCGCCCCCTGGCGGCCCTCTTGTTGTTTCAGGACGTCTT
Δ θ 1	1	CCTAGGCATCCTGGAACGATGTCAGCGGGATGC-----GCCTCTTGTTGTTTCAGGACGTCTT
WT θ 1	1	CCTAGGCATCCTGGAACGATGTCAGCGCCCCCTGGCGGCCCTCTTGTTGTTTCAGGACGTCTT
Δ θ 1 θ 2	1	CCTAGGCATCCTGGAACGATGTCAGCG-----CTCTTGTTGTTTCAGGACGTCTT
WT θ 2	1	CCTAGGCATCCTGGAACGATGTCAGCGCCCCCTGGCGGCCCTCTTGTTGTTTCAGGACATCTT
Δ θ 2/ Δ θ 1 θ 2	1	CCTAGGCATCCTGGAACGATGTCAGCGGGATGC-----GCCTCTTGTTGTTTCAGGACATCTT

A: Alignments show WT sequences of HS+44 and HS+48 CTCF binding sites, with the 20 bp core binding motif highlighted in blue (reverse orientation), and deleted sequences of HS+44 and HS+48 CTCF binding sites in Δ 44-48 mice (77 bp and 6 bp deletions respectively).

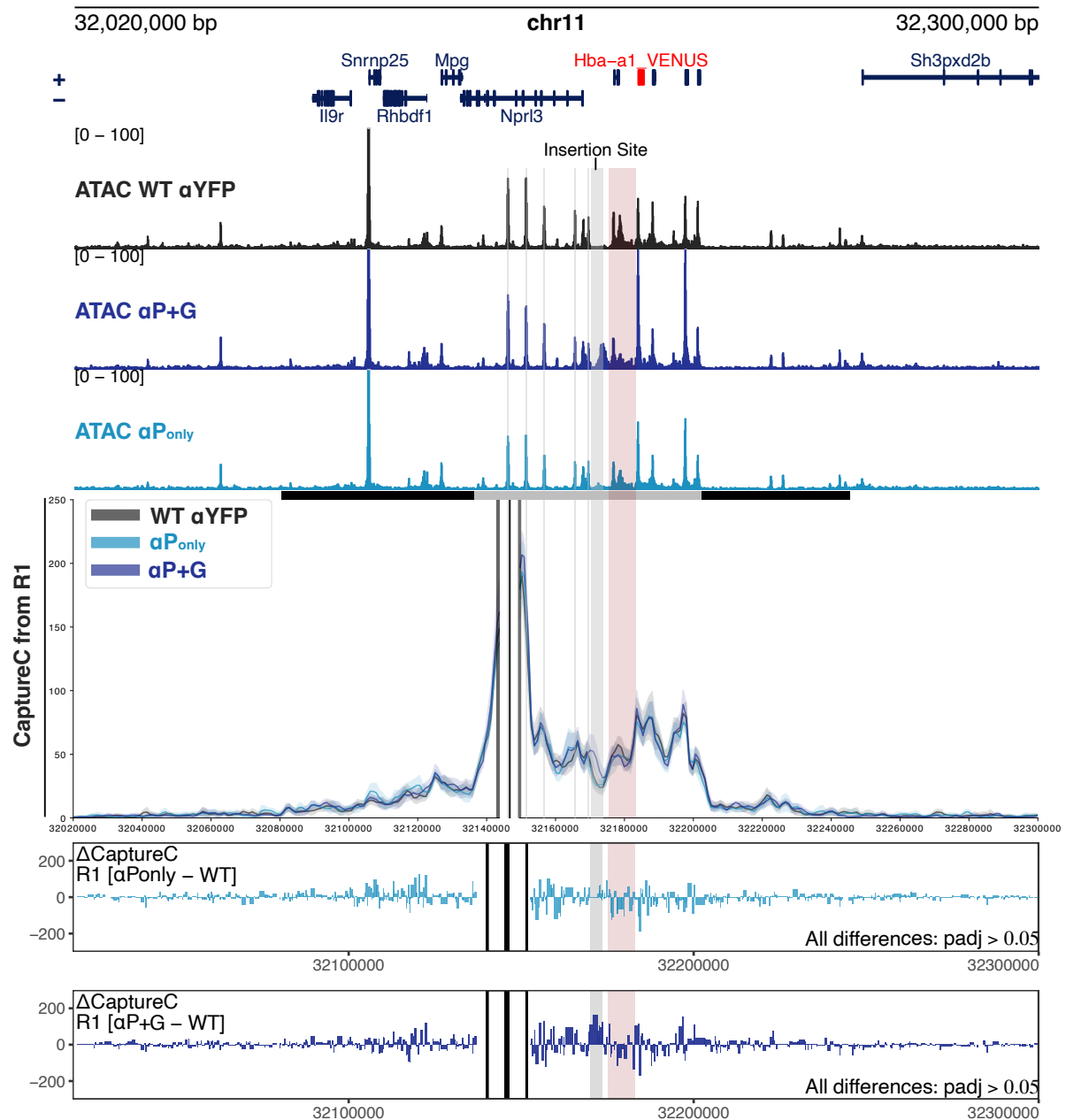
B: Alignments show WT sequences of θ 1 and θ 2 CTCF binding sites, with the 20 bp core binding motif highlighted in blue (reverse orientation), and mutated sequences of θ 1 CTCF binding site in Δ θ 1 (HDR mutation) and Δ θ 1 θ 2 mice (11 bp deletion), and the mutated sequence of θ 2 CTCF binding in Δ θ 2 and Δ θ 1 θ 2 mice (HDR mutation in both models).

Appendix Figure S2: Editing strategy for introducing α -globin sequences into the ectopic site in the α -globin locus, to report effect on Hba-a1 expression



The insertion site is located between the enhancers and *Hba-x* gene (chr11: 32171787 - 32171856 mm9); ~2.6 kb from the R4 enhancer element. The *Hba-a1::mVenus* hemizygous (α -YFP reporter) line was generated previously (Tsang et al. 2024; Francis et al. 2022). *Hba-a1* was edited C-Terminus of the resulting α -globin protein is fused with 2A-peptide:mVenus:3xSV40-NLS. On the other allele (not shown) an 86kb deletion spanning the entire α -globin locus was introduced, such that it can be considered hemizygous for the α -globin locus. The insert sequences were introduced to the locus using CRISPR-Cas9 mediated HDR; the HDR template contains two arms homologous for either side of the insert site and a HygromycinR: Δ TK selection cassette flanked by homotypic loxP sites (yellow triangles), allowing cassette removal upon Cre-mediated recombination. For sgRNA and inserted sequences see **Appendix Table 2**.

Appendix Figure S3: Extended characterisation of in vitro derived CD71+ cells with inserts.



Capture-C from the R1 enhancer in CD71+ erythroid cells derived from edited reporter cells. ATAC-seq in the top panel to show the positions of the elements of interest. Capture-C profiles represent the mean number of normalised unique interactions per restriction fragment from n=3 (WT αYFP), n=4 (αPonly and αP+G) independent differentiations. Differential tracks (Δ CaptureC) show subtractions (αPonly - WT αYFP and αP+G - WT αYFP) of the mean number of unique interactions per restriction fragment, scaled to a total of 100,000 interactions in cis. Note differences represented are not determined to be significant (threshold padj>0.05). The region around the insert (mm9: Insert region: chr11:32169922-32173392) is highlighted in grey, and region covering native targets (*Hba-x* - upstream of *Hba-a1* chr11:32175105-32182970) is highlighted in red. To note, read coverage over the *Hba-a1/2* genes is a combination of both inserted and native sources.

Appendix Supplementary information: Limitations on interpretation of Capture-C from insertion models

Whilst DpnII generated Capture-C profiles do not show a significant change in interactions between the inserts and the native genes, there are limitations to this approach used for this context; first, due to the homology of the native genes and inserts, it was not possible to count the most informative reporter, ie the restriction fragment over the α -globin promoter, as reporter reads from the inserted sequences and the native genes are identical. Secondly, a SNP that could have been used to distinguish between native or inserted α -globin promoters was unfortunately positioned too far from the nearest DpnII cut site, hence they were undetectable with this method. Thirdly, due to the sequence similarity of the inserts and native genes, interaction profiles over the native regions may be masked by multimapping of reporters from native and inserted sequences. Therefore, distributions of specific interactions between the enhancers with each promoter remain unclear. This could be rectified with the use of Mnase derived 3C (MCC), in which interaction reporters would not be limited by the position of restriction sites. Despite these limitations, quantification of the regions flanking the informative sites, shows a subtle trend toward an increase of interactions at the insert site in the inserted α P+G model and reciprocal decrease in interactions near the native α -globin promoter which supports the increased accessibility at the insert sites and decrease in *Hba-a1::Venus* expression. Further to this, R2 interactions with *Hba-x* show a subtle decrease in interactions however this is not observed in R1, further highlighting the limitation of this assay when looking for changes of interactions within regions of a few kb.

Appendix Table S1: DNA sequences for sgRNAs, ssODNs and screening primers used for targeting CTCF binding sites at the mouse α -globin locus

Target	Guide Sequence (5'-3')	ssODN Sequence (5'-3')
$\theta 1$	TGGAACGATGCAGCGCCCCC	CGCTCTGCCCCGCTGGCTGAGCTCAA AGACGTCCTGAAACACAAGAGGCG GATCCCGCTGCATCGTTCCAGGATG CCTAGGTGTTACAGATTCTGGTTC AGCTTTGAGCCCTCTGTTTCCCTGG GCTCCCCCTCC
$\theta 2$	TGAAACACAAGAGGCCGCCA	CTTGCGCTCTGCCCCCTGGCTGAGC TCAAAGATGTCCTGAAACACAAGA GGCGGATCCCGCTGCATCGTTCCAG GATGCCTAGGTGTTACAGATTCTG GTTTCTAGCTTTGAGCCCTCTGTTTCCC TGGGCTCCCT
$\theta 2$	GACATCTTTGAGCTCAGCCA	
HS+44	GAAAGCCAGTGGCGCCACCT	N/A
HS+44	CCCTGCAGGCCACTATAAGT	N/A
HS+48	TCCAAGGTCTCAAGCAGAC	N/A
HS+48	CGACGAGCACCCCCGTGTGG	N/A
Target	Screening primer Sequence (5'-3')	Notes
$\theta 1$	CCCTGGCGGCCTCTTG	
	GTCCAGGACAACAATGCAGC	
$\theta 1$	TGTGGAAGACAGAAGAGGGC	
	CTGAAACACAAGAGGCGCTG	
$\theta 2$	TTCCGAAGGACTCGGGAAGC	Product is BamHI digested : 206bp & 778bp ($\Delta\theta 2$ & $\Delta\theta 1\theta 2$ respectively)
	ATGCACAGAGGCAATGCAGC	
HS+44 /+48	TGTCTGGAAGTATCTGGAAACTGA	
	GGGAAGAATAGACTAGACACATCA	
HS+48	TGCTTGGCCTTCAGGTTTCA	
	AAGTTCCGAGTCCCGCCAC	

Appendix Table S2: DNA sequences for sgRNAs and regions used for targeting the insertion site.

Target	Sequence (5'-3')	Coordinates in mm9
5' sgRNA for insertion site	GCTGTAGTGTAATACTAAGTGC	
3' sgRNA for insertion site	GCTTCAAGAACTGCCTTCCTG	
5' Homology arm		chr11:32170381-32171787
3' Homology arm		chr11:32171856-32173092
α Promoter	AGGGACCCAGAGGGAGAGGTGGGG GGATGGGCGCTGCTCAGTTTGGTTT GAGGGACTTGCTTCTCTGACCAAGG TAGGAGGATACTAAGTCTTCCCAA ACTGCCATCACTGGAGACGTAAGTAA GGGGTAAGAAGTGTGTCCGGGCAA CTGATAAGGATTCCCTGCACCCAGG GGAAGCACAACCCAGCCCCAGAAT CTCAGGGGGCCCTAACAAGTTTACT GGGTAGAGCAAGCACAACCCAGCC AATGAGAACTGCTCCAAGGGCGTGT CCACCCTGCCTGGAGGACACGCCCT TGGAGGGCATATAAGTGTACTTGC TGCAGGACCAAGACACTTCTGATTCT TGACAGACTCAGGAAGAAACCATG GTG	chr11:32183340-32183712
α Gene Body	CTCTCTGGGGAAGACAAAAGCAAC ATCAAGGCTGCCTGGGGGAAGATTG GTGGCCATGGTGCTGAATATGGAGC TGAAGCCCTGGAAAGGTGAGAACA GGACCTTGATCTGTAAGGATCACAG GATCCAATATGGACCTGGCACTCGC TCAGTGGGCAGCTTCTAAGTATGCT TTTCTGTGACCTCAACTTCTTCTCTC TCCTTCTCCAGGATGTTTGCTAGCT TCCCCACCACCAAGACCTACTTCCC TCACTTTGATGTAAGCCACGGCTCT GCCCAGGTCAAGGGTCACGGCAAG AAGGTCGCCGATGCTCTGGCCAATG CTGCAGGCCACCTCGATGACCTGCC CGGTGCCCTGTCTGCTCTGAGCGAC CTGCATGCCCACAAGCTGCGTGTGG ATCCCGTCAACTTCAAGGTATGCGC TGGGACCTGGCAGGCGGCATCTGGG ACCCCTAGGAAGGGCTTGGGGGTCC TCGTGCCCAGGAGGGAACATAGT GGTCCCAGGAAGGGGAGCAGAGGC ATCAGGGTGTCCACTTTGTCTCCGC AGCTCCTGAGCCATTGCCTGCTGGT GACCTTGGCTAGCCACCACCTGCC GATTTACCCCCGCGGTGCATGCCT CTCTGGACAAATTCCTTGCCTCTGTG AGCACCGTGCTGACCTCCAAGTACC GTAAAGCTGCCTTCTGCGGGGCTTG CCTTCTGGCCATGCCCTTCTTCTCTC CCTTGACCTGTACCTCTTGGTCTTT GAATAAAGCCTGAGTAGGAAGAAG CCTGCA	chr11:32183713-32184493