

Variiegated gene expression caused by cell-specific long-range DNA interactions

Daan Noordermeer^{1,4,5}, Elzo de Wit^{1,5}, Petra Klous^{1,5}, Harmen van de Werken¹, Marieke Simonis¹, Melissa Lopez-Jones², Bert Eussen³, Annelies de Klein³, Robert H. Singer² and Wouter de Laat^{1,6}

Mammalian genomes contain numerous regulatory DNA sites with unknown target genes. We used mice with an extra β -globin locus control region (LCR) to investigate how a regulator searches the genome for target genes. We find that the LCR samples a restricted nuclear subvolume, wherein it preferentially contacts genes controlled by shared transcription factors. No contacted gene is detectably upregulated except for endogenous β -globin genes located on another chromosome. This demonstrates genetically that mammalian trans activation is possible, but suggests that it will be rare. Trans activation occurs not pan-cellularly, but in 'jackpot' cells enriched for the interchromosomal interaction. Therefore, cell-specific long-range DNA contacts can cause variegated expression.

High-resolution profiling of transcription-factor binding sites, the discovery of conserved genetic elements and identification of regulatory sites using technologies such as DNaseI hypersensitive site mapping¹, has demonstrated that the number of genomic sites with transcription regulatory potential far exceeds the number of genes. One of the main challenges of the post-genomic era therefore is to assign function to each element. This requires understanding of the capacity of regulatory sites to reach over distance and identify specific target genes at the single-cell level. It is known that mammalian DNA elements can modulate the activity of distant genes on the same chromosome, up to a genomic distance of over a megabase^{2,3}. The three-dimensional structure of the mammalian genome facilitates long-range gene regulation. This was first shown for the β -globin locus, which contains multiple β -globin genes arranged in the order of their developmental expression (Fig. 1a). Upstream of the genes resides a large regulatory DNA element that enhances expression of the β -globin genes up to ~ 100 -fold⁴. The element is called a locus control region (LCR), as it has the capacity to confer position-independent and copy-number-dependent expression on reporter genes in transgenic assays in addition to its classical enhancer activity⁵. Otherwise, the β -globin LCR is not different from classical enhancers: it upregulates gene expression over distance, it functions in a tissue-specific manner and genes compete for its activity^{6,7}. At its endogenous location, the LCR enhances expression of the β -globin genes through physical interactions, thereby looping out the intervening DNA that may carry non-responding genes^{8–10} (Supplementary Fig. S1). A similar mode of

action involving chromatin looping has been shown for enhancers in several other gene loci^{11,12}.

At a higher-order level of organization, microscopy studies have shown that genes can occupy differential positions in the nucleus depending on their expression status^{13–15} and that regulatory DNA elements are instrumental in targeting these genomic regions to specific nuclear positions^{16–19}. Moreover, recent observations suggest that functionally related genes on the same and different chromosomes may come together in the nuclear space for co-transcription²⁰. Finally, some reports have suggested functional communication between regulatory sites located on different chromosomes^{21–25}. Collectively, these studies raise questions on how genes and regulatory sequences explore the nuclear space to engage in functional crosstalk with preferred genomic partners.

RESULTS

Exploring mammalian transvection

To investigate the ability of regulatory DNA elements to relocate chromosomal regions in the nuclear interior and search for preferred target genes, we used transgenic mice with the human erythroid-specific β -globin LCR (hLCR) site-specifically integrated into a gene-dense region on chromosome 8. This region, 8C3–C4, contains many housekeeping genes¹⁸. Two transgenic lines were available, LCR-S and LCR-AS, each carrying the hLCR without globin genes at 8C3–C4, but in opposite orientations (Fig. 1b). In a previous study we showed that in erythroid cells each LCR detectably upregulates several housekeeping

¹Hubrecht Institute-KNAW and University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands. ²Department of Anatomy and Structural Biology, Albert Einstein College of Medicine, Bronx, New York 10461, USA. ³Department of Clinical Genetics, Erasmus Medical Centre, PO Box 2040, 3000 CA Rotterdam, The Netherlands. ⁴Present address: Laboratory of Developmental Genomics, School of Life Sciences, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland. ⁵These authors contributed equally to this work.

⁶Correspondence should be addressed to W.d.L. (e-mail: w.delat@hubrecht.eu)

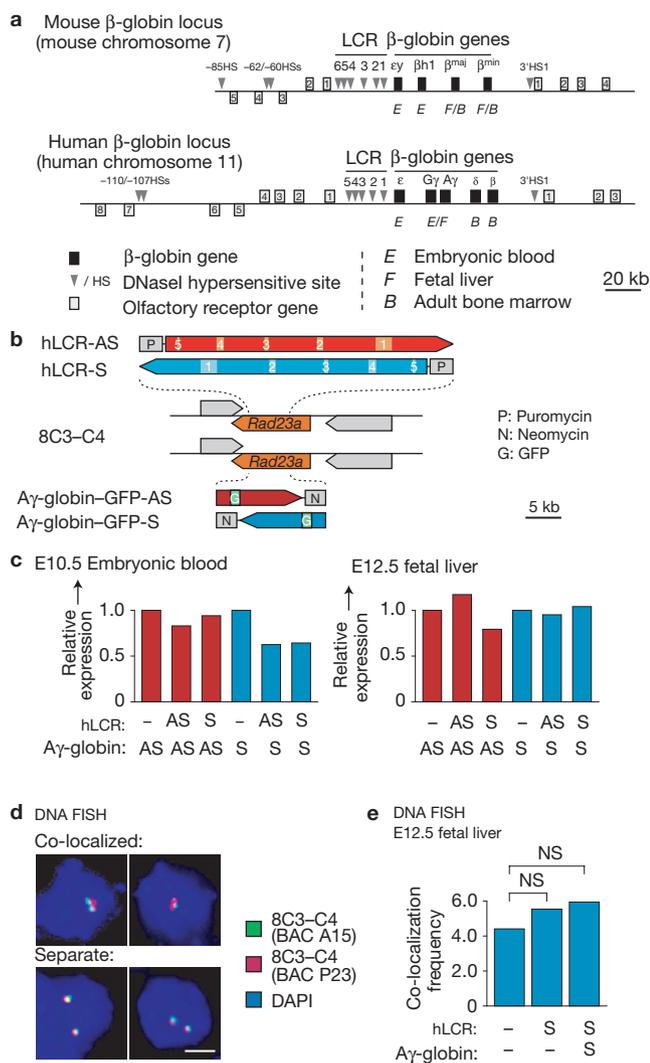


Figure 1 An ectopic LCR does not activate a natural target gene on the homologous chromosome. **(a)** Schematic representation of the endogenous mouse and human β -globin loci. Below each globin gene, gene activity in (transgenic) mice is indicated. **(b)** Targeting strategy for the insertion of the human β -globin LCR and a human $A\gamma$ -globin-green fluorescent protein (GFP) reporter gene into the 8C3-C4 locus on mouse chromosome eight. **(c)** RT-qPCR of $A\gamma$ -globin transcript levels, normalized to *Hprt1* transcript levels. Data are from at least two independent samples. **(d)** Representative examples of DNA FISH showing co-localized and separate 8C3-C4 alleles. DAPI, 4,6-diamidino-2-phenylindole. Scale bar: 2 μ m. **(e)** Co-localization frequencies of 8C3-C4 alleles. Significance levels are indicated above the graph (G-test).

genes that directly surround the integration site up to sixfold, with the most distal one being 150 kilobases (kb) away. Microscopy studies revealed that both LCRs positioned 8C3-C4 more often outside its own chromosome territory, raising the question of to where the LCR migrates its integration site.

We reasoned that, in the most extreme situation, the hLCR could 'search' for a natural target gene present anywhere in the genome, including on its homologous chromosome, as is seen in transvection. The term transvection was coined to describe transcriptional regulation across (paired) homologous chromosomes, a phenomenon mainly studied in *Drosophila*, where homologues frequently pair^{26,27}. Pairing

is generally not observed in mammals²⁸, but may occur in special instances where mono-allelic expression needs to be ensured²⁹⁻³¹. As mammalian chromosomes intermingle extensively³², individual chromosomal segments may indeed invade the territories of other, possibly homologous, chromosomes. We generated transgenic mice carrying a human $A\gamma$ -globin reporter gene in one or the other orientation at exactly the same location in 8C3-C4 (Fig. 1b). The $A\gamma$ -globin gene is a human fetal globin gene that, in mouse transgenics carrying a full human globin cluster including the hLCR, is most highly expressed between embryonic day 10.5 and 12.5 (ref. 33; Fig. 1a). Crossing the $A\gamma$ -globin reporter mice with transgenic mice carrying the hLCR on the homologue revealed no increase in $A\gamma$ -globin gene expression at any of the relevant developmental stages (Fig. 1c). Fluorescence *in situ* hybridization (FISH) experiments also did not reveal increased interaction between the homologues (Fig. 1d,e). Therefore, our experimental system provides no evidence for mammalian transvection involving the hLCR at 8C3-C4. Although these data do not rule out mammalian transvection at other genomic locations, they do strongly suggest that the ectopic hLCR does not have unlimited freedom to search the nuclear interior for natural target genes.

LCR motion is limited by chromosomal context

To investigate in more detail the ability of the hLCR to actively determine its genomic environment, we applied chromosome conformation capture³⁴ on chip (4C) technology to E14.5 fetal livers of wild-type and homozygous LCR-AS mice (Supplementary Fig. S2). 4C captures and identifies spatially proximal DNA sequences to enable an unbiased scan for DNA elements interacting with a locus of choice³⁵. Analysis of the 4C data revealed that the ratio of inter- over intrachromosomal captures increased in the LCR-AS mice (Supplementary Fig. S3), in agreement with the hLCR causing looping out from the chromosome territory (ref. 18). This might be a reflection of the LCR actively engaging in interactions with new interchromosomal partners. To identify specific DNA interactions, sliding-window algorithms were applied that scan the linear chromosome templates for significant clustering of independently captured sequences^{36,37}. Using both conventional analysis and a newly developed high-resolution analysis, we identified a highly similar set of interacting regions for 8C3-C4 with or without an integrated hLCR, both on the same chromosome (in *cis*) and on different chromosomes (in *trans*) (Fig. 2a-c and Supplementary Figs S4, S5a). Extensive fluorescence *in situ* hybridization experiments on cryo-sections (cryo-FISH; ref. 32) validated the 4C data (Fig. 2d,e). Some regions in *trans* showed increased interaction frequencies with the hLCR, again in agreement with the LCR causing looping out of the chromosome territory. However, we did not find chromosomal regions that exclusively interacted with the hLCR. Altogether, this demonstrates that the hLCR at 8C3-C4 does not search the genome to contact new preferred interaction partners. Rather, the chromosomal context of 8C3-C4 dictates the nuclear space that can be explored by the integrated hLCR.

The LCR contacts GATA-1 and EKLF regulated genes

Whereas the overall genomic environment did not change, a quantitative comparison of 4C results showed that the hLCR at 8C3-C4 captured a subset of pre-existing interchromosomal interaction partners more efficiently. This was most obvious for the α -globin locus on chromosome 11, but also for the endogenous β -globin

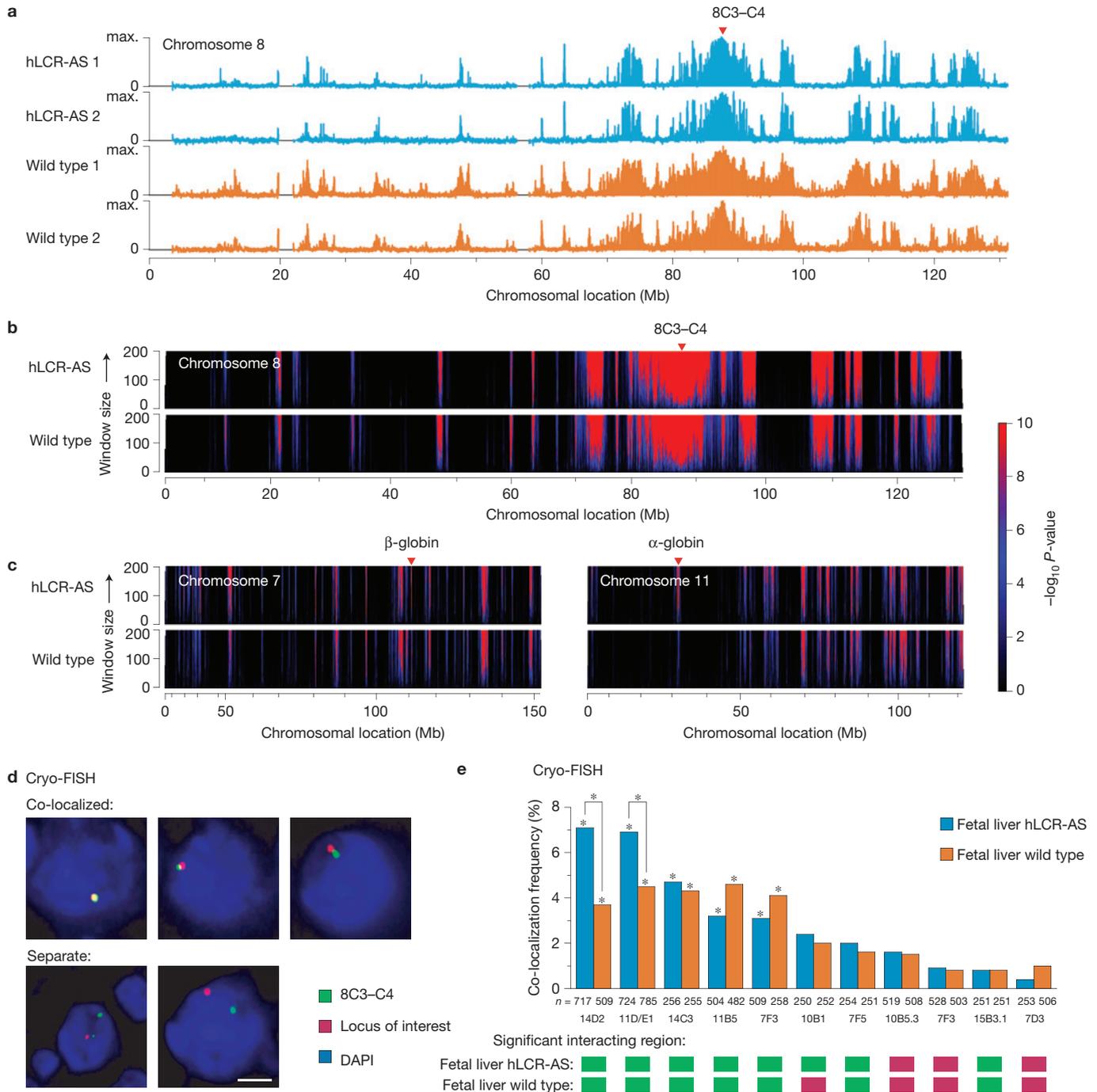


Figure 2 Contacts of 8C3–C4 with and without the LCR are similar. (a) Intrachromosomal DNA interactions of 8C3–C4 with (top) and without (bottom) an integrated β -globin LCR are essentially similar, as determined by 4C analysed with a running-mean analysis of microarray data (average probe spacing: 7 kb). (b) Intrachromosomal DNA interactions of 8C3–C4 with (top) and without (bottom) an integrated β -globin LCR are essentially similar, as determined by 4C analysed with domainograms that visualize probability scores (P -values indicated with colour codes) for the clustering

locus on chromosome 7 (Fig. 3a). Homology between the human and mouse LCRs is limited (<10%) and therefore not expected to underlie the latter contact. The two loci have in common that they are very highly expressed and carry genes controlled by EKLf and GATA-1, two transcription factors that also bind to the LCR.

of positive 4C signals over windows ranging in size from 1 to 200 probes. (c) Interchromosomal 4C data for two chromosomes (7 and 11), analysed as described above. (d) Validation of 4C results by cryo-FISH; examples of results. Scale bar: 2 μ m. (e) Interaction frequencies with a series of genomic regions, measured by cryo-FISH in wild-type and LCR transgenic fetal livers. The number of cells analysed (n) is indicated. Colour codes indicate the significance of the 4C signal (probe clustering), with green referring to $P < 0.01$ and red referring to $P \geq 0.01$.

Ranking genomic regions on the basis of their difference in 4C signal revealed significant enrichment ($P < 0.01$, hypergeometric test) of EKLf (refs 38,39)- and GATA-1 (ref. 40)-regulated genes, as well as of highly expressed genes, among regions contacted more strongly by the LCR (Supplementary Table S1). The same conclusion was drawn

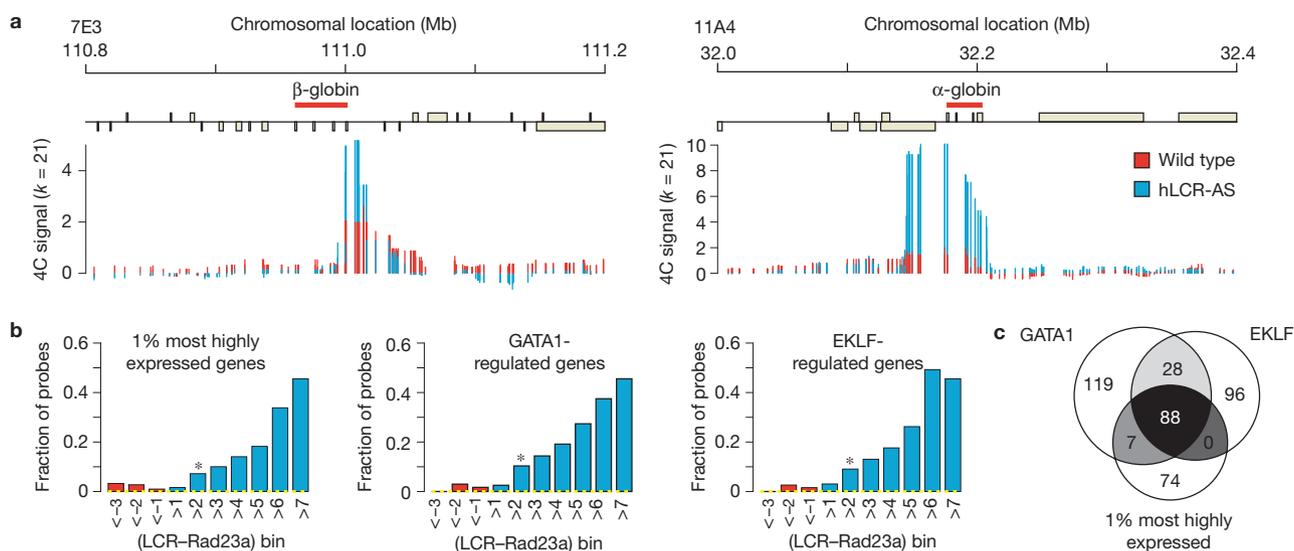


Figure 3 Within a predetermined genomic environment the ectopic LCR shows preferential interactions with specific genes. **(a)** 4C results (running median over sliding windows of 21 probes) for 8C3–C4 with (blue) and without (red) the integrated LCR, at the endogenous β -globin locus on chromosome 7 (left) and the α -globin locus on chromosome 11 (right). **(b)** Probes were binned according to increasing difference in LCR –

from an analysis per microarray probe (Fig. 3b). Each category of genes seems to independently attract the LCR (Fig. 3c and Supplementary Fig. S5b). We conclude that the LCR at 8C3–C4 behaves like a ‘dog on a lead’: the chromosomal context of 8C3–C4 imposes important constraints on its freedom to move, but within the restricted nuclear subcompartment that it occupies the LCR preferentially contacts genes that are controlled by shared transcription factors.

Interchromosomal gene activation by the LCR

Notwithstanding its restricted ability to change the spatial environment of 8C3–C4, the ectopic LCR is involved in many long-range interactions in *cis* and *trans*. We investigated whether any of the contacted genes were upregulated by the LCR. For this, we profiled the transcriptomes of wild-type and LCR-transgenic littermates using microarrays. A small number of genes was found to be upregulated more than twofold in transgenics. These were the previously characterized genes proximal to the LCR in *cis* (ref. 18), plus a single gene in *trans*, *Hbb-bh1*, that was confirmed to be upregulated by quantitative PCR with reverse transcription (RT–qPCR) (Fig. 4 and Supplementary Fig. S6a–d).

Interestingly, *Hbb-bh1*, or $\beta h1$, is one of the endogenous β -globin genes normally expressed at high, LCR-dependent levels earlier in development, in embryonic blood cells (Fig. 1a). In E14.5 fetal livers the gene is looped away from the endogenous LCR (Supplementary Fig. S1). However, three independent probe-sets on the Affymetrix microarray and two independent RT–qPCR primer-sets reveal that the gene is not fully silent, but expressed at basal levels in wild-type fetal livers (Fig. 4 and Supplementary Fig. S6). Importantly, increased *Hbb-bh1* expression was consistently found in transgenics carrying the LCR in either one of the orientations when compared with their wild-type littermates (Fig. 4a,b), but not when 8C3–C4 exclusively carried a *Neo* selection cassette (Fig. 4c). No other β - or α -globin gene detectably changed in expression in the presence of the ectopic LCR (but see below), as judged by microarray analysis and RT–qPCR

wild-type) 4C signal and characterized depending on their location relative to highly expressed and GATA-1- and EKLF-regulated genes. The yellow dashed line represents the expected frequency on the basis of all probes. **(c)** Venn diagram showing the number of, and overlap between, probes captured more frequently in the integrated LCR 4C experiment for each category of genes analysed in the population marked with an asterisk in **b**.

strategies that measure gene expression across cell populations (Fig. 4d and Supplementary Fig. S6e).

Trans-contacts cause variegated expression

Given that the endogenous β -globin locus carrying $\beta h1$ is among the chromosomal regions that are preferentially contacted by the ectopic LCR (Fig. 3a), we hypothesized that such interchromosomal LCR– $\beta h1$ contacts drive increased $\beta h1$ expression. DNA and RNA-FISH analysis showed that the interaction occurs in 5–10% of the cells, whereas higher-resolution cryo-FISH identified contacts between 2 and 3% of the alleles (Supplementary Table S2). Irrespectively, this raises the question of how these relatively rare interchromosomal contacts can account for the overall twofold increase in transcript levels measured across the entire cell population. We reasoned that interaction frequencies measured by FISH in fixed cells may reflect chromatin dynamics, such that over time a given interaction occurs in every cell. Alternatively, they reflect genome conformations that are cell specific and relatively stable after mitotic exit, implying that, in a given nucleus, genomic loci sample overlapping or non-overlapping nuclear subvolumes. If the latter were true, ‘jackpot’ cells should exist with accumulated $\beta h1$ messenger RNA levels in combination with frequent interchromosomal LCR– $\beta h1$ interactions. To investigate this we carried out RNA FISH using a mixture of probes visualizing both primary transcripts and accumulated mRNA (ref. 41; Fig. 5a,b). In wild-type E14.5 fetal liver cells, we failed to detect the $\beta h1$ primary transcript signals that were visible in E10.5 embryonic blood cells (Supplementary Fig. S7), consistent with its marked drop in expression during development. Accumulated cytoplasmic $\beta h1$ mRNA was seen in only a small percentage (3.5%) of cells. These cells also contained the highest adult β -globin transcript levels (*Hbb-b1*, or β -major) and we therefore assumed they represent the most differentiated erythroid cells in the liver that had most time to accumulate β -globin mRNA. Notably, in only one out of 50 of these cells was an interchromosomal interaction

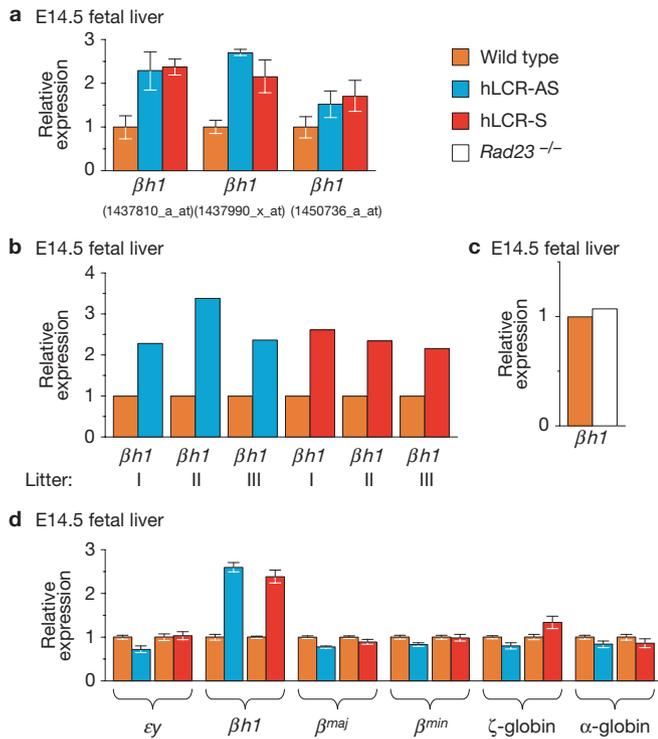


Figure 4 The ectopic LCR on chromosome 8 enhances the expression of the endogenous *beta h1* gene on chromosome 7. **(a)** Affymetrix gene-expression data for all probe-sets analysing *beta h1* transcripts ($n = 3$). **(b)** RT-qPCR comparison of *beta h1* gene expression between multiple wild-type and homozygous LCR-AS littermates, normalized to *Hprt1* transcript levels and to own wild-type littermates. **(c)** RT-qPCR analysis showing that the insertion at 8C3-C4 of a neomycin selection cassette instead of the hLCR does not lead to upregulation of the *beta h1* gene. Data from two independent samples. **(d)** RT-qPCR analysis of expression of β - and α -globin genes in wild-type and homozygous LCR littermates. Error bars: standard error on the basis of 3 littermates (n).

between 8C3-C4 and β -globin observed, which does not exceed the interaction frequency of 5.1% measured across the entire wild-type red-blood-cell population (Fig. 5c and Supplementary Table S2). On the other hand, in LCR transgenic fetal liver cells the percentage of cells with high *beta h1* mRNA levels is increased to 7.0%. More importantly, in these transgenics 15/50 cells (30%) with high *beta h1* levels showed an interchromosomal interaction between 8C3-C4 (with the LCR) and β -globin, a highly significant ($P = 6.6 \times 10^{-7}$, hypergeometric test) increase in interaction frequency when compared with the 8.5% measured across all red blood cells in the transgenics (Fig. 5c and Supplementary Fig. S8a,b). As a control, we analysed α - versus β -globin co-localization in transgenic cells and found no correlation between this interchromosomal interaction and *beta h1* expression levels (Supplementary Fig. S8c).

We next reversed the question and asked whether cells showing the interchromosomal LCR-*beta h1* interaction also had increased *beta h1* transcript levels in their cytoplasm as compared with other red blood cells in the transgenic fetal livers. For this, we developed automated image-analysis software (see Methods) and analysed thousands of cells with respect to their intensity of *beta h1* RNA FISH signal in the cytoplasm. The analysis showed that in transgenics, but not in wild type, cells with the interchromosomal interaction between 8C3-C4 and

the β -globin locus more often have high *beta h1* levels than the other cells from the same tissue (Fig. 6a). Interestingly, albeit less pronounced, the same was found for *beta-major* (Fig. 6b), the adult β -globin gene located next to *beta h1* in the endogenous globin locus (Fig. 1a). Its natural extremely high expression level precluded uncovering an extra contribution from the few cells with this interchromosomal interaction by cell-population-based expression assays such as microarrays or RT-qPCR. However, our analysis at the single-cell level revealed that also this adult β -globin gene benefits from contacts with the extra copy of the LCR. *beta h1* is looped away from the endogenous mouse LCR and as such may be available for contacts with the ectopic human LCR. *beta-major* dynamically forms and breaks contacts with the endogenous LCR (ref. 7), possibly providing opportunity for the ectopic LCR to also engage in contacts and further boost transcription. The genes encoding α -globin did not benefit from contacts with the ectopic LCR, as judged from the single-cell analysis strategy (Fig. 6c). We conclude that our transgenic mice contain a unique population of cells that have increased levels of mRNA for β -globin because these genes on chromosome 7 are contacted and trans activated by the ectopic LCR on chromosome 8.

DISCUSSION

One of the main challenges of the post-genomic era is to assign function to genomic sites, many of which have regulatory potential. Clearly, this cannot be done without considering the dynamics and spatial configuration of the genome. Here, we uncover properties of nuclear organization that dictate the action of regulatory elements in nuclear space. Our findings should contribute to a working model of genome function. The results demonstrate that regulatory DNA elements can search for preferred interaction partners, which in the case of the LCR are genes controlled by shared transcription factors. The ability to roam the nucleus is however heavily constrained by the chromosomal context. We predict the same to be true for almost all genomic locations, although the degree of constraint may vary. The concept of chromosomal context heavily influencing a gene's specific nuclear location seems to contradict more deterministic models of nuclear organization, where functionally related genes are proposed to meet at dedicated nuclear sites²⁰. We cannot exclude that the LCR would have a more notable effect when placed at other genomic locations, or that other enhancers exist that are better capable of repositioning chromosomes and forming specific interchromosomal interactions. We note however that very few, if any, regulatory elements have been described with such a strong influence on gene expression and chromatin organization as the β -globin LCR.

An important finding of this study was that the ectopic, orphan, LCR on chromosome 8 contacted many different genes in *cis* and in *trans*, including those sharing a similar set of regulatory proteins, but that no measurable effect on the expression of most of them was detected. This suggests that, in mammals, ultralong-range gene regulation within and between chromosomes will be rare, or at least difficult to measure in cell populations. Two endogenous β -globin genes on chromosome 7 were the exception, as they were both upregulated by the ectopic LCR in cells with the relevant interchromosomal interaction. As these are natural (mouse) target genes of the (human) LCR, promoter compatibility and spatial proximity seem essential for transcription regulation over distance. Interestingly, a few examples exist of endogenous tissue-specific enhancers activating not only target

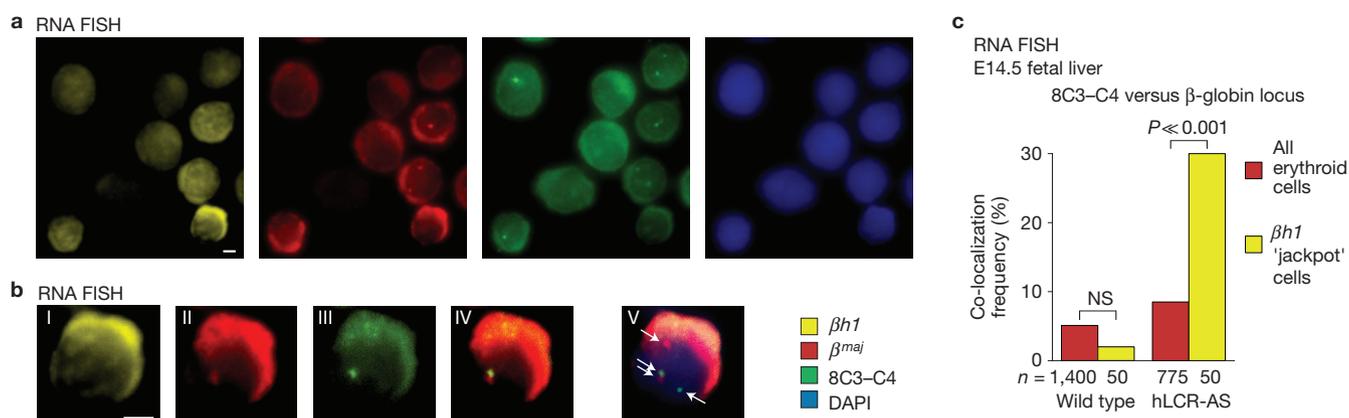


Figure 5 Increased $\beta h1$ mRNA levels in cells showing interchromosomal LCR– $\beta h1$ interactions. (a) RNA FISH on E14.5 fetal liver cells, with one cell showing strongly increased $\beta h1$ mRNA levels in the cytoplasm ('jackpot cell'). Scale bars: 2 μm . (b) Enlargement of the 'jackpot cell', revealing an interchromosomal interaction between the endogenous β -globin locus on chromosome 7 and the ectopic LCR on chromosome 8. (I–IV)

Probes from one focal plane are shown separately and merged. (V) Z stack showing all RNA signals for β^{maj} and 8C3–C4. (c) Quantification of RNA FISH. Determining interchromosomal interaction frequencies between the endogenous β -globin locus and 8C3–C4 without (wild type) and with an integrated LCR (hLCR-AS), in all red blood cells and in ' $\beta h1$ jackpot cells'. The number of cells analysed (n) is indicated. NS: no significant difference.

genes but also non-target genes that happen to be in physical proximity to the enhancer^{42,43}. Our results open the possibility that such bystander activation may be more common in the genome, but appreciable only in individual cells that have their genome folded such that an enhancer and gene happen to be within contacting distance.

Our data provide genetic evidence for classical enhancer activity between mammalian chromosomes, where the genetic addition or deletion of a regulatory DNA element on one chromosome causes increased or reduced expression of a physically interacting gene on another chromosome. As such, we provide formal *in vivo* evidence that mammalian regulatory sites do not need an intervening chromatin fibre to propagate activating signals to responding gene promoters elsewhere in the genome, but that spatial proximity, in combination with enhancer–promoter compatibility, is sufficient for gene activation. Interchromosomal interactions between mammalian regulatory sites and genes have been observed before, but genetic evidence for *trans* activation was lacking so far. For example, the alternatively expressed T_H1 and T_H2 cytokine loci were seen to come together before their activation in naive T cells. On differentiation to T-helper 1 or 2 cells, the interactions between these signature loci were lost and the respective genes turned on²⁵. The functional consequences of this interaction seemed complex, however, and different from classical enhancer activity. The deletion of a regulatory element in the T_H2 locus caused a delay, rather than a reduction, in the expression of the T_H1 gene, and intriguingly this effect was measurable only in differentiating T_H1 cells that no longer showed the interchromosomal interaction. The interaction was proposed to prepare loci for proper expression during subsequent T-helper cell specification, an activity not previously assigned to regulatory sites²⁵. Interchromosomal gene regulation by a single enhancer was suggested to control the expression of all ~1,200 olfactory receptor genes spread across the genome²³, but deletion of the enhancer demonstrated that the enhancer only affects genes in *cis*^{44,45}. In another study the activation of human interferon beta (*IFN- β*) expression in response to viral infection was reported to coincide with interchromosomal interactions with three Alu repeat elements harbouring cryptic NF- κ B sites²¹. Although transfection experiments with plasmids carrying these elements supported the idea

that the DNA interactions boost *IFN- β* expression, formal evidence for interchromosomal enhancer activity awaits demonstration that the chromosomal deletion of one of these repeats causes a drop in *IFN- β* expression. Finally, the imprinting control region (ICR) of the *H19-Igf2* locus has been the subject of several studies on interchromosomal DNA interactions^{22,24,46}. The data did not reveal *trans* activation and were not necessarily consistent, as each study identified different interchromosomal interactions with different functional outcomes, possibly owing to the use of different cell types and/or experimental approaches.

An interesting observation from our artificial system is that interchromosomal interactions can lead to variegated levels of accumulated transcripts in the individual cells. We propose to term the observed phenomenon that cell-specific long-range DNA interactions cause variable gene expression levels among otherwise identical cells 'spatial effect variegation' (SEV; Fig. 6d). Stochastic cell-to-cell variation in gene expression, or transcriptional noise, is common in cell populations⁴⁷. Our data open the possibility that SEV may be one of the underlying mechanisms of transcriptional noise. In such a scenario, the nature of the interacting region will determine whether gene expression goes up or down in the corresponding cell. This is different from position effect variegation⁴⁸, where variable expression of ectopically placed genes is classically thought to be caused by repressive effects from directly surrounding chromatin. Future research should indicate if SEV is acting on endogenous genes. If so, it may provide specific cells within a larger population with a mechanism to make autonomous cell-fate decisions, without the need for external signalling. □

METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/naturecellbiology>

Note: Supplementary Information is available on the Nature Cell Biology website

ACKNOWLEDGEMENTS

We thank J. Marteiijn and W. Vermeulen for providing *Rad23a* knockout material, Y. Oz for counting FISH slides, V. Buckle for providing BAC probes, J. van Rheenen and the Hubrecht Imaging Center for help with image analysis and E. Splinter and other members of the group for assistance. This work was financially supported by

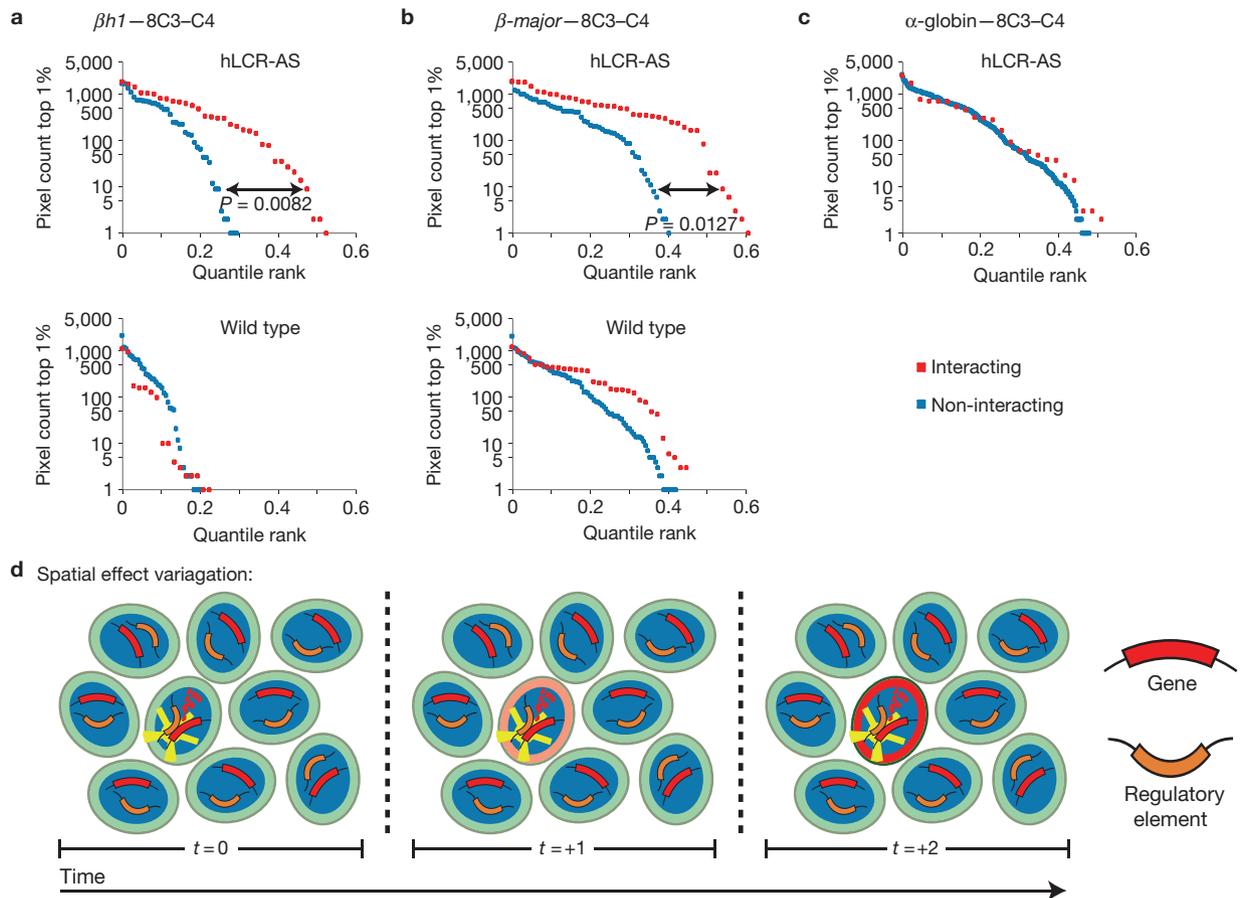


Figure 6 Increased $\beta h1$ and $\beta\text{-major}$ mRNA levels in cells showing interchromosomal LCR- $\beta h1$ interactions. **(a,b)** Automated RNA-FISH image-analysis (see Methods) results, showing that cells in which the ectopic LCR interacts *in trans* with the endogenous β -globin locus more often have high $\beta h1$ **(a)** or $\beta\text{-major}$ **(b)** transcript levels than cells that have the loci apart. **(c)** Cells in which the ectopic LCR interacts *in trans* with the endogenous α -globin locus do not

differ in their levels of mRNA for α -globin from cells without this interchromosomal interaction. The probability score for the difference in distributions of interacting and non-interacting cells is calculated by a one-sided Kolmogorov-Smirnov test. **(d)** SEV: variegated expression among otherwise identical cells caused by cell-specific long-range DNA interactions (intra- or interchromosomal) that are relatively stable during interphase.

grants from the Dutch Scientific Organization (NWO) (91204082 and 935170621) and a European Research Council Starting Grant (209700, '4C') to W.d.L.

AUTHOR CONTRIBUTIONS

D.N. and W.d.L. designed the experiments, analysed the data and, with help from E.d.W., wrote the manuscript. D.N. and P.K. carried out experiments. E.d.W. analysed 4C data and developed the automated FISH image analysis. H.v.d.W. analysed 4C and microarray expression data. M.S. carried out 4C experiments. M.L.-J. and R.H.S. designed and synthesized RNA-FISH probes. B.E. and A.d.K. helped with the FISH experiments.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at <http://www.nature.com/naturecellbiology>

Reprints and permissions information is available online at <http://www.nature.com/reprints>

- Boyle, A. P. *et al.* High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**, 311–322 (2008).
- Amano, T. *et al.* Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. *Dev. Cell* **16**, 47–57 (2009).
- Kleinjan, D. J. & Coutinho, P. Cis-rupture mechanisms: disruption of cis-regulatory control as a cause of human genetic disease. *Brief Funct. Genomic Proteom.* **8**, 317–332 (2009).
- Epner, E. *et al.* The β -globin LCR is not necessary for an open chromatin structure or developmentally regulated transcription of the native mouse β -globin locus. *Mol. Cell* **2**, 447–455 (1998).

- Grosfeld, F., van Assendelft, G. B., Greaves, D. R. & Kollias, G. Position-independent, high-level expression of the human β -globin gene in transgenic mice. *Cell* **51**, 975–985 (1987).
- Wasylyk, B., Wasylyk, C., Augereau, P. & Chambon, P. The SV40 72 bp repeat preferentially potentiates transcription starting from proximal natural or substitute promoter elements. *Cell* **32**, 503–514 (1983).
- Wijgerde, M., Grosfeld, F. & Fraser, P. Transcription complex stability and chromatin dynamics *in vivo*. *Nature* **377**, 209–213 (1995).
- Carter, D., Chakalova, L., Osborne, C. S., Dai, Y. F. & Fraser, P. Long-range chromatin regulatory interactions *in vivo*. *Nat. Genet.* **32**, 623–626 (2002).
- Tolhuis, B., Palstra, R. J., Splinter, E., Grosfeld, F. & de Laat, W. Looping and interaction between hypersensitive sites in the active β -globin locus. *Mol. Cell* **10**, 1453–1465 (2002).
- de Laat, W. & Grosfeld, F. Spatial organization of gene expression: the active chromatin hub. *Chromosome Res.* **11**, 447–459 (2003).
- Spilianakis, C. G. & Flavell, R. A. Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nat. Immunol.* **5**, 1017–1027 (2004).
- Vernimmen, D., De Gobbi, M., Sloane-Stanley, J. A., Wood, W. G. & Higgs, D. R. Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *EMBO J.* **26**, 2041–2051 (2007).
- Brown, K. E. *et al.* Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. *Cell* **91**, 845–854 (1997).
- Chambeyron, S., Da Silva, N. R., Lawson, K. A. & Bickmore, W. A. Nuclear re-organization of the Hoxb complex during mouse embryonic development. *Development* **132**, 2215–2223 (2005).
- Skok, J. A. *et al.* Nonequivalent nuclear location of immunoglobulin alleles in B lymphocytes. *Nat. Immunol.* **2**, 848–854 (2001).
- Hewitt, S. L. *et al.* Association between the Igk and IgH immunoglobulin loci mediated by the 3' Igk enhancer induces 'decontraction' of the IgH locus in pre-B cells. *Nat. Immunol.* **9**, 396–404 (2008).

17. Lundgren, M. *et al.* Transcription factor dosage affects changes in higher order chromatin structure associated with activation of a heterochromatic gene. *Cell* **103**, 733–743 (2000).
18. Noordermeer, D. *et al.* Transcription and chromatin organization of a housekeeping gene cluster containing an integrated β -globin locus control region. *PLoS Genet.* **4**, e1000016 (2008).
19. Ragozy, T., Bender, M. A., Telling, A., Byron, R. & Groudine, M. The locus control region is required for association of the murine β -globin locus with engaged transcription factories during erythroid maturation. *Genes Dev.* **20**, 1447–1457 (2006).
20. Schoenfelder, S. *et al.* Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat. Genet.* **42**, 53–61 (2010).
21. Apostolou, E. & Thanos, D. Virus infection induces NF- κ B-dependent interchromosomal associations mediating monoallelic IFN- β gene expression. *Cell* **134**, 85–96 (2008).
22. Ling, J. Q. *et al.* CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science* **312**, 269–272 (2006).
23. Lomvardas, S. *et al.* Interchromosomal interactions and olfactory receptor choice. *Cell* **126**, 403–413 (2006).
24. Sandhu, K. S. *et al.* Nonallelic transvection of multiple imprinted loci is organized by the H19 imprinting control region during germline development. *Genes Dev.* **23**, 2598–2603 (2009).
25. Spilianakis, C. G., Lalioti, M. D., Town, T., Lee, G. R. & Flavell, R. A. Interchromosomal associations between alternatively expressed loci. *Nature* **435**, 637–645 (2005).
26. Lewis, E. B. The theory and application of a new method of detecting chromosomal rearrangements in *Drosophila melanogaster*. *Am. Naturalist* **88**, 225–239 (1954).
27. Pirrotta, V. Transvection and chromosomal trans-interaction effects. *Biochim. Biophys. Acta* **1424**, M1–8 (1999).
28. Bolzer, A. *et al.* Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol.* **3**, 826–842 (2005).
29. Bacher, C. P. *et al.* Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. *Nat. Cell Biol.* **8**, 293–299 (2006).
30. Hewitt, S. L. *et al.* RAG-1 and ATM coordinate monoallelic recombination and nuclear positioning of immunoglobulin loci. *Nat. Immunol.* **10**, 655–664 (2009).
31. Xu, N., Donohoe, M. E., Silva, S. S. & Lee, J. T. Evidence that homologous X-chromosome pairing requires transcription and Ctf protein. *Nat. Genet.* **39**, 1390–1396 (2007).
32. Branco, M. R. & Pombo, A. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol.* **4**, 780–788 (2006).
33. Strouboulis, J., Dillon, N. & Grosveld, F. Developmental regulation of a complete 70-kb human β -globin locus in transgenic mice. *Genes Dev.* **6**, 1857–1864 (1992).
34. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
35. Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.* **38**, 1348–1354 (2006).
36. de Wit, E., Braunschweig, U., Greil, F., Bussemaker, H. J. & van Steensel, B. Global chromatin domain organization of the *Drosophila* genome. *PLoS Genet.* **4**, e1000045 (2008).
37. Simonis, M., Kooren, J. & de Laat, W. An evaluation of 3C-based methods to capture DNA interactions. *Nat. Methods* **4**, 895–901 (2007).
38. Drissen, R. *et al.* The erythroid phenotype of EKLF-null mice: defects in hemoglobin metabolism and membrane stability. *Mol. Cell Biol.* **25**, 5205–5214 (2005).
39. Hodge, D. *et al.* A global role for EKLF in definitive and primitive erythropoiesis. *Blood* **107**, 3359–3370 (2006).
40. Welch, J. J. *et al.* Global regulation of erythroid gene expression by transcription factor GATA-1. *Blood* **104**, 3136–3147 (2004).
41. Femino, A. M., Fay, F. S., Fogarty, K. & Singer, R. H. Visualization of single RNA transcripts *in situ*. *Science* **280**, 585–590 (1998).
42. Cajiao, I., Zhang, A., Yoo, E. J., Cooke, N. E. & Liebhaber, S. A. Bystander gene activation by a locus control region. *EMBO J.* **23**, 3854–3863 (2004).
43. Lower, K. M. *et al.* Adventitious changes in long-range gene expression caused by polymorphic structural variation and promoter competition. *Proc. Natl Acad. Sci. USA* **106**, 21771–21776 (2009).
44. Fuss, S. H., Omura, M. & Mombaerts, P. Local and cis effects of the H element on expression of odorant receptor genes in mouse. *Cell* **130**, 373–384 (2007).
45. Nishizumi, H., Kumasaka, K., Inoue, N., Nakashima, A. & Sakano, H. Deletion of the core-H region in mice abolishes the expression of three proximal odorant receptor genes in cis. *Proc. Natl Acad. Sci. USA* **104**, 20067–20072 (2007).
46. Zhao, Z. *et al.* Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.* **38**, 1341–1347 (2006).
47. Raj, A. & van Oudenaarden, A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**, 216–226 (2008).
48. Muller, H. J. Types of visible variations induced by X-rays in *Drosophila*. *J. Genet.* **22**, 299–334 (1930).

METHODS

Gene targeting and generation of transgenic mice. Targeting of the human β -globin LCR to the mouse *Rad23a* gene has been described¹⁸. The human A γ -globin gene was targeted to the mouse *Rad23a* gene by substituting the ClaI Neo-resistance cassette from an existing *Rad23a* targeting construct by a 7.6 kb ClaI fragment containing a TK-Neo resistance cassette coupled to a ClaI-SmaI fragment containing the human A γ -globin gene with a gene encoding GFP at the translational start. Constructs with the ClaI fragment in two orientations were obtained: γ -globin-S (5'- γ -globin at the 5'-end of the *Rad23a* gene) and γ -globin-AS (5'- γ -globin gene at the 3'-end of the *Rad23a* gene). Targeting in Ola129-derived embryonic stem cells, blastocyst injection and breeding to obtain homozygous transgenic animals in an FVB background has been described¹⁸. Genotyping was carried out by Southern blot. Animal experiments were carried out according to institutional and national guidelines (Committee on Experiments with Laboratory Animals (DEC-Consult); Ministry of Agriculture, Nature and Food Quality, The Hague, The Netherlands).

Gene expression analysis. Total RNA was isolated using Trizol (Invitrogen) from tissues of at least two independent embryos. Complementary DNA was synthesized using SuperScriptII Reverse Transcriptase and oligo(dT)12–18 primer according to the manufacturers' instructions (Invitrogen). Products were quantified by qPCR, using platinum *Taq* DNA polymerase (Invitrogen) and SYBR Green (Sigma) on an Opticon 2 Real-Time PCR detection system (BioRad). Sequences of primers are provided (Supplementary Table S3). Transcript levels were normalized to the *Hprt1* transcript.

FISH. DNA FISH and cryo-FISH was carried out as described before^{18,35,49}. Bacterial artificial chromosome (BAC) clones (BACPAC Resources Centre) used to visualize genomic regions are listed in Supplementary Table S3. For DNA-FISH, two adjacent BAC probes, RP24-136A15 and RP24-319P23, were labelled with SpectrumGreen-dUTP (green, Vysis) or ChromaTide Texas Red-12-dUTP (red, Invitrogen). BAC probes for cryo-FISH were labelled with SpectrumGreen-dUTP (green) or ChromaTide Alexa Fluor 594-5-dUTP (red; Invitrogen). Probe specificity was confirmed on mouse spleen metaphase spreads. 500 ng of labelled probe was co-precipitated with 5 μ g of mouse Cot1 DNA (Invitrogen). Images were collected with a Zeiss Axio Imager Z1 epifluorescence microscope (100 \times plan apochromat, 1.4 oil objective) equipped with a CCD (charge-coupled device) camera. DNA-FISH images were analysed with Zeiss AxioVision software (Zeiss). Cryo-FISH images were analysed with Isis FISH Imaging System software (Metasystems). Filters used for DNA FISH: DAPI (Zeiss), fluorescein isothiocyanate, AF594 (Chroma). Filters used for cryo-FISH: DAPI, fluorescein isothiocyanate, RD-TR-PE (Zeiss). No bleed-through was detected and images were collected without saturation of intensities. The significance of co-localization (DNA FISH and cryo-FISH) was determined by applying a replicated goodness-of-fit test (G-statistic). The null hypothesis in DNA-FISH experiments was that integration of the LCR or LCR and A γ -globin gene resulted in similar co-localization frequencies as the wild-type 8C3–C4 alleles. The null hypothesis in cryo-FISH experiments was that associating regions identified in the 4C analysis had co-localization frequencies comparable to those of negative regions. In DNA FISH a minimum of 150 alleles were counted for each genotype; in cryo-FISH a minimum of 500 alleles were counted by a person not knowing the probe combination applied to the sections.

To simultaneously detect accumulated mRNA in the cytoplasm and the nuclear position of transcribed loci, RNA FISH was carried out⁵⁰ with some adjustments. E14.5 fetal livers were resuspended in 200 μ l PBS and 20 μ l was spotted on a poly-L-lysine slide (Sigma) and air-dried for 5 min at RT before fixation. For hybridization, slides were rehydrated into PBS plus magnesium and equilibrated in 50% formamide/2 \times SSC for 20 min. Hybridization was carried out three times overnight at 37 $^{\circ}$ C, followed by a series of post-hybridization washes and mounting in Prolong Gold antifade reagent (Molecular Probes). Images were collected on a Leica DM6000 Fluorescence microscope (100 plan apochromat, 1.4 oil objective), using Leica filters 11513888 (Y5), 11513887 (Y3), 11513890 (GFP) and 11513874 (DAPI), and analysed using Leica Application Suite software. RNA probes for *Hbb-bh1* and *Hbb-b1* were synthesized as described previously⁵⁰. A mixture of six different probes was labelled with Cy3 or Cy5 (GE Healthcare). Probe sequences are listed in Supplementary Table S3. BAC clones RP23-319P23 and 14567 (Incyte), analysing respectively the actively transcribed locus 8C3–C4 and the active α -globin locus, were labelled with SpectrumGreen-dUTP (Vysis) as described previously³⁵. For each combination of probes (*Hbb-bh1* and *Hbb-b1* with either 8C3–C4 or α -globin), two to five slides were independently hybridized and analysed. Images with five to five cells were checked for 'jackpot cells', with high levels of β *h1* (*Hbb-bh1*) transcript in their cytoplasm. After identification of these 'jackpot' cells, images for other probes were opened and cells were analysed for co-localization between the

active endogenous β -globin locus (visualized using *Hbb-b1* transcript probes) and the locus of interest (8C3–C4 or α -globin). Loci were scored as co-localizing if two spots overlapped or touched in a focal plane without black pixels in between. Counting results are summarized in Supplementary Table S2. Results were reproducible between different slides and between different investigators analysing RNA-FISH slides.

Automated image analysis. For this study, we developed software to automatically identify and analyse large numbers of microscopy images. In summary: we used the DAPI and *Rad23a* signal to define cell boundaries in each image (on average \sim 10–20 cells/image). Of all pixels inside cells in an image, we selected pixels with the 1% highest intensity in the β *h1* and β -major signal. Next, we determined the number of high-intensity pixels for every cell, which is expected to be roughly equal under the null hypothesis of homogeneous expression.

The R package EImage (ref. 51) was used for the automated image analysis. Within a z-stack, the focal point was determined by calculating the absolute difference between the actual and the blurred version of the image. Because this is highest in the image with the highest contrast, the focal point is the image where the sum of this difference is maximal. We binarized *Rad23a* and DAPI images by applying a threshold function (thresh) on both channels. This yields two images with binary information, which are combined into a single binary image by setting to a value of one pixels that are one in either or both of the original images (logical OR function). Subsequently, a distance map (distmap) is calculated using the binary image⁵². Cell boundaries were defined by carrying out a watershed (watershed) function on the distance map⁵³. 83 wild-type images yielded 1,757 cells; 149 hLCR-AS images yielded 1,563 cells; 214 α -globin images were analysed, yielding 2,108 cells.

Because of the differences in illumination settings between images, a non-parametric threshold rather than an absolute threshold was used for scoring expression levels. Of all pixels found within cells we selected the 1% with highest intensity. Homogeneous expression levels would return an equal distribution of pixels in the cells. However, we see enrichment of high pixels in a population of high-expressing cells. Cells were then classified as 'interacting' (active 8C3–C4 and β -globin loci overlap or touch), or 'non-interacting' (four nascent transcript signals (two 8C3–C4 and two β ^{nas}), but no overlap). The top 1% intense pixels in the β *h1* channel ('intense pixels') were defined and counted per cell. Per category, cells were sorted according to their number of intense pixels. Probability scores for the difference in distributions for interacting and non-interacting cells were calculated by a one-sided Kolmogorov–Smirnov test.

4C analysis. 4C sample preparation was carried out as described³⁵, with HindIII as a primary restriction enzyme, DpnII (wild type) or NlaIII (LCR-AS) as secondary frequent cutter and EcoNI as tertiary restriction enzyme for linearization. Sequences of primers are in Supplementary Table S3. Custom Nimblegen arrays carrying 4C probes for chromosomes 7, 8, 10, 11, 12, 14 and 15 were used (release mm6 of the UCSC build mouse genome; ref. 2). The liftOver tool (UCSC) was used to transfer probe position to the locations in release mm9 (NCBI37). Competitive hybridization of 4C product versus genomic DNA was carried out as described³⁵. Normalization of the array data was carried out with 'lowess', using standard settings⁵⁴. Subsequently, we corrected for differences in ranges between datasets. Because of the asymmetric, non-normal distribution of the data, standard variance normalization is not suitable; therefore, we used the median absolute deviation (a non-parametric estimator for the variance). For the visualization of DNA interactions we adapted the domainogram algorithm³⁶, enabling multiscale visualization of 4C signals. First, 4C data are binarized by calling probes that show a more than threefold increase of median absolute deviation over the median positive. Next we take the union (U) of the positive signals in the two replicates, which is used as input for subsequent analysis. A probability score for enrichment of positive 4C signals with a given genomic region is calculated by a binomial test, and for every chromosome we determine the probability (*P*) of observing positive 4C signals by dividing the total number of positive signals by the total number of probes. For a range of window sizes (*w* from 2 to 200) enrichment of positive signals is determined in window *w* with given *P*. By carrying out this analysis using sliding windows, we can analyse entire chromosomes and construct a multiscale chromosomal map by visualizing the matrix of $-\log_{10}$ *P*-values of chromatin interactions.

To identify differentially interacting regions we calculated the average of the two median absolute deviation normalized replicates followed by the running median (window size of 21) for both viewpoints for every *trans* chromosome. Windows with a difference of two (log 2 scale) were scored as differentially interacting. Genes within 50 kb of a differentially contacted probe were scored as differentially contacted genes. The resulting genes were intersected with a list of genes that were differentially regulated on EKLF (refs 38,39) or GATA-1 (ref. 40) deletion, or the 1% most highly expressed genes. Enrichment for EKLF and/or GATA-1 regulated genes among

differentially contacted genes was calculated by comparing this ratio to that of all EKLF and/or GATA-1 regulated genes among all genes.

Accession numbers. Data are available from Gene Expression Omnibus (GEO): GSE24614 and GSE5891.

49. Solovei, I. *et al.* Spatial preservation of nuclear chromatin architecture during three-dimensional fluorescence *in situ* hybridization (3D-FISH). *Exp. Cell Res.* **276**, 10–23 (2002).
50. Levsky, J. M., Shenoy, S. M., Pezo, R. C. & Singer, R. H. Single-cell gene expression profiling. *Science* **297**, 836–840 (2002).
51. Pau, G., Fuchs, F., Sklyar, O., Boutros, M. & Huber, W. EImage—an R package for image processing with applications to cellular phenotypes. *Bioinformatics* **26**, 979–981 (2010).
52. Kolountzakis, M. N. & Kutulakos, K. N. Fast computation of the Euclidian distance maps for binary images. *Inform. Process. Lett.* **43**, 181–184 (1992).
53. Vincent, L. & Soille, P. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *Pattern Analysis and Machine Intelligence, IEEE Trans.* **13**, 583–598 (2002).
54. Smyth, G. K. & Speed, T. Normalization of cDNA microarray data. *Methods* **31**, 265–273 (2003).

DOI: 10.1038/ncb2278

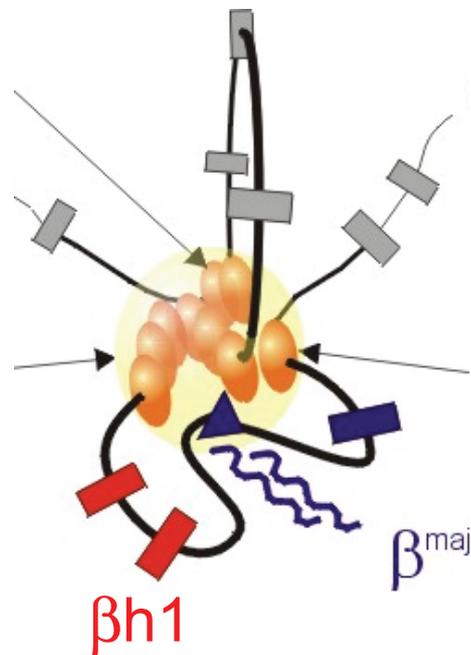


Figure S1 Spatial organization of the endogenous mouse β -globin locus. Cartoon showing the spatial topology of the mouse endogenous β -globin locus in definitive red blood cells at E14.5 and later stages of development, as inferred from 3C data. At this stage, $\beta h1$ is ignored by its own LCR in the fetal liver.

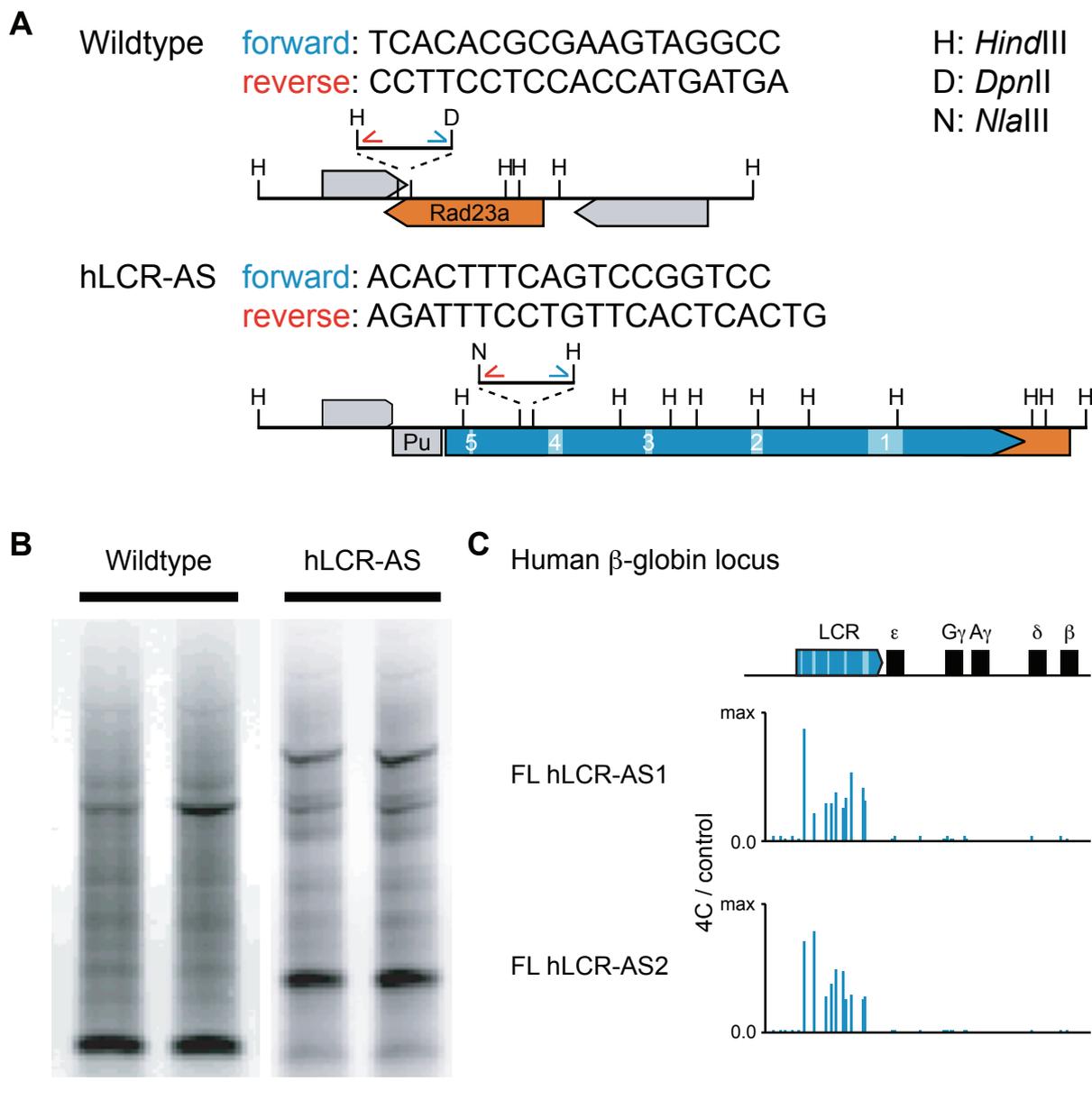


Figure S2 4C analysis of 8C3/C4 with and without the human globin LCR. (a) Location of restriction fragments and primer sets used for the amplification of 4C material from wildtype and transgenic animals. Different restriction fragments and primer sets have been used to analyze DNA interactions with wildtype 8C3/C4 and 8C3/C4 carrying the integrated human LCR in either orientation. Only relevant restriction sites of frequent cutters have been indicated. (b) Gel electrophoresis analysis of PCR amplified wildtype

and transgenic 4C material shows the reproducibility between replicate samples and the differences between amplified 4C material generated with different primer sets. (c) Normalized, non-smoothed, 4C signal intensities for probes mapping to the human δ -globin locus which was present as a control locus on the dedicated 4C microarray carrying probes for 7 mouse chromosomes. Positive signal is restricted to the human LCR, while probes mapping to other parts of the human globin locus are negative.

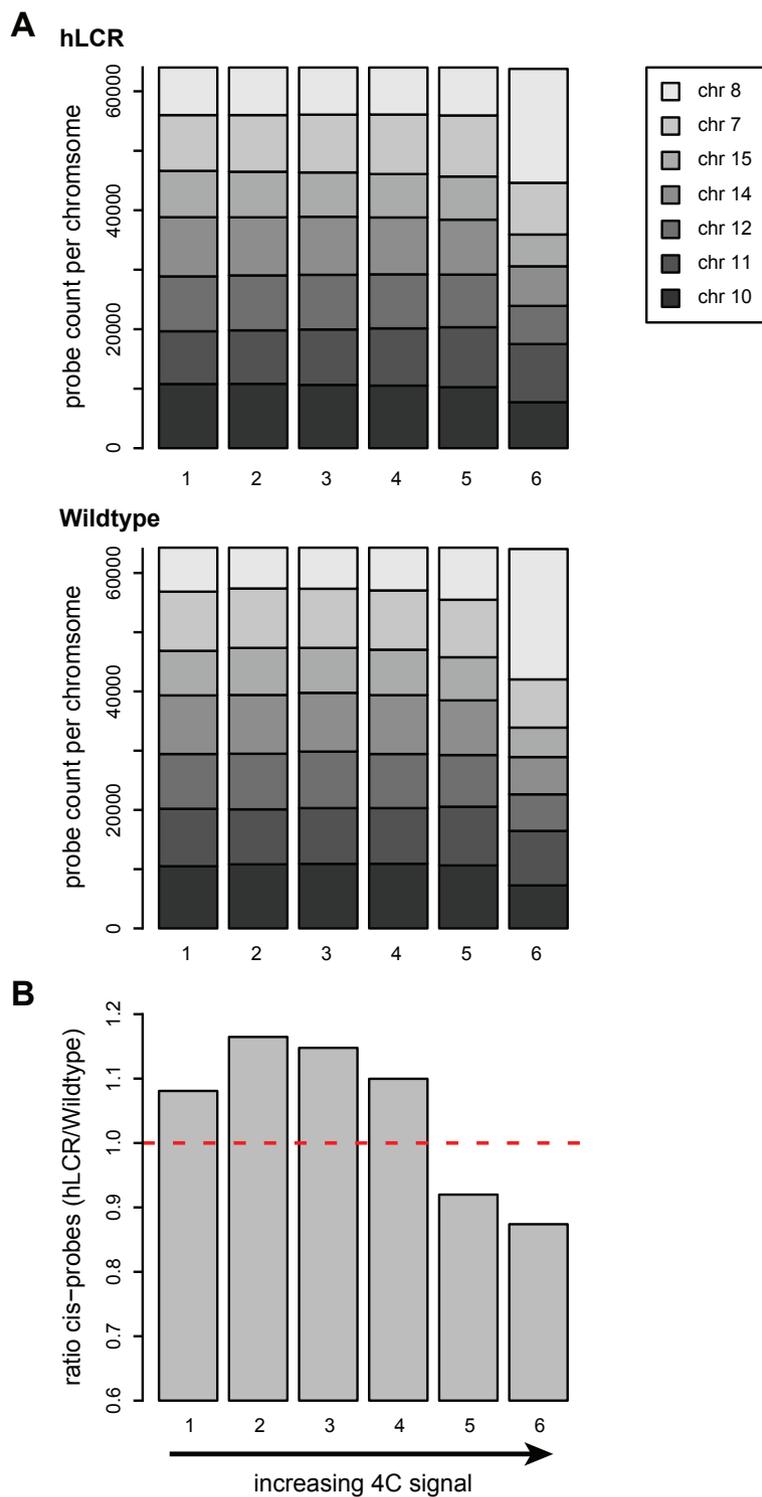
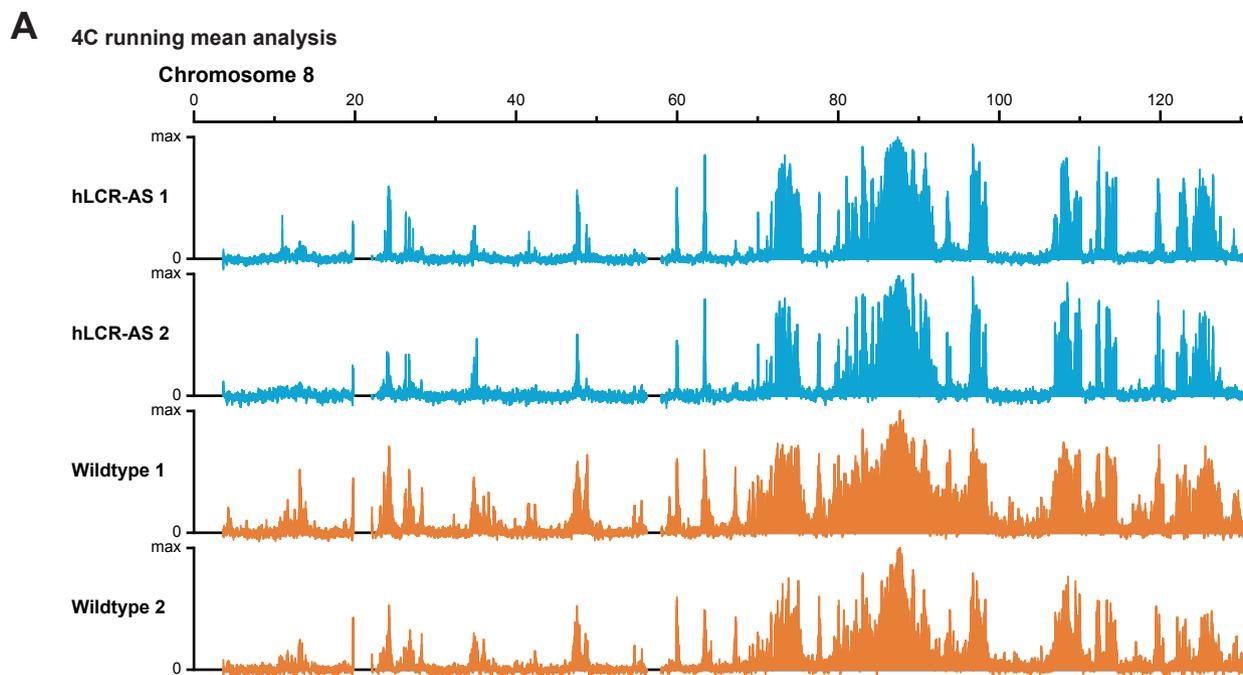


Figure S3 With the LCR, the ratio of high 4C signal on inter- versus intra-chromosomal probes increases. **(a)** Average normalized 4C signals over two replicates were sorted in an ascending manner for the wildtype and LCR experiment. Probes were subsequently divided into six equal-sized bins. Barplots show for each bin the frequency of probes for every chromosome. Note that in

bin 6 (probes with highest 4C signal) chromosome 8 (the *cis* chromosome) is overrepresented, but less so with the LCR than in wildtype. **(b)** Barplot showing the ratio of *cis* probe frequencies between hLCR and WT for the 4C signal bins. The relative decrease in the amount of *cis* probes with high intensities is proportional to more contacts with *trans* chromosomes by the hLCR.



B Correlation between 4C replicates:

		hLCR		Wildtype	
		<u>hLCR-AS 1</u>	<u>hLCR-AS 2</u>	<u>WT 1</u>	<u>WT 2</u>
hLCR	<u>hLCR-AS 1</u>	1	0.64	0.65	0.65
	<u>hLCR-AS 2</u>		1	0.64	0.59
Wildtype	<u>WT 1</u>			1	0.77
	<u>WT 2</u>				1

Figure S4 Intra-chromosomal contacts of 8C3/C4 without and with LCR are essentially similar. (a) 4C running mean data in *cis* of 4 independent E14.5 fetal liver samples, 2 with and 2 without the ectopic LCR. Despite

amplification with different primer sets, the overall chromosome-wide interaction pattern is comparable. (b) Pairwise correlation coefficients (Spearman's rank) between all experiments.

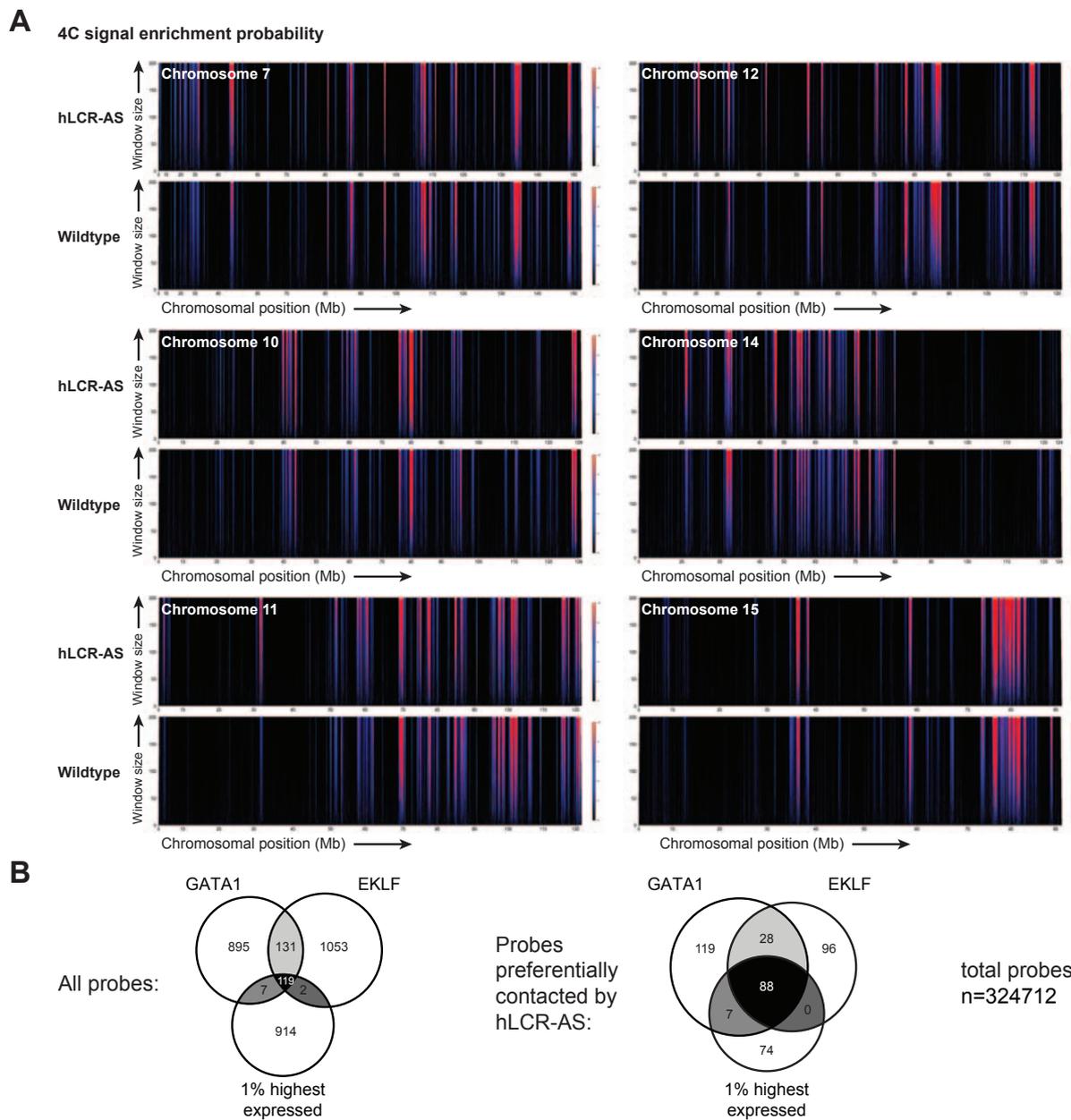


Figure S5 The genomic environment of 8C3/C4 without and with LCR is essentially similar, but the LCR preferentially interacts with genes controlled by shared transcription factors. (a) Domainograms showing interactions of 8C3/C4 with (top) and without (bottom) the integrated human LCR for all chromosomes

analyzed. Window size: 0-200 probes. (b) Venn diagrams showing the total number of probes at or around (-/+ 50 kb) the most highly expressed genes, GATA1-regulated genes and EKLf-regulated genes on the microarray (left) and among the regions preferentially contacted by the LCR (right).

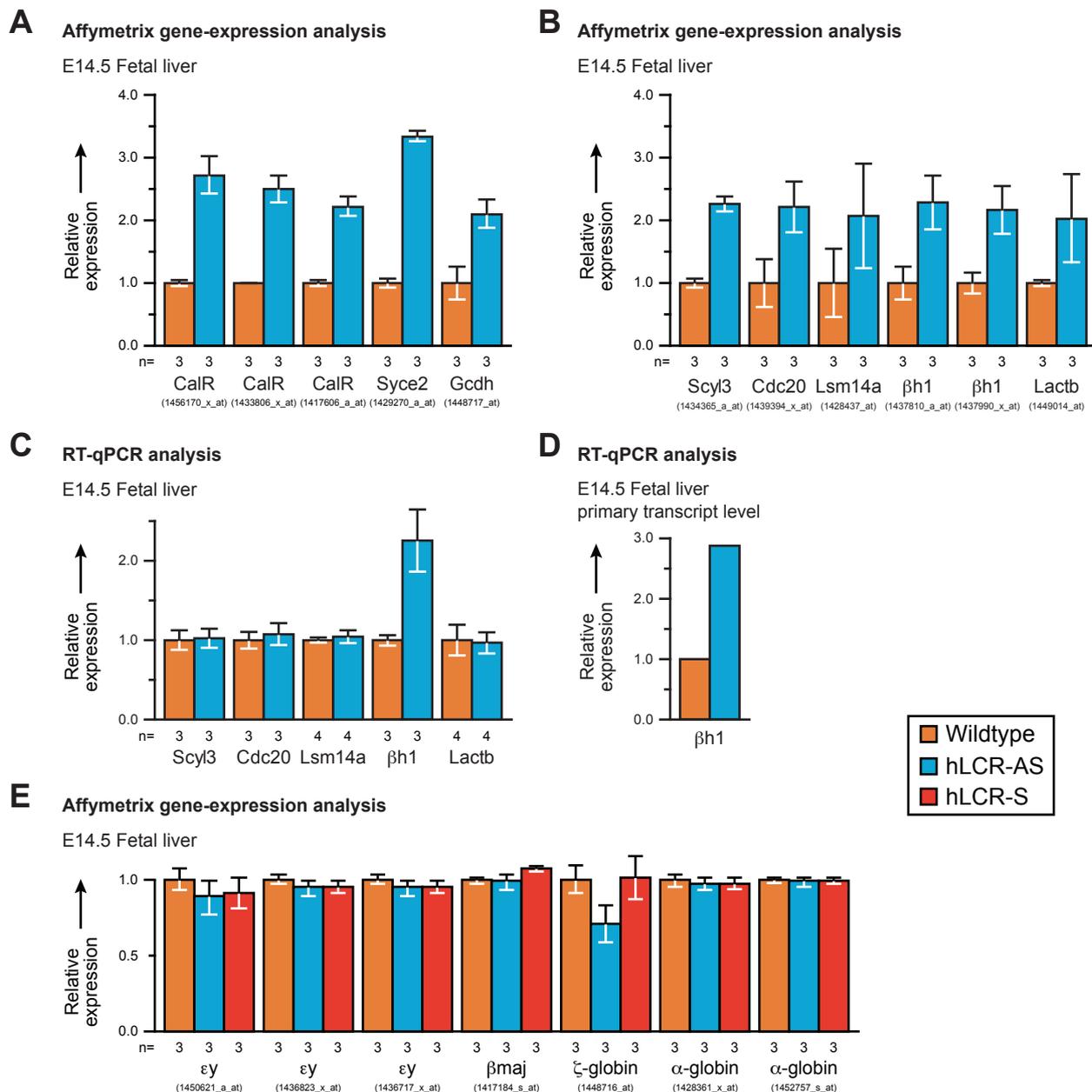


Figure S6 The ectopic LCR on chromosome 8 activates a single gene in *trans*. (a) Affymetrix gene-expression data of previously identified upregulated genes surrounding the integrated LCR. (b) Affymetrix gene-expression data identifying over twofold upregulated genes in *trans*. (c) Verification of upregulation in *trans* by RT-qPCR analysis identifies the endogenous beta1 gene as the only gene upregulated over twofold in *trans*.

(d) RT-qPCR of beta1 primary transcript levels demonstrating increased transcription activity of the beta1 gene at the nascent RNA level. Data from two independent samples. (e) Affymetrix gene-expression data for all probe-sets analyzing additional genes in the beta- and alpha-globin loci. Error bars: Standard Error based on the number of littermates indicated below (n).

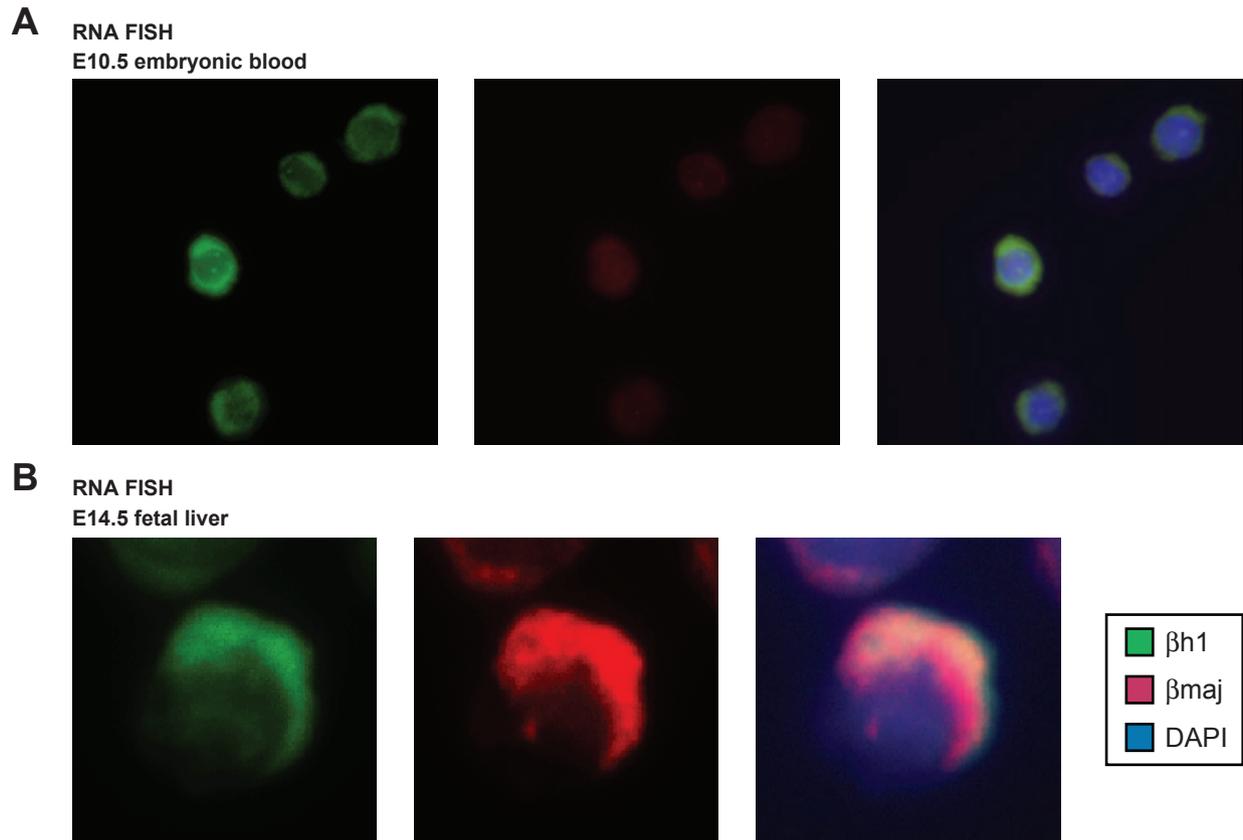


Figure S7 RNA FISH analysis with probes specific for $\beta h1$ and βmaj transcripts. In E10.5 embryonic blood, $\beta h1$ primary transcripts and mRNA is detected, but βmaj transcripts not. Scale bars: 2 μm

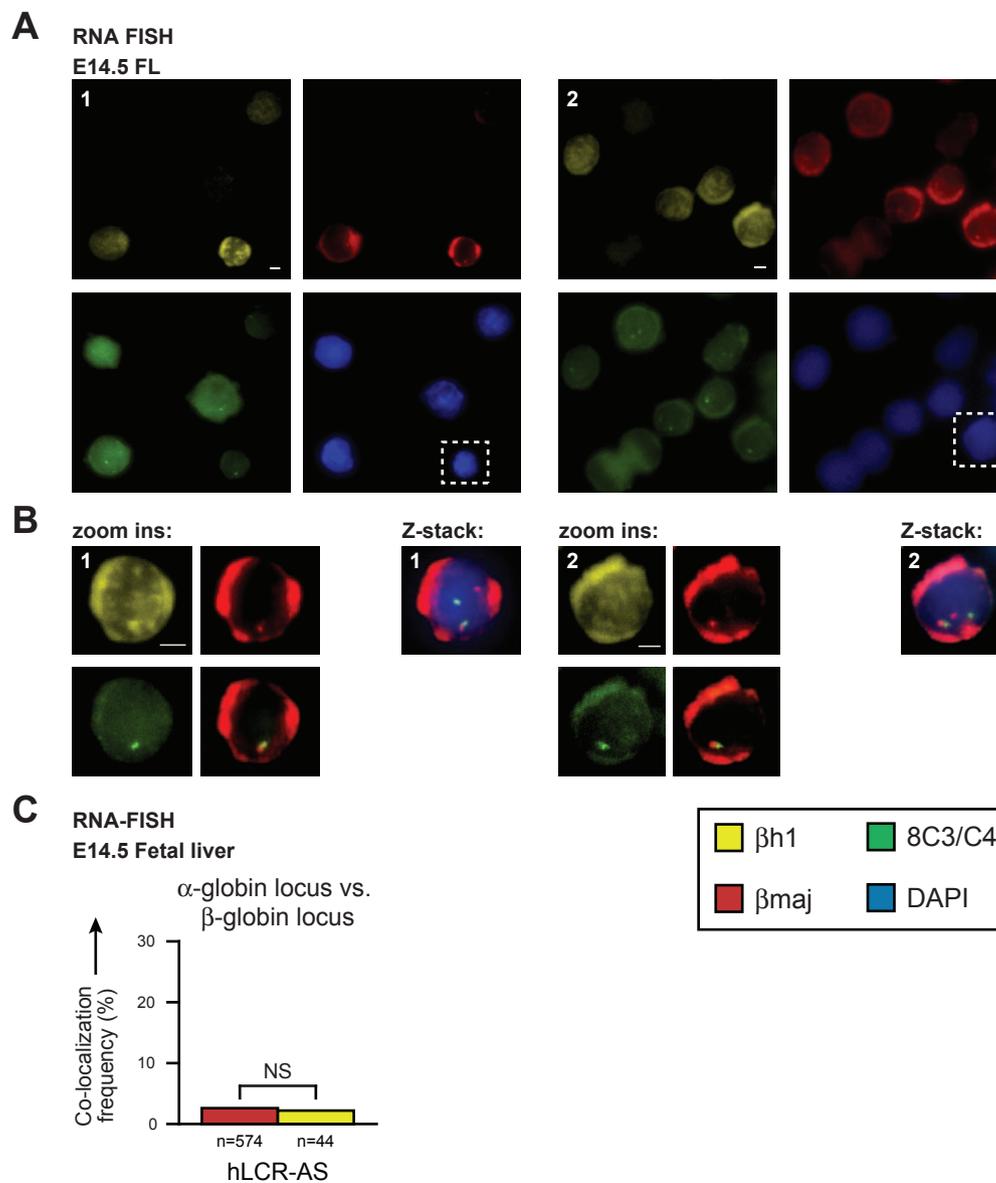


Figure S8 RNA-FISH showing increased β h1 mRNA levels in cells with inter-chromosomal LCR- β h1 interactions. (a) RNA-FISH on E14.5 fetal liver cells, with one cell in both fields showing strongly increased β h1 mRNA levels in the cytoplasm ('jackpot cell'). (b) Zoom in on the 'jackpot cell' highlighted in a., showing an inter-chromosomal interaction between the endogenous β -globin locus on chromosome 7 and the ectopic LCR on chromosome 8. The zoom-in shows separate images of three channels from a single focal plane and in the bottom right a image of the merged β maj and

8C3/C4 is shown. The z-projection shows the entire z-stack highlighting all primary transcript-signals for β maj and 8C3/C4. Scale bars: 2 μ m (c) No increased β h1 transcript levels in cells showing an interchromosomal interaction between the β - and α -globin locus. Quantification of inter-chromosomal interactions in LCR-AS transgenics between the β -globin locus on chromosome 7 and the α -globin locus on chromosome 11, in e14.5 fetal liver total erythroid cells and in ' β h1 jackpot cells'. Number of cells analyzed (n) is indicated.

Supplementary Tables

Table S1 Genomic regions that are differentially contacted when the LCR is integrated in 8C3/C4. Genomic regions are ranked based on their difference in 4C signal. The presence of EKLF and GATA-1 regulated genes in these regions is indicated.

Table S2 FISH results. (a) Interaction frequencies measured in E14.5 fetal liver cells as measured by RNA-FISH. (b) Interaction frequencies measured in E14.5 fetal liver cells, between the endogenous β -globin locus on chromosome 7 and 8C3/C4 on chromosome 8, without (WT) and with (hLCR-AS) the integrated hLCR. Measurements were made by cryo-FISH (which necessarily focuses on alleles), DNA-FISH (where interactions were scored only in cells showing signals for all 4 alleles) and RNA-FISH (where interactions were scored only in the 50-80% of red blood cells in the fetal liver (i.e. cells showing cytoplasmic β -major transcripts). Note that cryo-FISH gives much lower interaction frequencies, as expected from a technique providing ~5-fold higher resolution. Note also that frequencies measured by different techniques are difficult to compare, as they all score interactions differently.

Table S3 List of primers and BACs.

A. RNA-FISH results

Analyzed loci	Tissue	Total β -globin expressing cells	Colocalization in β -globin expressing cells		β h1 positive cells (among β -globin expressing cells)		Colocalization in β h1 positive cells	
			cells	%	cells	%	cells	%
8C3/C4 vs β -globin	hLCR-AS	775	66	8.5 %	50	6.5 %	15	30.0 %
8C3/C4 vs β -globin	Wildtype	1400	71	5.1 %	50	3.6 %	1	2.0 %
α -globin vs β -globin	hLCR-AS	574	15	2.6 %	44	6.7 %	1	2.2 %

B. Comparison of FISH results (interaction frequency 8C3/C4 – β -globin locus)

FISH technique	Tissue	Total cells	counted	Colocalization		Significance of difference (Fisher exact test)
				n	%	
Cryo-FISH	hLCR-AS	534	8C3/C4 alleles	12	2.3 %	Not significant
	Wildtype	262		6	2.3 %	
DNA-FISH	hLCR-AS	129	Cells with DNA-FISH signals	8	6.2 %	Not significant
	Wildtype	90		8	8.9 %	
RNA-FISH	hLCR-AS	775	Cells with β maj mRNA signal	66	8.5 %	p = 0.00437
	Wildtype	1400		71	5.1 %	

4C primers

Primer set	Sequence
Rad23a	TCACACGCGAAGTAGGCC CCTTCCTCCACCATGATGA
hLCR-AS	ACACTTTCAGTCCGGTCC AGATTCCTGTTCACTCACTG

Gene ID and Affymetrix probes

Gene name	Gene ID (ENSEMBL)	Affymetrix Probes
Rad23a	ENSMUSG00000003813	-
CalR	ENSMUSG00000003814	1417606_a_at 1433806_x_at 1456170_x_at
Syce2	ENSMUSG00000003824	1429270_a_at 1448717_at
Scyl3	ENSMUSG00000026584	1434365_a_at
Cdc20	ENSMUSG00000006398	1439394_x_at
Lsm14a	ENSMUSG00000066568	1428437_at
β h1 (Hbb-bh1)	ENSMUSG00000052217	1437810_a_at 1437990_x_at 1450736_a_at
Lactb	ENSMUSG00000032370	1449014_at
ϵ y (Hbb-y)	ENSMUSG00000052187	1436823_x_at 1436717_x_at 1450621_a_at
β maj (Hbb-b1)	ENSMUSG00000073940	1417184_s_at
β min (Hbb-b2)	ENSMUSG00000052305	-
ζ -globin (Hba-x)	ENSMUSG00000055609	1448716_at
α -globin (Hba-a1)	ENSMUSG00000069919	1428361_x_at 1452757_s_at

Gene expression analysis

Primer set	Sequence
mRNA Hprt	AGCCTAAGATGAGCGCAAGT ATGGCCACAGGACTAGAACA
mRNA Gapdh	TTCACCACCATGGAGAAGGC GGCATGGACTGTGGTCATGA
mRNA human A γ -globin	AGGTGCTGACTTCCTTGGG GGGTGAATTCTTTGCCGAA
mRNA β h1	TGGACAACCTCAAGGAGAC AGTAGAAAGGACAATCACCAAC
mRNA ϵ y	GAACTTGTCTCTGCCTCT ATCACCAGCACATTACCCA
mRNA β maj	ATGCCAAAGTGAAGGCCCAT CCCAGCACAATCACGATCAT
mRNA β min	ATCCCAAGGTGAAGGCCCAT CCCAGCACAATCACGATCGC
mRNA ζ -globin	GAGAGAGCTATCATCATGTCC AAGTAGGTCTTCGTCTGGG
mRNA α -globin	TGGCCATGGTGTGAATATG TCTTGCCGTGACCCTTGAC
Primary transcripts β h1	TCTGGGAGTTGAGACTGTGA TGGACCCATGGACTCTAACA
mRNA Scyl3	CTTACCATCTGGACTTGCTG GGGTGACGGAGTGTCTTTA
mRNA Cdc20	TGCTCCATCCTCTGGTCT CGTGCTGTGTGCCTTTG
mRNA Lsm14a	CCACCCAAACCACAATGT GGACTGAACTGACTGTATGC
mRNA Lactb	CGTGGTTGGAGTTTCTGTAG TGCTGATGCTTGCGATTC

BAC clones for DNA-FISH and Cryo-FISH

BAC clone	Locus (Cytological band)	Chromosome	Position (Mb) (NCBI assembly m37)
RP23-317H16	7D3, negative control	7	93
RP23-370E12	7E3, β -globin locus	7	111
RP23-32C19	7F3, negative control	7	131
RP23-265I23	7F3	7	134
RP23-27B18	7F5	7	148
RP24-136A15	8C3 (previously 8C3/C4)	8	87
RP24-319P23	8C3 (previously 8C3/C4)	8	87
RP23-87K3	10B1	10	41
RP24-130O14	10B5.3, negative control	10	74
RP23-375D18	11A4, α -globin locus	11	32
14567 (Incyte)	11A4, α -globin locus	11	32
RP24-236L11	11B5	11	78
RP23-311P1	11D/E1	11	102
RP23-258M10	14C3	14	56
RP23-450E9	14D2	14	70
RP24-255K10	15B3.1	15	37

RNA-FISH probes

Probe	Sequence
β h1 5'UTR	GTGAGGTCTAGAAGCTTGGAGATGATCTCAAGTGTGCAAAAGCCAGAATG
β h1 Exon 1	TCCAAGTCCACTTTATCCCAGATGCTTGTGATAGCTGCCTTCTCCTCAGC
β h1 Intron 1	TGTGCCACAAAACCCTATAGAAACCCTGGAAATTTCTGCCATGCATAAGG
β h1 Intron 2	TCAAACCCAAGGCCCCAGAAATGCTGGGCGCTCACTCAAATCTGCACCCA
Alternate β h1 3'UTR 1	GGGCATCATAGACACATGGGATTGCCAGTGTACTGGAATGGAGTTTAATG
Alternate β h1 3'UTR 2	TTAAACAAAATTTGTGCTCTCAATGCCTAATCCAGTCCCCATGGACTCAA
β maj 5'-UTR1	CTGTTTCTGGGGTTGTGAGTCAACACAACATATGTCAGAAGCAAATGTGAG
β maj Exon1	TTCATCGGAGTTCACCTTCCCCACAGGCAAGAGACAGCAGCCTTCTCAG
β maj Intron1	GGAAAGCAGACCTCTGTCTCCAAGCACCCAACCTTCTTCTGTGAGCTGCC
β maj Intron2	ACTCCACACACAGTCATGGAGACTGCTCCCTAGAATCGCTTCCCCTTTTT
Alternate β maj 3'-UTR1	CTTGGGAACAATTAACCATTGTTTACAGGCAAGAGCAGGAAAGGGGGTTT
Alternate β maj 3'-UTR2	AGAAGACAGATTTTCAAATGTCTATCATTTTGCCAACAACCTGACAGATGC